

Navigation Efficiency of Two Dimensional Auditory Menus Using Spearcon Enhancements

Dianne K. Palladino and Bruce N. Walker

Sonification Lab, School of Psychology

Georgia Institute of Technology

Atlanta, GA 30332-0170

dianne.palladino@gatech.edu, bruce.walker@psych.gatech.edu

A total of 28 undergraduates navigated to specific items in a two dimensional menu that was displayed using only sound. The auditory menu consisted of either text-to-speech (TTS) only, or TTS enhanced with spearcons. Spearcons are brief sound cues created by compressing the original TTS sound file. Speed of navigation to target items in the auditory menu was found to be significantly faster in the spearcon condition than in the condition using only TTS. There was also a smaller per-item cost in terms of speed for the spearcon-enhanced menu, leading to increasingly better performance as menu length increased. These results provide further evidence that spearcon enhancements can lead to faster navigational performance in auditory menus, when compared to text-to-speech alone.

INTRODUCTION

Auditory displays are representations of information often using sounds (Walker & Kramer, 2004). Various types of auditory displays, using both speech and non-speech sounds, have been studied as either enhancements to, or primary display modalities for, the navigation systems of small electronic devices such as cell phones and personal digital assistants (PDAs) (Brewster, Leplatre, & Crease, 1998; Brewster, Raty, & Kortekangas, 1996; Gaver, 1986; Leplatre & Brewster, 2000; Palladino & Walker, 2007, 2008; Vargas & Anderson, 2003; Walker & Kramer, 2004; Walker, Nance, & Lindsay, 2006). Auditory displays are of benefit to people with persistent vision impairments (i.e., blind users of a computer system), as well as those with limited visibility due to environmental conditions or circumstances. For example, auditory menus on electronic devices can allow firefighters to access important information in a smoke-filled room. Since users interact with most current devices through the use of various types of visual-only menus, it is important to determine the auditory equivalents to interface navigation structures that will allow the greatest level of efficiency, ease of learning, and dissemination of navigational information.

There are four primary auditory menu cues that previously have been suggested as feasible: regular speech, auditory icons (Gaver, 1986), earcons (Blattner, Sumikawa, & Greenberg, 1989), and most recently spearcons (Palladino & Walker, 2007, 2008; Walker et al., 2006). Auditory icons are representations of the sound naturally produced by or associated with a menu item, and earcons are hierarchical representations of menu items created using musical elements. Spearcons are a form of condensed speech with a unique, fingerprint-like acoustical representation. All of these auditory menu cues have advantages and limitations; research continues to search for the optimum auditory enhancement (Palladino & Walker, 2007, 2008; Walker & Kramer, 2004; Walker et al., 2006; Yalla & Walker, 2007).

Auditory Menus

Auditory menus allow navigation of functionality on electronic interfaces by using sound (Yalla & Walker, 2007). 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. 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 sometimes include sounds that are prepended to the TTS portion of a menu item, 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 create a greater 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 often is (but does not necessarily need to be) included after the cue, so that if the user has any confusion about the meaning of the cue, the entire word or phrase can be heard 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 themselves 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 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 (Asakawa, Takagi, Ino, & Ifukube, 2003). 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 to be able to discern the fastest path to reach another position in the menu (Leplatre & Brewster, 2000). 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 the ability for the user to learn the auditory cues quickly. A shorter learning curve will decrease the time it takes for the user to use the phone's functionality.

Evidence for the most feasible auditory menu enhancement cue type has been provided by two previous experiments. Recently, Walker, Nance, and Lindsay (2006) found that spearcons outperform auditory icons, earcons, and speech alone in time to target efficiency. Palladino and Walker (2007) 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. Spearcons were found to be the most favorable cue type with respect to efficiency and learning rate. The present study collected more evidence about the usability of the spearcon.

Auditory Icons and Earcons

Although auditory icons (Gaver, 1986) and earcons (Blattner et al., 1989) 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 the spearcon.

An auditory icon is a representation of the natural sound produced by an item (Gaver, 1986). 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 problem arises when designers attempt to use auditory icons to represent actions or objects that are intangible. For example, what would be the auditory icon for "Save to Desktop" or "Options" on a typical electronic interface? 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 (Gaver, 1986; Walker & Kramer, 2004). 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 electronic menu enhancements.

Earcons (Blattner et al., 1989) 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 (1994) 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 (Brewster et al., 1996; Leplatre & Brewster, 1998).

To create a 5-row by 5-column 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.

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. As Walker et al. (2006) have stated, the arbitrary nature of the earcon is considered both its strength and its weakness. Additionally, Palladino and Walker (2007) 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 (Palladino & Walker, 2007; Walker et al., 2006) as well as to increase navigation efficiency, but earcons seem to fall short on these criteria (Walker et al., 2006).

Spearcons

A spearcon (Palladino & Walker, 2007; Walker et al., 2006) 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 (2006) likened the spearcon to a fingerprint because each word or phrase creates a unique sound when compressed that distinguishes it from other spearcons.

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 often 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 easily can be 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 (Palladino & Walker, 2007), 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. (2006), who found that spearcons resulted in significantly more efficient navigation than hierarchical earcons, even when using spearcons with no hierarchical information.

In a study performed by Palladino and Walker (2007), spearcons were found to be significantly easier and less frustrating to learn than earcons when users were trained on associations with the words and phrases they represented. With these advantages for spearcons over other enhancement types, the focus for auditory menu enhancement research has narrowed to comparing the specific benefits of using spearcons as prepended cues to TTS phrases in auditory menu systems. The first study to compare such differences used a one dimensional 1x50 auditory menu such as what would be found on a typical mobile phone contact book. Palladino and Walker (2008) found that spearcon enhancements led to performance that was at least as efficient as menu items created with only TTS.

The present experiment was designed to determine if navigation efficiency would increase when using prepended spearcons on two dimensional menus. It compares navigation rates in a simulated two dimensional menu with auditory menus constructed with only TTS to those using prepended spearcon auditory enhancements. This study’s hypothesis was that speed of navigation would be faster when the menu items were prepended with spearcons than when using only TTS.

METHOD

Participants

A total of 28 undergraduate participants (9 male, 19 female, mean age = 19.14) with normal or corrected to normal hearing and vision participated for extra credit in psychology courses. English was the native language of all participants.

Design

This experiment used a between-subjects design with two conditions. The independent variable was sonification type (TTS Only, Spearcon Cue + TTS) and the dependent variable was average time in milliseconds to selection of requested target item. There were 14 participants in each condition.

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 to adjust volume for personal comfort before commencing the experiment.

An auditory menu structure was created using menu items included in the menus of a Nokia N91 mobile phone. This structure consisted of six menu categories (Messaging, Music, Connectivity, Tools, Camera, Gallery), and from 5 to 9 items that were associated with each category. This created an irregular menu structure similar to what a user would encounter using any hierarchical menu structure on a mobile phone or computer operating system. Table 1 illustrates the items included in the menu structure.

Auditory TTS was generated for all 44 items using the AT&T Labs, Inc. Text-To-Speech Demo program (<http://www.research.att.com/~ttsweb/tts/demo.php>). TTS files were converted to spearcons using a MATLAB algorithm that compressed each word or phrase while maintaining sound frequency. This procedure ensured a consistent perception of pitch. The algorithm uses a logarithmic compression method that compresses long words or phrases more than it does short ones. Doing this eliminates the transformation of very small monosyllabic words into “clicks” that lose their resemblance to the original TTS. Phrases or words with many syllables are compressed at a much higher ratio because they contain more contextual linguistic information in their original form. This allows more of this context to be maintained when compressing at a higher ratio.

For the Spearcon+TTS condition, the spearcons were prepended to the TTS with approximately 250 ms between the two sounds. No visual tools needed to be created for this experiment except for the screens that provided instructions to, and collected information from, the participants.

Procedure

Participants were presented with ten blocks of 22 trials each. Two stimulus lists were created from the original 44 items, and each list was alternated throughout the ten blocks. The lists were also randomized before each block. Using this procedure, each participant was tested on each menu item five times during the course of the experiment, for a total of 220 trials per participant. The order of presentation of the list halves was counterbalanced among subjects.

After a brief explanation of the auditory menus by the experimenter, the participants were shown an instruction

	Messaging	Music	Connectivity	Tools	Camera	Gallery
1	New Message	All Songs	Bluetooth	File Manager	New Image	Images
2	Inbox	Playlists	Data Cable	Application Manager	Delete	Video Clips
3	My Folders	Artists	Sync	Data Transfer	Send	Tracks
4	Mailbox	Albums	Device Manager	Profiles	Set as Wallpaper	Sound Clips
5	Drafts	Genres	Connectivity Manager	Settings	Add to Contact	Streaming Links
6	Sent	Composers		Themes	Rename Image	Presentations
7	Outbox	Options			Go to Gallery	All Files
8	Reports				Settings	Help
9	Options				Help	

Table 1. Visual representation of the auditory menu navigated by participants in each condition. The left column shows the level number corresponding to each item in the row of the menu.

screen that explained how to navigate the auditory menu using the keyboard. They were instructed that their task was to find the item requested as quickly as possible without sacrificing accuracy. For each trial, the participant was presented with an item to find by navigating through the menu using the arrow keys on the keyboard (e.g., “Find Genres on the Music Menu”). Participants then navigated through the menu system to find the requested item. Upon the first key press of each trial, the timer would start, and the timer ended when the participant hit the “enter” key. Time to target was found by calculating the difference between each start and end time. Each new trial in a block was followed immediately by the next trial, but the participants could control the start of each new block. After completing the tenth block of trials, participants filled out a brief questionnaire and were debriefed.

RESULTS

Error trials, arising from incorrect item selection, were removed from analyses; this meant 1.10% of trials (26 in Spearcons condition, 42 in TTS only condition) were eliminated. One outlier was also eliminated due to an extreme time to target. Upon further analysis of the path taken on this one trial, it was determined that the participant navigated the entire grid on the first trial to get a feel for the menu structure. Since this was clearly not the expected task, the trial was eliminated from consideration. After these eliminations, data from 6092 trials remained with which to perform statistical analysis. An alpha level of .05 was used for all statistical analysis.

A t-test on the mean time to target for each of the two conditions revealed that performance by participants was significantly faster in the spearcons condition ($M = 3821, SD = 3917$) than for those in the TTS only condition ($M = 5344, SD = 3918$) $t(6089) = 17.89, p < .001$. Figure 1 shows the mean time to target for each condition with 95% confidence intervals shown as error bars.

Due to the significant difference in navigation time between the two conditions, further analysis was performed based upon the level of the item on the menu. The number on the left hand side of the menu structure on Table 1 shows the number associated with each level of the menu structure. The menu structure used a maximum of nine levels of depth, and every menu category had at least five levels. For levels six through nine, the number of menu categories having each of

the levels decreased until level nine, in which case only two menu categories had an item on that level.

Regression lines created using the mean times to target by level for both conditions revealed that the navigation time was faster for every level of the auditory menu in the spearcons condition (see Figure 2). The slope of the TTS condition (slope = 518.59) was significantly steeper than for the spearcons condition (slope = 338.84) ($z = 5.064, p < .05$).

DISCUSSION

Palladino and Walker (2008) found no significant difference in performance for TTS-only and spearcon-enhanced one-dimensional menus. In contrast, in the present study with two-dimensional menus spearcon enhancements were found to improve navigation speed significantly when compared to auditory menu items created solely with TTS. In addition to faster performance across the board, the significantly flatter increase in average time to target as the level down a menu category increased indicates a lower per-item cost in navigation time in auditory menus using spearcon enhancements.

There are several possible reasons for this discrepancy. In the one dimensional menu, it was much more efficient for users to hold down the arrow key to get to an item that was alphabetically at the end of the list of contacts. Once they were used to how far they had to go they could forego listening to each item and estimate how long it would take to get to that part of the menu. This strategy potentially resulted in the auditory nature of each item between the top of the contact book and the target item to be sometimes negligible, because the user could “slide” past it without listening to it. In the two dimensional menu, the user had to take more care in navigating the menu. The menu structure consisted of two possible lines of navigation, and even the longest path to the target item was much shorter than in the one dimensional menu for the names alphabetically toward the end of the list.

The data in this study support the conclusion that using spearcon enhancements can lead to faster navigation of two dimensional auditory menus. The lower cost per navigational unit also suggests that spearcon enhancements increase efficiency on two dimensional menus at an even greater rate as the level of menu increases down a category. Future research is planned to determine if there is a limit to the size of a two-dimensional menu on which the spearcon

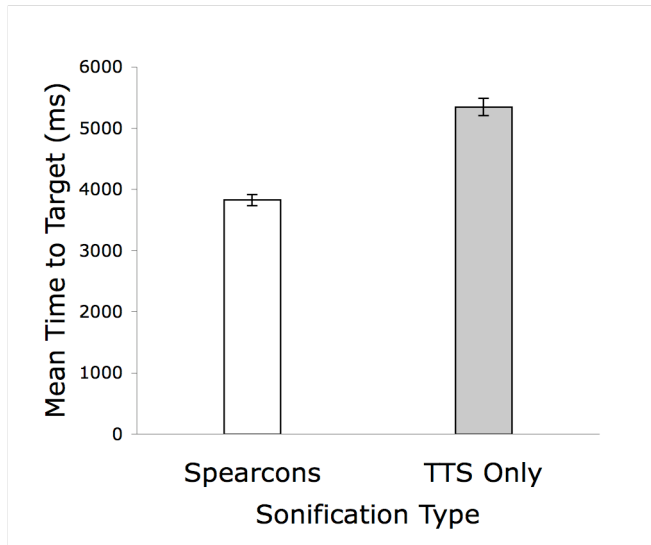


Figure 1. Mean time to target (ms) for navigating auditory menus with text-to-speech only versus TTS menu items with spearcon enhancements. Participants in spearcons condition performed significantly better than those in the text-to-speech condition.

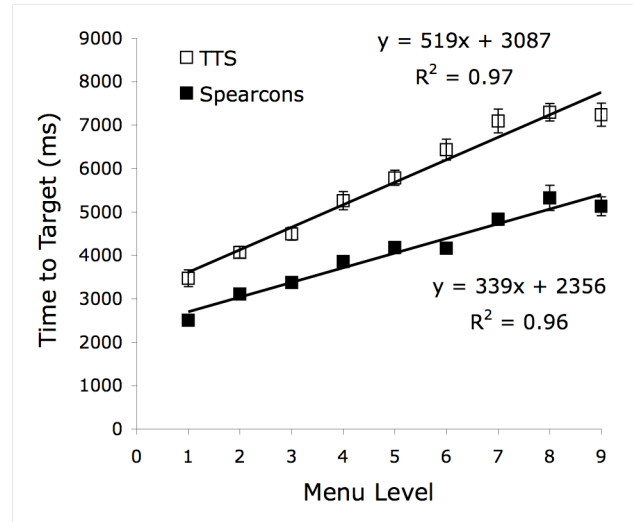


Figure 2. Mean time to target (ms) as a function of menu level. Spearcons led to faster performance at all menu depths, and there was a lower per-item cost for spearcons-enhanced items as depth in the menu increased.

enhancement results in such improvements in navigational speed.

Other planned research includes replicating this study and the one-dimensional spearcons study on a mobile phone and to replicate the spearcon studies using participants who are visually impaired or blind. These studies will provide data from the demographic of participant who is likely to benefit the most from enhanced navigational efficiency on auditory menus.

Incorporating accessible technology into small electronic devices is essential to allowing individuals with visual disabilities more opportunities to enjoy ever-advancing technology. It is equally important to make every effort to make these accessible devices as efficient and usable for all potential end users based upon their own requirements and abilities. Investigating technologies like the spearcon is one step on the path to increasing the numbers of people who can take advantage of technological advancements with ease.

REFERENCES

Asakawa, C., Takagi, H., Ino, S., & Ifukube, T. (2003). *Maximum listening speeds for the blind*. Proceedings of the International Conference on Auditory Display, Boston, MA.

Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and icons: Their structure and common design principles. *Human-Computer Interaction*, 4(1), 11-44.

Brewster, S., Leplatre, G., & Crease, M. G. (1998). *Using non-speech sounds in mobile computing devices*. Proceedings of the First Workshop

on Human Computer Interaction with Mobile Devices, Glasgow, UK.

Brewster, S., Raty, V., & Kortekangas, A. (1996). *Earcons as a method of providing navigational cues in a menu hierarchy*. Paper presented at the HCI'96 Conference, Imperial College, London, UK.

Gaver, W. W. (1986). Auditory icons: Using sound in computer interfaces. *Human-Computer Interaction*, 2(2), 167-177.

Hereford, J., & Winn, W. (1994). Non-speech sound in human-computer interaction: A review and design guidelines. *Journal of Educational Computing Research*, 11(3), 211-233.

Leplatre, G., & Brewster, S. (1998). *An Investigation of Using Music to Provide Navigation Cues*. Proceedings of the International Conference for Auditory Display, Glasgow, UK.

Leplatre, G., & Brewster, S. (2000). *Designing non-speech sounds to support navigation in mobile phone menus*. Proceedings of the International Conference for Auditory Display, Atlanta, GA.

Palladino, D. K., & Walker, B. N. (2007). *Learning rates for auditory menus enhanced with spearcons versus earcons*. Proceedings of the International Conference on Auditory Display, Montreal, Canada.

Palladino, D. K., & Walker, B. N. (2008). *Efficiency of spearcon-enhanced navigation of one-dimensional electronic menus*. Proceedings of the International Conference on Auditory Display, Paris, France.

Vargas, M. L. M., & Anderson, S. (2003). *Combining speech and earcons to assist menu navigation*. Proceedings of the International Conference for Auditory Display, Boston, MA.

Walker, B. N., & Kramer, G. (2004). Ecological psychoacoustics and auditory displays: Hearing, grouping, and meaning masking. In J. G. Neuhoff (Ed.), *Ecological Psychoacoustics* (pp. 150-175). New York: Academic Press.

Walker, B. N., Nance, A., & Lindsay, J. (2006). *Spearcons: Speech-based earcons improve navigation performance in auditory menus*. Proceedings of the International Conference on Auditory Display, London, U.K.

Yalla, P., & Walker, B. N. (2007). *Advanced Auditory Menus* (No. GIT-GVU-07-12.). Georgia Institute of Technology GVU Center.