Graphics and User's Exploration via Simple Sonics (GUESS): Providing Interrelational Representation of Objects in a Non-visual Environment

Hesham M. Kamel

The Group for User Interface Research
EECS, UC Berkeley
Berkeley, Ca. 94720, USA
Hesham@eecs.berkeley.edu

Patrick Roth

Computer Science Department
CUI, University of Geneva
CH - 1211 Geneva 4, Switzerland
Patrick.Roth@cui.unige.ch

Rashmi R. Sinha

SIMS, UC Berkeley
Berkeley, CA 94720, USA
sinha@sims.berkeley.edu

ABSTRACT

In this research we investigated the use of the GUESS system in the exploration of auditory pattern perception by blind and visually impaired people.

We have compared three different techniques for presenting graphical scenes via non-speech sounds: one based on the physical tablet, one on the virtual-sonic grid, and one on sound localization techniques. In each technique we utilized a 2D sound plane to represent different geometric shapes. As an input device, we used a graphical tablet in order to explore the images rendered.

We have conducted a pilot study with three groups of four participants each. Our results have shown that with the second and third techniques, blind people were able, within a relatively short space of time, to precisely identify the interrelation of simple geometric shapes. They have also shown that, in the second technique, assigning a non-speech sound to a region located in the center of the tablet reduced the navigation time when relocating specific shapes. As to the first technique, it received the lowest time rating for relocating objects.

Our findings indicate that the method of presenting interrelation in auditory interface designs does indeed play an important role in assisting users comprehend the diagrams communicated.

1. INTRODUCTION

Vision plays a preponderant role in our knowledge of the world. The majority of image representations are created from visual perception. This natural method of graphical interaction seems to exclude blind and visually impaired people who lack the sense of vision.

Recently, triggered by the advent of multimedia computing, various investigations have been made into the use of non-speech sounds to convey visual information. Mynatt [14] has investigated the mapping of spatial graphical display into a hierarchical auditory interface. In the Mercator system, which is an X-window-based screen reader for the blind, the user interacts with the graphical user interface (GUI) via a standard computer keyboard. The system provides an auditory feedback that corresponds to the graphical component (e.g. menu, dialog box) selected. Mynatt prefers to work with everyday sounds as elements of the auditory feedback. James [8] similarly utilizes audio as a method of non-visual communication through which to represent HTML. This method allows blind people to conceptualize the different layout of Web pages, and, furthermore, creates a framework for understanding how to represent document structure with audio that combines both speech (textual content) and non-speech (tag) sounds.

The above two systems share a similar method of feedback: transforming spatial information into one-dimensional response. These methods of transformation, however, cannot be utilized to fully communicate meaningful spatial representations such as maps or geometric patterns.

In this paper we describe the design and use of GUESS, an audio Haptic system that represents spatial layout of images to blind and visually impaired people. These rendered images are composed of simple geometrical shapes (e.g., square, circle, and triangle). Our system employs three different techniques with which to provide the blind user with access to information regarding both the type of shape and the spatial organization of the different shapes. In the first technique, the user depends for his/her perception of spatial relations on the navigational dimension (absolute positions) of the physical tablet. In the second technique, user-perception is assisted by our mapping a 3×3 virtual-sonic grid with unique central point of reference onto the interactive area of the tablet. In the third technique, the absolute position of the object is determined by sound localization in the 2D sound spectrum, localization manipulated by the relative position of the pointing device.

In this paper we report the preliminary evaluation, which we conducted with twelve blind and visually impaired participants. In section 2, we introduce other work that has been done to represent spatial information of graphics via auditory feedback. In section 3, we describe the design of the GUESS system. Section 4 covers the system implementation and hardware configuration. Sections 5 and 6 report our experiments and present a discussion of our preliminary findings. Finally, in section 7, we draw our conclusion.

2. DESIGNS OF AUDITORY INTERFACES VIA NON-SPEECH SOUNDS

Other work has been done addressing the spatial representation of graphical information. Bly [3] was the first to report the use of audio in human computer interfaces. The non-visual representation of maps and diagrams has been investigated by linking touch (using graphical tablets) with auditory feedback [7, 11]. Kennel [11] represented diagrams (e.g., flowcharts) to blind people using multi-level audio feedback and a touch panel. Touching particular objects (e.g., diagram frames) and applying different pressures triggered three types of feedback. The first type of feedback was information regarding the frame; the second was the interrelation between frames; the third expressed the textual content of the frame via the use of speech. Jacobson [7] used a similar technique to represent maps. He designed the feedback functionalities, both speech and non-speech sounds, as different sectors surrounding the area of the represented map on a touch-sensitive screen. Depending on the sector selected, the blind user can obtain different types of information while exploring the map. For instance, by selecting the non-speech functionality, then touching any of the structured areas (e.g., a building or park), a corresponding non-speech sound is produced. The user can then go on to obtain more details, such as verbal explanation, by choosing the corresponding sector.

Directly using the physical properties of the sound was another method used to represent spatial information. Mansur [12] and Flowers [5] represented a graph by using a single tone where its pitch corresponded to the Y coordinate and its duration to the X coordinate. Meijer [13] designed a system that used time-multiplexed sound to represent a gray level picture. In his system, each pixel was associated with a sinusoidal tone, where the frequency corresponded to the pixel's vertical position and the amplitude corresponded to its brightness. Each column of the picture was defined by superimposing the vertical tones. The final signal was obtained by concatenating each resulting column.

In contrast, Hollander [6] represented shapes using a virtual-sound space. This space was defined by an array of speakers that directly mapped the visual counterpart. Each pattern was rendered by a moving sound source, which traced out the segments that belonged to the pattern.

3. THE GUESS SYSTEM

In the GUESS system we utilize an audio-Haptic approach that is based on a 2D sound environment, to provide blind users with the topological information of a given image. Using a stylus as an input device, the blind user interacts with a graphical tablet to explore an image that is composed of three simple geometric shapes (e.g., square, circle, triangle). During this exploration, the user can obtain two types of auditory information. The first type represents the contour of the shape (see section 3.1.), while the second type provides the spatial layout of each shape relative to the whole image. Thus, there are two distinct components of the GUESS output system: the sonic representation of the shapes, and the orientation rendering techniques (see sections 3.2.1-3.2.3).

3.1. The GUESS Interface

In order to represent a geometrical object in terms of auditory cues, we investigated Blauert's approach based on sound localization in a 3D sound space using headphone [2]. In our approach, a shape is rendered by a moving sound that is acoustically drawn within a 2D virtual sound space [16]. This plane is vertically oriented and located in front of the blind listener. For instance, if the rendered shape is a right triangle, the user hears a tone descending vertically in the right channel. Next, the sound moves horizontally from the bottom right to the bottom left channel. Finally, it ascends from the bottom left back to the top right channel, reaching the initial position (see figure 1).

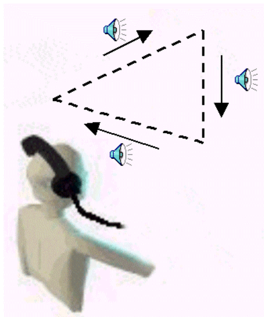


Figure 1. *Auditory rendering of the triangle*.

Concerning auditory perception, much research points out the difficulty of localizing a sound on the elevation plane [18]. As a solution to this problem, we incorporated two unique auditory features into our initial design:

- We reinforced the elevation rendering by means of a frequency variation [5, 12, 17]. Thus, a decrease in height will be perceived commensurate with frequency decrease.
- We utilized an alarm sound (beep) to code particular line junctions for patterns composed of several lines (e.g. square, triangle) (see figure 2.).

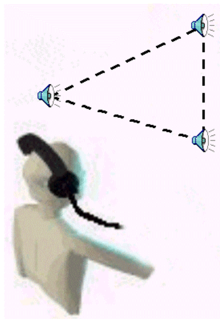


Figure 2. *Auditory coding for line junction.*

3.2. The Location and Visualization of Objects

We have implemented the GUESS system using the auditory shape-rendering method discussed above. In the following subsections we describe the three different techniques of the graphical pattern localization that we employ in the GUESS system.

3.2.1. Localization Using the Tablet Positions

In order to locate shapes using the first technique, we depend on the unique physical positions of the tablet. We assign auditory cues to non-overlapping areas. These cues take on different characteristics depending on the geometric shapes (i.e., we assign meaningful auditory labels to each of the areas [7, 11]). When the stylus touches a labeled area (e.g., a triangle), it triggers its unique auditory representation. Silence is introduced when the stylus is located in a non-labeled area. In order to locate a shape using this technique, the user must search for its physical location on the tablet: the only guidance that the blind user has is provided by the stylus entering the area of the shape.

3.2.2. Localization Using The Virtual Grid

In the second technique, we utilize a grid-based model to increase panel exploration precision [9, 10]. We map a 3×3 virtual grid onto the tablet. The grid consists of nine non-overlapping regions, which are based on the layout of the telephone keypad, where "1" is the top left cell of the grid and "9" is the bottom right cell (see figure 3) [9]. We assign to the vertical and horizontal axes the sound of two different musical instruments [1]. In order to communicate the location of any individual cell on the grid, two different notes played consecutively characterize the boundary between any two cells. When crossing all vertical boundaries from left to right (e.g., from cell 1 to 2), two clarinet tones (from low to high) of long duration are played. When crossing horizontal boundaries from top to bottom (e.g., from 1 to 4), two vibe tones (from high to low) of short duration are played. The two clarinet notes are increased in pitch as the stylus moves from left to right; similarly, the vibe notes decrease as the stylus moves from top to bottom. As the stylus moves in the opposite directions, so all of the tones are played in the reverse order. Since the center cell (position 5) is a unique point of reference [9], we assign it a drum sound for easy and fast identification (see figure 3).

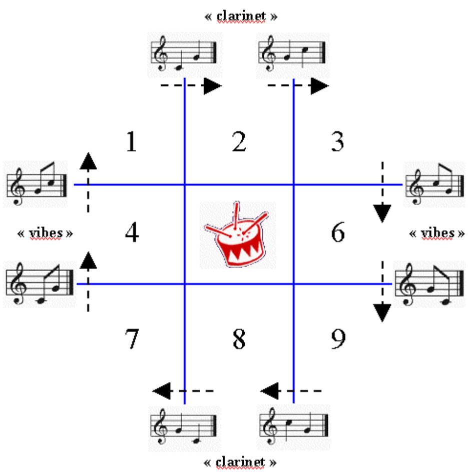


Figure 3. *Audio grid rendering*.

Using this technique, users can locate a shape by having the stylus touch its corresponding area, as described in section 3.2.1. In addition, the grid can be used as a frame of reference when relocating a shape by recognizing that shape's corresponding cell position (see figure 4).

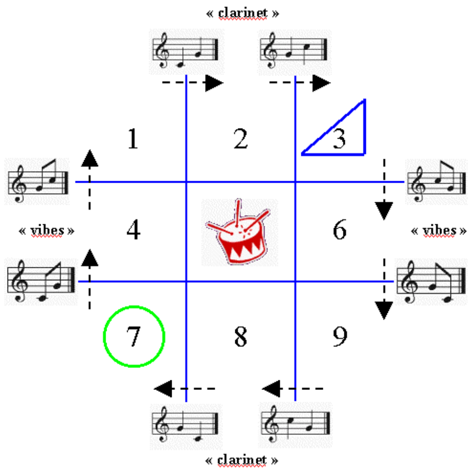


Figure 4. *Locating shapes via grid positions*.

3.2.3. Localization Using the 2D Sound Spectrum

In the third and final exploration technique, the presentation of the full image is incorporated into the 2D virtual sound plane [15]. All of the shapes that compose the image are represented by their corresponding auditory cues (see section 3.1). Each location of an auditory cue in the 2D sound plane is manipulated by the relative position of the stylus to the corresponding shape on the tablet. As the stylus gradually approaches or moves away from the center of a shape, its audio rendering becomes louder or softer respectively. Thus, the user obtains the complete audio rendering of the shape when the stylus is at its exact center position (see figure 5).

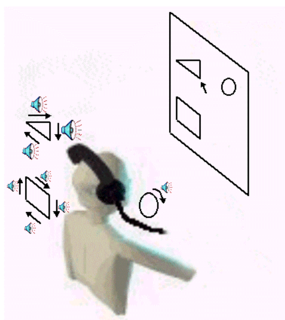


Figure 5. *2D Sound localization rendering.*

Relocating a shape using this technique is guided by the amount of the distance between the stylus and the shape position.

4. HARDWARE CONFIGURATION

The evaluation of the GUESS interface was made on a Pentium PIII equipped with a SoundBlaster Live sound card with 64MB RAM under Windows 98. A set of Sennheiser headphones was used. In the 3D sound rendering, we used the DirectSound3D library available from Microsoft. Finally, for the Haptic interaction, we used an Aiptek A5 graphical tablet.

5. EXPERIMENT

The participants in our experiment consisted of both totally blind and partially sighted people. Their ages ranged between twenty-two and fifty-nine years of age. No participant had prior experience using a physical tablet. We assigned one technique to each of three groups (G1, G2 and G3), with four participants per group. Each group was asked to complete the same three tasks using the rendering technique it was testing. G1 participated in testing the localization via the physical tablet technique (see section 3.2.1). G2 examined the localization using the virtual grid (see section 3.2.2). Finally, G3 tested the technique of sound localization using the 2D sound spectrum (see section 3.2.3). Upon completing all tasks, each participant was asked to give his or her personal comments about the technique they used.

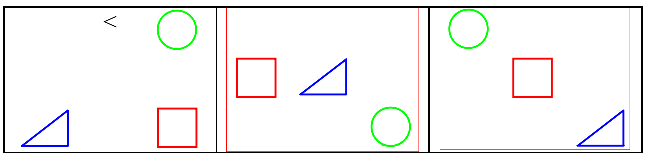
5.1. Methodology

All participants were given ten minutes in which to familiarize themselves with the rendering technique they were about to test, as well as with the auditory representation of the shapes. The participants of G2 were also introduced to the rendering of the virtual grid. All participants were allowed as much time as they needed to complete each task. While testing, participants were not allowed to feel the tablet with the hand they were not using. The partially-sighted participants were blindfolded for the entire duration of the experiment. Upon completing the tasks, each participant was asked to give a description of the spatial layout, then rate their confidence level. No hints were given at any time. Task completion times and the participants' comments were noted down. All sessions were audio recorded for further analysis.

5.2. Tasks

Each of the participants was given the three tasks consecutively. After exploring each task, they were instructed to take their hand off the tablet, then relocate one specific shape out of the three that were rendered.

For the first task, they were asked to find the square shape (see figure 6.a). In the second task, they were asked to find the triangle (see figure 6.b). In the third task they were asked to find the circle (see figure 6.c).



Figures 6.a. *Task1*; 6.b. *Task2*; 6.c. *Task3*.

6. RESULTS AND DISCUSSION

Preliminary evaluation of the GUESS system is described in figures 7 - 9. Figure 7 shows the mean time taken to explore and relocate shapes averaged across for all tasks by the three groups.

The results shown in figure 7 reveal that Group G2, using the virtual grid, spent the shortest time in locating and relocating shapes in all three tasks. On average, Group 2 spent 1.6 minutes locating shapes and 0.78 minutes relocating them. Group G3, using the localization technique, obtained the second best time in locating and relocating shapes: 2.9 and 2.3 minutes respectively. Group G1, using the absolute positioning technique, spent the most time: 3.9 minutes locating shapes, and 3.8 minutes relocating shapes. We did Pairwise Independent Sample t-tests to find out if the above differences were significant at the .05 level. For the exploring stage, t-tests show that the three groups differed significantly from each other at the .05 level: G1 & G2 $[t(6)=-14.45, p<.05]$; G1 & G3 $[t(6)=7.36, p<.05]$; G2 & G3 $[t(6)=-8.46, p<.05]$. For the Relocation Stage, Independent Sample Pairwise t-tests show that the three groups differed significantly from each other at the .05 level: G1 & G2 $[t(6)=-16.64, p<.05]$; G1 & G3 $[t(6)=-8.66, p<.05]$; G2 & G3 $[t(6)=-10.05, p<.05]$.

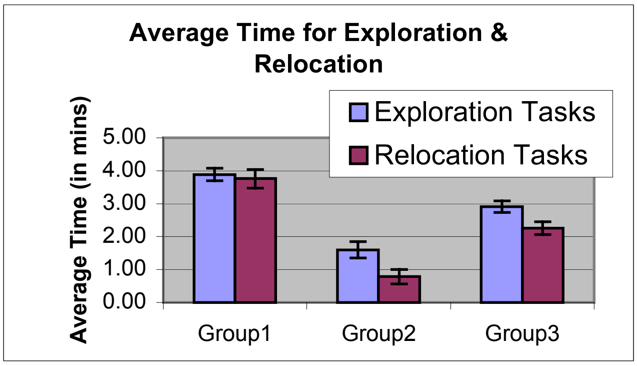


Figure 7. *Grand means and standard deviation of time taken in minutes to explore and relocate shapes.*

We also computed mean percentage decrease in time from exploring to relocation. The percentage decrease in time is the relative difference in time taken to complete the exploration and relocation for all tasks. G1 has the smallest decrease in time, 3.2%, while G2 has the largest decrease in time, 51.3%. G3 shows an intermediate difference in time, 22.3%. Thus the most learning occurred for G2.

Next, we analyzed participants' confidence in performance. As Figure 8 shows, G2 participants were the most confident about their performance. G3 participants showed the next highest level of confidence, while G1 participants showed the lowest level of confidence. Independent Sample Pairwise t-tests showed that the G1-G2, and G2-G3 differed significantly from each other, but G1-G3 did not significantly differ from each other. G1 & G2 $[t(6)=-3.97, p<.05]$; G1 & G3 $[t(6)=-1.80, p>.05]$; G2 & G3 $[t(6)=2.21, p<.05]$.

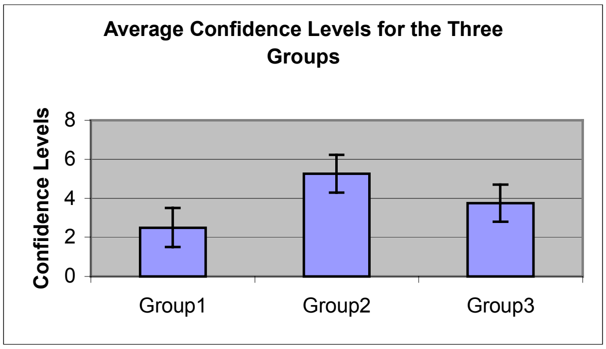


Figure 8. *Average confidence levels for the three groups.*

Finally, we looked at the accuracy of the task descriptions provided by the study participants. Figure 9 shows that two out of the four participants in G1 were not able to correctly describe the different patterns of task presented to them. One participant in G3 was unable to provide a correct task description. None of the G2 participants encountered any difficulty in describing the tasks presented.

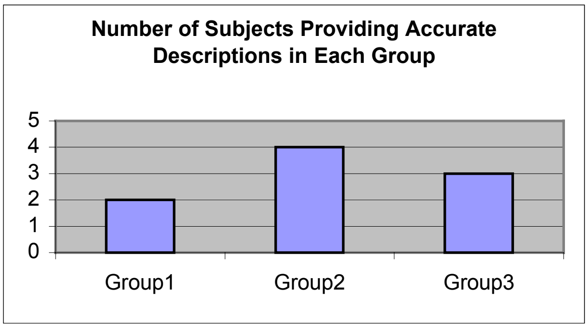


Figure 9. *Number of participants who could provide accurate task descriptions in each group.*

Overall, participants in G2 obtained the best performance in terms of time takes, accuracy of task descriptions, and confidence levels. One participant in G2 commented that the grid was useful when navigating, and the center point played an especially important role in building the mental model of the images rendered. Most of the participants in G2 appreciated the fact that the center cell was "marked" with an assigned sound. One participant commented, "once I found the center, it was easy for me to find any shape around it". However, another said, "when I was going from square one to two and from two to three I could not tell the difference between the tones … I had the same problem going from one to four and four to seven". A possible solution to this problem would be to add stereo effect on the current audio horizontal boundaries rendering. The effect of this would be that the user, while going from, for example, cell 1 to cell 3, would hear two clarinet tones in his left channel (from 1 to 2), and followed by two clarinet tones in his right channel (from 2 to 3).

Group G3 (using the localization technique) obtained the second best overall performance. One participant commented "I knew where one shape was in relation to another shape, but I did not know precisely where it was in relation to the tablet". This problem can be resolved by adding an auditory cue to inform the user when the stylus enters a shape.

Others expressed their concern with what is called "the cocktail-party effect" [4]: residual sound from one source interfered with their visualization of another sound source.

Group G1 (using the absolute positioning technique) produced the least successful performance. They were the least confident; half the group was not able to provide accurate task descriptions. This technique required much effort and concentration on the part of the users. The absolute positions on the interactive area of the tablet were not sufficient for G1 to precisely determine the locations of each shape. It was difficult for them, therefore, to visualize the location of each shape in relation to the others. In addition, members of this group were least comfortable with being confined to the use of one hand.

These preliminary findings suggest that spatial information should be communicated when designing auditory interfaces. We should provide blind users with a technique to relate one shape to another, similar, perhaps, to the "landmark" technique: blind people who are cane-travelers use a "landmark" to assist them in identifying a specific target area (e.g., count two trees on the right side then the third door on the left is the pet shop). In other words, for blind people, visual cues are replaced by casting a visual area of interaction with a memorable discrete structure. The interpretation of such cues, however, is as crucial to blind people as to sighted people, but in a non-visual manner.

7. CONCLUSION

In this paper we have presented the design of an auditory Haptic interface as well as its users' evaluations. We have shown that the lack of interrelational representation of diagrams in a non-visual environment could affect the final mental perception level of blind people. Audio representation alone of a given shape, we discovered, is not sufficient for relocating it when relying on the use of the entire physical area. Additional and clear information regarding absolute and relative positions of shapes rendered must be indicated. It is preferable to partition the interactive workspace into meaningful discrete subsections when using a physical device as a means of graphical interaction. Creating a label to designate the center position as a unique point of reference can further reduce exploration time. The use of sound localization is marginally effective in identifying objects: additional features would be required in order to determine positions in relation to the whole diagram.

We can provide most blind users with access to GUIs equal to that of sighted people. This can be achieved by designing an interface to match their capabilities and limitations, rather than obliging them to adapt to an uncongenial user-interface.

8. ACKNOWLEDGMENTS

The authors are grateful to Lori Stefano Petrucci and Professor Thierry Pun from the University of Geneva for their help with the design of the GUESS program. Many thanks to Rose Mary Obrist from Zurich, Diana Sallam from California, and Frederike Breuning from Berlin for their personal support. Thanks to J. Finbarr Nugent and Peter Baxter for their help with this paper.

9. REFERENCES

[1] Alty, J.L., Rigas, D.I. Communicating Graphical Information to Blind Users Using Music: The Role of Context in Design, CHI 98 Proceedings, April 1998, Los Angeles, CA, pp. 574-581.

[2] Blauert, J. Spatial Hearing, MIT Press, MA, 1983.

[3] Bly, S. A., Communicating with Sound, CHI 82 Proceedings, ACM Press, Addison-Wesley, New York, pp. 371-375.

[4] Bregman, A. S. Auditory Scene Analysis. Cambridge, MA: MIT Press, 1990.

[5] Flowers, J. H., and Hauer, T. A. Musical versus visual graphs: Cross-modal equivalence in perception of time series data, Human Factors, 37(3), 1993, pp. 553-569.

[6] Hollander, Ari. (1994). An Exploration of Virtual Auditory Shape Perception, Masters' Thesis, University of Washington, online http://www.hitl.washington.edu/publications/hollander/home.html.

[7] Jacobson, R.D. Navigating maps with little or no sight: A novel audio-tactile approach. Proceedings of Content Visualization and Intermedia Representations 98, August 98, Montreal, online http://www.fsu.edu/~geog/jacobson/dan_jacobsoncvir.pdf.

[8] James, F., Lessons from Developing Audio HTML Interfaces, ASSETS 98 Proceedings, April 1998, Marina del Rey, CA, pp. 27-34.

[9] Kamel, H., Landay, J. The Integrated communication 2 Draw (IC2D) system, CHI 99 Proceedings, May 1999, Pittsburgh, PA, pp. 222-223.

[10] Kamel, H., Landay, J. A Study of Blind Drawing Practice: Creating Graphical Information Without the Visual Channel, ASSETS 2000 Proceedings, November 2000, Washington, DC, pp. 34-41.

[11] Kennel, A. AudioGraf: A diagram reader for blind people, ASSETS 96 Proceedings, April 1996, Vancouver, pp. 51-56.

[12] Mansur, D.L., Blattner, M.M., and Joy, K.I. Sound graphs: a numerical data analysis method for the blind, Journal of Medical Systems, 9(3), 1985, pp. 163-174.

[13] Meijer, P.B.L., "An Experimental System for Auditory Image Representations," IEEE Transactions on Biomedical Engineering, Vol. 39, No. 2, pp. 112-121, Feb 1992.

[14] Mynatt, Elizabeth and Gerhard Weber. Nonvisual Presentation of Graphical User Interfaces: Contrasting Two Approaches, CHI 94 Proceedings, April 1994, Boston, MA, pp. 166-172.

[15] Petrucci, L. S., Harth, E., Roth, P., Assimacopoulos, A., Pun, T. WebSound: a generic Web sonification tool, and its application to an auditory Web browser for blind and visually impaired users, ICAD 2000 Proceedings, April 2000, April 2000, Atlanta, GA, pp. 6-9.

[16] Roth, P., Petrucci, L. S., Pun, T. From Dots to Shapes: an auditory Haptic game platform for teaching geometry to blind pupils, ICCHP 2000 Proceedings, July 2000, Karlsruhe, pp. 603-610.

[17] Shimizu, M., Itoh, K., Nakazawa, T., Pattern Representation System using Movement Sense of Localized Sound, HCI 99 Proceedings, August 1999, Munich, pp. 990-994.

[18] Wenzel E M et al. Localization using nonindividualized head-related transfer functions. Journal of the Acoustical Society of America, 94, 1993, (1) 111-123.