EFFICIENCY OF SPEARCON-ENHANCED NAVIGATION OF ONE DIMENSIONAL ELECTRONIC MENUS

Dianne K. Palladino and Bruce N. Walker

Sonification Lab, School of Psychology
Georgia Institute of Technology
Atlanta, GA 30332
dianne.palladino@gatech.edu, bruce.walker@psych.gatech.edu

ABSTRACT

This study investigated navigation through a cell phone menu in the presence of auditory cues (text-to-speech and spearcons), visual cues, or both. A total of 127 undergraduates navigated through a 50-item alphabetically listed menu to find a target name. Participants using visual cues (either alone or combined with auditory cues) responded faster than those using only auditory cues. Performance was not found to be significantly different among the two auditory only conditions. Although not significant, when combined with visual cues, spearcons improved navigational efficiency more than both text-to-speech cues and menus using no sound, and provided evidence for the ability of sound to enhance visual menus. Research results provide evidence applicable to efficient auditory menu creation.

1. INTRODUCTION

Various types of auditory displays have been studied as either enhancements or primary means of display for the menus on small electronic devices, such as cell phones and personal digital assistants (PDAs). Auditory displays on mobile devices are of potential benefit to all users, but the visually impaired may stand the most to gain because of the potential for auditory displays to make the latest technology more readily accessible [1]. Since most devices are currently designed for use via a purely visual interface (i.e., various types of visual menus), it is important to determine the auditory enhancements that may lead to more effective use of these menus by individuals for whom a primarily visual interface is impractical.

Four primary auditory menu cues have been previously suggested as feasible: regular speech, auditory icons [2], earcons [3], and most recently, spearcons [4, 5]. All of these auditory menu cues have their advantages and limitations, so continuing research attempts to determine the optimum auditory enhancement in terms of efficiency, learning rates, and usability [4-7].

1.1. Auditory Menus

Auditory menus facilitate navigation of available functionality on electronic interfaces by using sound [7]. Using sound to enhance menus on electronic systems, whether small electronic devices or desktop systems, widens potential uses for the devices, and increases the number of potential users. In its simplest form, an auditory menu typically consists of electronic Text-To-Speech (TTS) conversion of the words or phrases included in the menu hierarchy. Users of auditory menus typically navigate the menu using arrow keys provided on the device, and menu items are presented using sound. Sound alone or sound combined with visual menu cues can be used to assist the user with navigation through the device's functions. In most cases, when the user lands on the desired item, a button such as the "enter" key on the device or keyboard is used to select the item.

Auditory enhancements to a menu are sometimes prepended with cues to assist in efficient navigation. Since speech alone is relatively slow and inefficient, the goal of these cues is to provide faster recognition of the menu item in question and to improve navigational efficiency. It is possible for the auditory cue (or a portion of the cue) to be sufficient information for the user to determine if the current location on the menu is the desired destination or if it is necessary to navigate further. The unaltered TTS of the menu item can be (but does not necessarily need to be) included after the cue, so that if the users have any confusion about the meaning of the cue they can listen to the entire spoken word or phrase to verify menu location. It is possible that with moderate usage of the auditory cues, the original TTS phrase will be used less frequently, and the option to remove the TTS phrases completely and utilize solely the cues to navigate the auditory menu is a potential option for users. If the auditory cues take less time to perceive than the original TTS phrases, then once the TTS is no longer needed navigation should become more efficient for the user.

The transient nature of sound causes several unique usability challenges for designers of auditory menus. The first is the differences in speech comprehension speed among individuals. There is limited information available on this topic, but one study found that blind listeners can understand speech at up to 2.8 times faster than the standard rate of TTS [8]. These differences in range challenge designers to create renditions that will be at a comfortable and understandable speed for most users. A second challenge is location awareness. Users need to know their current position in an auditory menu and be able to discern the fastest path to reach another position in the menu [9]. Unlike a visual menu, which can be scanned quickly to determine the current position relative to the hierarchy of the menu, an auditory menu can require a considerable amount of the user's working memory to maintain the same information. The third challenge to auditory menu design is enabling the user to learn the auditory cues quickly. A shorter learning curve will allow the user to begin taking advantage of the functionality of the phone in the shortest amount of time possible.

Evidence for the most feasible auditory menu enhancement cue type has been provided by two previous experiments. Walker, Nance, and Lindsay [4] found that spearcons outperform auditory icons, earcons, and speech alone in time to target efficiency. Palladino and Walker [5] compared rates of learning associations between earcons and spearcons and the items that they represent,

and found that earcons were significantly more difficult and frustrating to learn than spearcons. The current study collected evidence about the usability of only spearcon enhancements as compared to TTS alone. Note that the present study did not include auditory icons as cues, because the lack of natural sounds available to represent menu items in mobile devices often makes auditory icons less effective in practical applications.

1.2. Auditory Icons and Earcons

Although auditory icons [2] and earcons [3] are not empirically investigated in this experiment, a brief explanation of their composition and their advantages and disadvantages is worthwhile. Both have been proposed in the past as solutions to auditory menu challenges but have disadvantages that have been at least partially overcome by spearcons.

An auditory icon is a direct or metaphorical representation of the natural sound produced by an item [2]. From infancy we learn that cows "moo" and that cats "meow," and there are a large number of items for which we have a natural automatic association between the sound and the item. For certain words, such as animals, musical instruments, and people sounds, a direct connection between the sound and the word is obvious to most people.

A challenge arises when designers attempt to use auditory icons to represent actions or objects that do not produce natural sounds. For example, what would be the auditory icon for "Save to Desktop" or "Options" on a typical electronic interface? The more metaphorical an auditory icon becomes, the longer it may take for a user to learn an association between the representation and the item, even though once the association is learned little difference is seen in performance [2]. There have been somewhat successful attempts to create auditory icons for some computer-related functions, as illustrated in the sound associated with Microsoft's Recycle Bin. Although this is not a natural sound, it seems somewhat logical. Most people agree that the sound is like a crumpled up piece of paper being thrown into a metal waste paper basket. What happens when a computer user tries a Mac, however? The sound for the Trash icon on the Mac interface defaults to a completely different sound. If the item represented does not make a natural sound, it is difficult to reach a consensus because the auditory icon needs to become more metaphorical [6]. It then is less useful due to conflicting opinions of the most appropriate auditory representation for the item. This lack of ecological validity to most electronic menu items makes an auditory icon an undesirable option for creating electronic menu enhancements.

Earcons [3] are systematically produced representations of menu items using musical elements and can be created by varying frequency, timbre, tempos, rhythmic patterns, or combinations of any aspect of music to represent unique items on a menu. Guidelines suggested by Hereford and Winn [10] suggest that earcons are most effective when each item represented in a group differs in as many musical elements as possible from the other members of the group. Earcons can be created to represent a hierarchy of items in a menu system by combining musical elements systematically [9, 11, 12].

To create a 5-row by 5-column hierarchical menu system, a designer might consider using a different timbre of sound (piano, trumpet, flute) to represent every item in each column, and a different overlying rhythmic pattern (two quarter notes on snare drum, eighth notes on a cowbell, triplets on a wood block) to represent each row. An item on the menu grid would be represented by the simultaneous play of the two musical elements of the row and column for that particular grid position. Once the user has memorized the order of each musical element for each row and column, it can be an effective way for users to determine their position in a particular menu hierarchy, and participants in prior studies have had success in identifying and understanding this hierarchical information [11, 13]. In 2003, Vargus and Anderson [14] combined earcons with speech to find that the combination increased efficiency of menu navigation without additional burden on the user.

Advantages of earcons include their usefulness in providing hierarchical menu information and their ability (unlike auditory icons) to be applied to menus containing any type of information. Earcon hierarchy can be a disadvantage, however, because the rigid nature of the menu setup makes it difficult to add or subtract an item within the hierarchy. For example, if an item is added to the fourth column, second row of the grid, it is debatable whether it would make more sense to move everything else in that column down a row and change its earcon representation or to create an entirely new row and leave that row blank in the other columns. It is not clear which (if any) of these two solutions would be the most effective. As Walker et al. [4] have stated, the arbitrary nature of the earcon is considered both its strength and its weakness. Additionally, Palladino and Walker [5] found that it is difficult for users to learn earcon/word associations, and this difficulty can cause frustration for the user. Auditory enhancement cues are intended to decrease user frustration and annoyance [15] as well as to increase navigation efficiency, but earcons seem to fall short on these criteria [4, 5]. For this reason, earcons are not considered in this study as possibilities for auditory cues.

1.3. Spearcons

A spearcon [4] is created by compressing a spoken phrase (created either by a TTS generator or by recorded voice) without modifying the perceived pitch of the sound. Some speech is compressed to the point that it is no longer comprehensible as a particular word or phrase. Walker et al [4] compared the spearcon to a fingerprint because each unique word or phrase creates a unique sound when compressed that distinguishes it from other spearcons. After a brief learning session, the associations between a spearcons and their related words or phrases are easy to recognize [5].

In order to create spearcons for use as auditory menu cues, a sound file containing the speech must first be created by using TTS generation software or by simply recording a voice speaking the words or phrases. The spearcon is created from that file, and prepended to the original TTS file in the form of a "cue." A small duration (250 ms) of silence is inserted between the spearcon cue and the original word or phrase. More information on spearcon creation is provided in the methods section of this document.

Spearcons are naturally briefer than the words and phrases they represent, are fast and easy to produce, and can be easily inserted into any menu structure in any position because they are direct representations and do not depend upon hierarchical positioning in a menu. Although spearcons do not provide natural hierarchical information

to the user, such as those that are inherent in hierarchical earcons [3], it would be possible to create hierarchical information for the user by implementing some sort of augmentation to the spearcons, such as adding volume cues or pitch cues to provide position information to the user. This addition may not be absolutely necessary for efficiency of navigation, however, as shown by Walker et al [4], who found that spearcons resulted in significantly more efficient navigation than hierarchical earcons, even when using spearcons with no hierarchical information.

Palladino and Walker [5] found spearcons to be significantly easier to learn than earcons when users were trained on associations with the words and phrases they represented. Half of the participants were trained and tested on spearcon associations, and the other half of participants were tested on earcon associations. Participants found spearcon/word associations easier to learn and the learning process for earcon associations more arduous and frustrating. With these advantages for spearcons over other enhancement types, the focus for auditory menu enhancement research has narrowed to comparing the benefits of using spearcons as prepended cues to TTS to using TTS alone in an auditory menu system. This comparison is the focus of the current study.

This experiment included conditions with visual menu cues, either alone or in combination with one of the auditory representations. For an individual with normal vision, the conditions with visual cues are expected to enhance the speed to the target menu item. Visual cues, however, may not be useful to visually impaired individuals, and this experiment will focus more on the length of an auditory stimulus and its effect on the time it takes to reach a requested target item on a menu. It is of interest, however, to compare the visual and auditory stimuli to have a basis of comparison for future planned studies with visually impaired individuals.

This experiment compared navigation rates of a simulated cell phone contact book created with various combinations of visual and auditory elements. It compared auditory cues created with TTS only, TTS with a spearcon enhancement cue, and no audio at all. Each auditory condition also was tested combined with a visual menu. The hypothesis of this study was that conditions with visual menus will outperform those with only auditory cues, and that spearcon enhancement prepended to the TTS will significantly outperform the other auditory conditions.

2. METHOD

2.1. Participants

A total of 127 undergraduates (55 men and 72 women, mean age = 19.74) with normal or corrected-to-normal hearing and vision participated for extra credit in psychology courses. English was the native language of all participants. There were either 25 or 26 participants in each condition.

2.2. Design

This experiment used a between-subjects design. The first independent variable was sonification type (TTS Only, Spearcon Cue + TTS, or No Audio), and the second independent variable was visual cue (On or Off). The condition in which auditory and visual cues are simultaneously off obviously is not a valid condition, which leaves five appropriate experimental conditions. The dependent variable was average time to selection of target menu item.

2.3. Materials

Participants were tested with a computer program written with Macromedia Director MX and Lingo on a Windows XP platform listening through Sennheiser HD 202 headphones. They were given an opportunity at the beginning of the experiment to adjust volume for personal comfort.

A random name generator (http://www.xtra-rant.com/gennames/) created the 50 names used for the contact book stimuli (e.g., “Allegra Seidner”). Auditory TTS was generated for all of the names using the AT&T Labs, Inc. Text-To-Speech Demo program (http://www.research.att.com/~ttsweb/tts/demo.php).

Spearcons were created for the TTS conversion of each name by running them through a MATLAB algorithm that compressed each name logarithmically while maintaining original sound frequency. Logarithmic compression is currently considered the preferred compression technique for creating spearcons because it compresses longer phrases more than shorter phrases. Shorter words (particularly those that are monosyllabic) tend to sound more like “clicks” if they are compressed too much and become indistinguishable. Since they are very short to begin with, the advantage of compression of very short words is much less than for a longer phrase. Phrases of several words or syllables can be compressed at a much higher ratio since they contain a higher level of language context. Higher compression makes the spearcons shorter and more efficient without losing the context needed to identify them as unique.

Stimuli for the Spearcon Cue + TTS condition were created by using Audacity software to prepend the cue to the TTS with a 250 ms post-cue interval between them. Visual stimuli consisted of a list of names displayed to the participant in 30-point text. Names were displayed in alphabetical order by first name in a “window” ten at a time, and the list scrolled downward or upward based upon the key presses of the participant. For both the auditory and visual components, if the participant reached the top or bottom of the list, the list did not wrap around. Although this design does not simulate the exact functionality of the screen on a cell phone or PDA contact book menu, this feature is necessary to control for distance to the target name on the list. As the focus changed to each menu item, auditory and visual menu cues were presented simultaneously in conditions including both modes of display.

2.4. Procedure

A simulated cell phone contact book menu was presented that contained items constructed with auditory, visual, or both representations. The contact book consisted of 50 names (first and last) in alphabetical order by first name. The up and down arrow keys were used to navigate the menu, and the enter key was used to select the appropriate item. Participants were assigned to one of five conditions. Two conditions provided only auditory cues for each menu item: one with TTS cues and one with spearcons prepended to the TTS. The other three conditions all combined visual cues with sound: one with no auditory cues, one with TTS, and one with spearcon cues prepended to the TTS. In a given block of trials, half of the names were used as targets. The resulting two types of blocks were alternated five times for a total of 10 blocks of 25 trials each. All participants experienced the same procedure for each block, regardless of the assigned menu display condition. The order of appearance of the list halves was counterbalanced among subjects.

Participants first saw a brief instruction screen that taught them about menu navigation and that the required task was to find the requested target name on the menu as quickly as possible without sacrificing accuracy. The participant was then presented with a name (e.g. "Allegra Seidner") on the top of the screen that indicated the target name. When the first up or down key was pressed, the timer started. Participants navigated through the menu system to find the assigned target name and hit the "enter" key to indicate selection of the requested target. Hitting the enter key recorded the end time. Each participant immediately was shown the next target name, and the procedure was repeated for all 25 names in the block. Participants were then shown a screen that indicated that the next block of 25 trials was about to start. Each of the nine subsequent blocks proceeded in the exact same way. After the tenth block, participants filled out a brief demographics questionnaire regarding age, gender, ethnicity, and musical training information. A free-format opportunity was also provided to comment on their experience with the experiment and any strategies they may have used to complete the task.

3. RESULTS

An alpha level of .05 was used for all statistical analysis. After disqualifying 1.53 % of trials due to incorrect item selection (37 in Visuals Off/Spearcons+TTS condition, 16 in Visuals Off/TTS condition, 112 in Visuals On/No Sound condition, 96 in Visuals On/TTS condition, and 36 in Visuals On/Spearcons+TTS condition), a total of 31272 trial records remained with which to perform the data analysis. A one-way ANOVA was performed on the data to check for significant differences among the different experimental conditions. Results of this analysis are illustrated in Figure 1, which plots mean times to target for each condition in each block of the experiment. Not surprisingly, overall performance on all conditions including visual cues were significantly faster than those including only auditory cues, F (1, 31270) = 4963.665, $p < 0.001$.

As expected, the plotlines for the auditory-only conditions show consistently longer mean times to target throughout the blocks than the conditions that contained both visual and auditory cues. A Tukey honestly significant difference analysis of Block 10 data for each

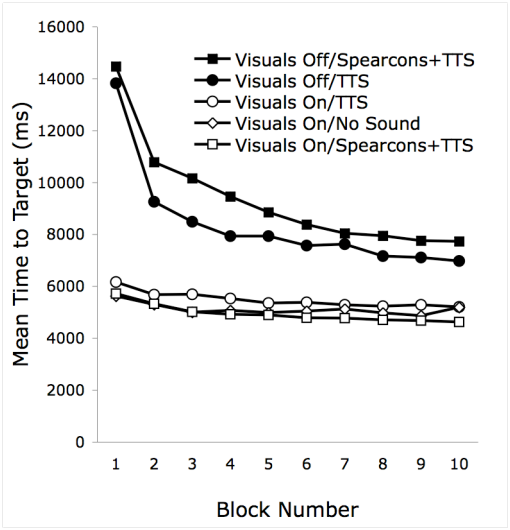

Figure 1. *Mean time to target in milliseconds for all conditions over all blocks. Learning effects were found for all conditions, and were most significant for the two conditions that did not use visual cues. The TTS condition outperformed the spearcon+TTS condition in auditory-only conditions, and spearcon+TTS conditions outperformed both of the conditions using visual cues consistently, although not significantly. The Visuals On/Spearcons+TTS condition outperformed the condition that did not use auditory cues, though not significantly. This may provide evidence that auditory cues potentially enhance the performance of menu navigation if used in conjunction with visual information.*

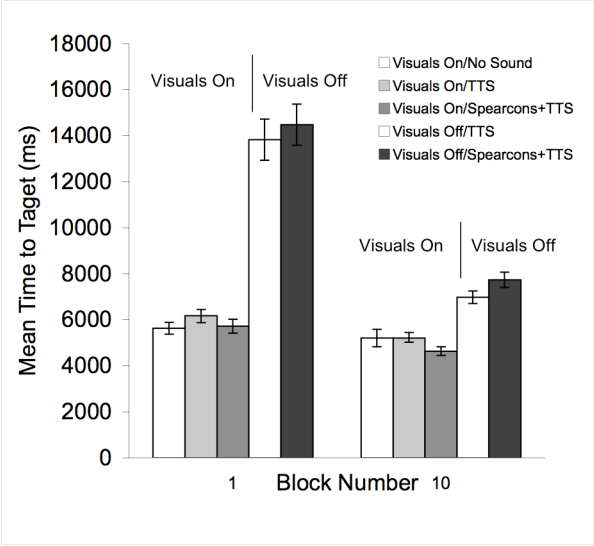

Figure 2. *Mean time to target in milliseconds for all conditions in Blocks 1 and 10. Graph shows the difference in performance between auditory only conditions and those including visual cues decreases by the last block of the experiment. This is evidence that solely auditory menu cues have the ability to approach the efficiency of menus with visual elements. Error bars show 95% Confidence Intervals.*

	Block 1			Block 10			
Condition	Mean	SD	N	Mean	SD	N	Δ (ms)
Visual On	5828	3535	1845	5014	3635	1842	814
Visual Off	14147	11552	1291	7350	3961	1297	6797
Spearcon	10279	9796	1239	6245	3826	1241	4034
TTS	10089	9408	1258	6111	3436	1264	3978
No Sound	5618	3326	639	5194	4909	634	424

Table 1. *Means, Standard Deviations, and Change of Time (ms) to Target Name for Blocks 1 and 10 Collapsed Over Visual and Sound Conditions*

condition found no significant difference between any of the three conditions including visual cues at the $p < 0.05$ level, although the differences in means between the Visuals On/TTS (M = 5201, SD = 3028) and Visuals On/Spearcons + TTS (M = 4627, SD = 2365) conditions in Block 10 showed a nearly significant contrast $p = 0.062$. It is clear from Figure 1, however, that even though the differences between the conditions using auditory-only and auditory and visual cues in Block 10 are significant, there is much less of a difference between the auditory-only and visual conditions than existed in the first block of the experiment.

Figure 2 illustrates the mean time to target for the five categories in the first and tenth blocks. There was a significant difference in the means collapsed over all conditions between the first (M = 9253, SD = 8890) and tenth (M = 5979, SD = 3944) blocks, $F(1, 6273) = 355.635$, $p < 0.001$, indicating learning across blocks.

Table 1 summarizes mean and standard deviation information comparing visual and auditory conditions and their performance improvements between the beginning and end of the trials. Comparison of the change in performance among the auditory cues between the first and tenth block revealed a main effect of sonification type, $F(2, 6269) = 86.113$, $p < 0.001$ with an interaction of sonification type and block number, $F(2, 6269) = 35.761$, $p < 0.001$ indicating a more significant improvement from Block 1 to Block 10 for the spearcon conditions than for the No Sound condition. Post-hoc analysis indicated that Spearcons + TTS and TTS conditions did not show significantly different performance improvements.

Comparing the conditions with visual cues to those without visual cues revealed a main effect of visual cue, $F(1, 6271) = 1128.36$, $p < 0.001$ between the first and tenth block with the non-visual conditions facilitating a larger improvement in performance by the end of the experiment, as indicated with a significant interaction between visual cue condition and block, $F(1, 6271) = 355.75$, $p < 0.001$.

4. DISCUSSION

The results confirm that the conditions including visual cues lead to faster performance overall when compared to conditions with only auditory cues. Expectations that the spearcon cues would outperform TTS-only cues, regardless of visual cues absence or presence, were not corroborated. Nevertheless, performance in conditions using spearcons *was* consistently, but not significantly, faster when auditory cues were combined with visual cues. This finding provides evidence that spearcons combined may enhance performance when navigating auditory-enhanced visual menus. More research should be conducted to determine if this effect persists.

There are potential reasons why, in the case of the auditory-only conditions, spearcons were not found to lead to shorter navigation times. Since the spearcons were presented as cues prepended to the TTS phrase, some participants may have felt compelled to listen through the spearcon and the silent interval to hear the TTS phrase, rather than concentrating on the spearcon itself. This may have certainly increased time to target in the spearcon+TTS condition. It would be interesting to run the study again without the convenience of the TTS phrase inclusion. Another option would be to scramble the names on the phone between each trial, rather than leaving the names in alphabetical order the entire time. This setup would test the spearcon enhancements more purely, but would not be an accurate replication of the real-life setup of such menus since they do not scramble in practice. Replicating this study in this fashion may nevertheless provide useful information for non-alphabetical menus. Also, perhaps including a training session before starting the experiment on the associations between the words and the sounds would decrease the impulse to wait for the TTS as well. These considerations should be tested in future studies.

There was a strong learning curve in the auditory-only cue conditions, but after 10 blocks the performance in these conditions had improved to a point that remained significantly different from that in the conditions that used both auditory and visual cues combined. Figure 2 shows a compelling picture, however, that reveals the level of performance to be much more level for all five conditions than in the first block of the experiment. One would probably expect performance on a strictly auditory menu to be worse than on one including visual cues for a person without a visual impairment. The fact that performance improved to such a degree for individuals accustomed to a visual world lends interest to a replication of this study with visually impaired individuals; that study is in preparation. The replication will provide an even more complete picture of navigational performance in different contexts. Our next studies also include studying additional auditory enhancements, particularly spearcons usage on multi-dimensional menus and submenus, replication of this study on actual cell phone devices, and replication and focus groups with visually impaired and blind users.

In conclusion, utilizing auditory and multimodal menus and enhancements in small electronic devices is clearly feasible, and the electronics industry appears ready to take on the challenge of incorporating accessible technology into their interfaces, particularly for cell phone

menus. With strong empirical science backing up the feasibility of the spearcon, it is hoped that it will not be long before those with temporary and permanent visual disabilities will more easily be able to enjoy the productivity of electronic devices to the same extent as individuals with normal vision. From the viewpoint of both the manufacturers and the potential users, this research is expected to lead to positive advancements in accessible technology.

5. REFERENCES

[1] M. A. Nees and B. N. Walker, "Auditory interfaces and sonification," in *The Universal Access Handbook*, C. Stephanidis, Ed. Mahwah, NJ: Lawrence Erlbaum Associates, In Press, p. TBD.

[2] W. W. Gaver, "Auditory icons: Using sound in computer interfaces," *Human-Computer Interaction,* vol. 2, pp. 167-177, 1986.

[3] M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg, "Earcons and icons: Their structure and common design principles," *Human-Computer Interaction,* vol. 4, pp. 11-44, 1989.

[4] B. N. Walker, A. Nance, and J. Lindsay, "Spearcons: Speech-based earcons improve navigation performance in auditory menus," in *International Conference on Auditory Display*, London, U.K., 2006.

[5] D. K. Palladino and B. N. Walker, "Learning rates for auditory menus enhanced with spearcons versus earcons," in *International Conference on Auditory Display*, Montreal, Canada, 2007.

[6] B. N. Walker and G. Kramer, "Ecological psychoacoustics and audtory displays: Hearing, grouping, and meaning masking," in *Ecological Psychoacoustics*, J. G. Neuhoff, Ed. New York: Academic Press, 2004, pp. 150-175.

[7] P. Yalla and B. N. Walker, "Advanced Auditory Menus," Georgia Institute of Technology GVU Center GIT-GVU-07-12., October 2007.

[8] C. Asakawa, H. Takagi, S. Ino, and T. Ifukube, "Maximum listening speeds for the blind," in *International Conference on Auditory Display*, Boston, MA, 2003.

[9] G. Leplatre and S. Brewster, "Designing non-speech sounds to support navigation in mobile phone menus," in *International Conference for Auditory Display*, Atlanta, GA, 2000, pp. 190-199.

[10] J. Hereford and W. Winn, "Non-speech sound in human-computer interaction: A review and design guidelines," *Journal of Educational Computing Research,* vol. 11, pp. 211-233, 1994.

[11] S. Brewster, V. Raty, and A. Kortekangas, "Earcons as a method of providing navigational cues in a menu hierarchy," in *HCI'96 Conference*, Imperial College, London, UK, 1996.

[12] S. Brewster, P. C. Wright, and A. D. N. Edwards, "An evaluation of earcons for use in auditory human-computer interfaces," in *SIGCHI Conference on Human Factors in Computing Systems*, Amsterdam, 1993.

[13] G. Leplatre and S. Brewster, "An Investigation of Using Music to Provide Navigation Cues," in *International Conference for Auditory Display*, Glasgow, UK, 1998.

[14] M. L. M. Vargas and S. Anderson, "Combining speech and earcons to assist menu navigation," in *International Conference for Auditory Display*, Boston, MA, 2003.

[15] S. Brewster and M. G. Crease, "Correcting menu usability problems with sound," *Behaviour & Information Technology,* vol. 18, pp. 165-177, 05 1999.