Spearcons Improve Navigation Performance and Perceived Speediness in Korean Auditory Menus

Hyewon Suh[1], Myounghoon Jeon[2] and Bruce N. Walker[12]
Sonification Lab, School of Interactive Computing[1] & School of Psychology[2]
Georgia Institue of Technology
Atlanta, GA, 30332 USA

For decades, auditory menus using both speech (usually text-to-speech, TTS) and non-speech sounds have been extensively studied. Researchers have developed situation-optimized auditory menus involving such cues as auditory icons, earcons, spearcons, and spindex. Spearcons have generally outperformed other cues in terms of providing both contextual information and item-specific information. However, little research has been devoted to exploration of spearcons in languages other than English, or the use of spearcon-only auditory menus. In this study, we evaluated the use of spearcons in Korean menus, as well as the use of spearcons alone. Twenty-five native Korean speakers navigated through a two-dimensional auditory menu presented via TTS, with or without spearcon enhancements. Korean spearcons were successful. Participants also rated the spearcon-enhanced menu as seeming speedier and more fun than the TTS-only menu. After a short learning period, mean time-to-target in the auditory menu was even faster with spearcons alone, compared to traditional TTS-only menus.

INTRODUCTION

Introduction of Graphical User Interfaces (GUIs) made it easy for most computer users to navigate and access information on the computer. However, some benefits of GUIs are not shared with users with visual impairments. In order to help visually impaired users access computers and electronic devices, auditory menus and screen readers (e.g., Job Access With Speech (JAWS) from Freedom Scientific; Window-Eyes from GW Micro; or VoiceOver by Apple Inc.) have become available. However, plain text-to-speech (TTS) is not always the optimal auditory presentation for orientation, navigation, and manipulation.

Researchers in the field have been working to develop new auditory displays, using both speech and non-speech sounds, while studying the situation-optimized usage of existing display methods. Various types of auditory menus and enhancements, such as auditory icons (Gaver, 1986), earcons (Blattner, Sumikawa, & Greenberg, 1989), spearcons (Walker, Nance, & Lindsay, 2006), and spindex cues (Jeon & Walker, 2009) have been explored. Spearcons and spindex cues have bee particularly successful, and have been studied extensively in recent years. However, both of these cues leverage the language of the user, so it is important to determine how effective they are in languages other than English. One study of spearcons in German and Hungarian has appeared (Wersenyi, 2010), but to our knowledge, none has been conducted in other languages, including Korean.

Also, spearcons are typically used as non-speech enhancements to regular TTS menus. As a user becomes more familiar with the menu, they come to rely more on the spearcon, and less on the TTS (Walker, Nance, & Lindsay, 2006, Palladino & Walker, 2007, 2008). However, it is also interesting to study whether auditory menus can be constructed entirely from spearcons.

Auditory Menu

Many types of auditory displays, particularly auditory menus, have been studied either as enhancements to visual displays or as the primary means for interacting with a system or device. With the trend toward smaller displays, mobile devices, and increasing technology use by the visually impaired, it becomes more important to improve usability of non-GUI interfaces such as auditory menus (Walker, Nance, & Lindsay, 2006). The benefits of having a robust auditory communication channel include increasing mobility of the users and decreasing bandwidth requirements on the often-overloaded visual channel (Frauenberger & Stockman, 2006). Auditory menus (especially with enhanced cues) can also be effectively used in eyes-free situations (e.g. while jogging, driving) or in devices with small or non-existing displays (e.g., Apple's iPod Shuffle). Although the most obvious users of auditory-enhanced systems are the visually impaired, auditory displays have the potential to benefit users with all levels of vision. For instance, Jeon and Walker (2010) showed that adding spindex cues to the auditory menu improves performance and subjective satisfaction, and reduces perceived workload with both visually impaired and sighted users.

Various auditory enhancements have been proposed in order to maximize the functionality of interaction and improve the basic TTS representation. There are several designs that have been explored including: basic speech with no enhancement, adding auditory icons to a speech-based menu, adding earcons, adding spearcons, and adding spindex. All of these have their advantages and disadvantages. The present study focuses on the use of spearcons either as an enhancement to TTS or as an independent auditory menu presentation after a short learning period. We compared the menu navigation efficiency, learning, perceived speediness, and subjective satisfaction using the auditory version menus of the Microsoft Internet Explorer 9 (IE9) menu.

Types of Non-speech Auditory Cues

Auditory Icons

Auditory Icons (Gaver, 1986) are "caricatures of naturally-occurring sounds, which could be used to provide information about sources of data". Each auditory icon is usually an everyday sound that can be easily associated with its source (e.g., the sound of a typewriter). Because an auditory icon is a fairly direct representation of an object or action, auditory icons provide a natural way to represent dimensional data as well as many objects in a system. As an example, emptying a recycle bin on computer makes the sound (i.e., has the auditory icon) of crumpling paper in order to help users identify and have additional feedback about the event. The biggest advantage of auditory icons is that they require little learning to develop a connection between the sound and the event because there is already an underlying connection developed. However, it is sometimes difficult to match all functions of devices with proper auditory icons. This is partly due to problems in representing abstract concepts with no natural sound associated with them (e.g., "Save a file"). Additionally, there are some difficulties in creating two distinguishable yet similar natural sounds to represent related actions in a computing system (e.g. "Save" and "Save As"). Therefore, the utility of auditory icons in computer applications is limited.

Earcons

Earcons (Blattner et al., 1989) are defined as "non-verbal audio messages that are in the computer-user interface to provide information to the user about some computer object, operation, or interaction". Earcons are composed of musical motives, which are short, rhythmic sequences of pitches with variable intensity, timbre, and register (Brewster, Raty, & Kortekangas, 1996). An important advantage of earcons is that they can represent not only the menu item content, but also its location in a menu logically. Generally, it is well-known that for auditory menus, unlike a visual menu where users can easily scan and determine location, it takes considerable time for a user to form a mental model of the entire menu structure. However, by considering each earcon as a node on in a hierarchy, and inheriting all the properties of the earcons on the upper level, Brewster et al. (1996) showed that earcons are a robust and extensible method of communicating hierarchical information in auditory interfaces. In addition, because earcons use an arbitrary mapping, they can represent any menu item, which is not always possible with auditory icons. Unfortunately, this arbitrary mapping for earcons means there is no intuitive link between the earcon and what is represented. Thus, linking of the associations requires users to undergo considerable learning to be effective.

Spindex Cues

A spindex (Jeon & Walker, 2009) is a speech-based index that uses brief audio cues to inform users of their position in a long menu list, such as a mobile phone contact list or the song list of an MP3 player. It is created by associating auditory cues with each menu item, in which the cues are based on the pronunciation of the first letter of each menu item. The set of spindex cues are analogous to the visual tabs that facilitate flipping to the right section of a thick reference book such as a dictionary. Because alphabet sounds ("A", "B", "C", etc.) are sufficiently informative to users when sorting out the non-target items, spindex cues help users reach the alphabetical area proximal to the target faster. In addition, because the cues are made out of natural sounds (based on speech) and part of the original word, they do not require training or learning. Because spindex cues are more appropriate for long list type menus, they are not considered in the current study, which involves a two-dimensional hierarchical menu.

Spearcons

Spearcons (Walker et al., 2006), the main focus of the present study, are produced by speeding up the regular speech, even to the point where the resulting sound is no longer comprehensible as speech. Each spearcon is unique due to the specific underlying speech phrase, which allows them to be distinct, but at the same time allows similar phrases to form families of related sounds, much like earcons.

Because spearcons can be created using an algorithm to speed up a TTS phrase, they can be created automatically and on the fly. Palladino and Walker (2007) found that learning rates for an auditory menu scheme were faster when the menu was enhanced with spearcons, compared to auditory icons or earcons. Because the mapping between spearcons and the objects they represent is not arbitrary, less learning is required. Even though several studies have shown that spearcons + TTS outperformed TTS-only menus, there has been no research using spearcon-only menus (i.e., even without the TTS phrase), which is of interest in the current study. Based on this background, we investigated whether prepending spearcons, or using spearcons alone (with practice) would improve navigation efficiency on a two-dimensional auditory menu in Korean. We compared navigation time (i.e., time-to-target), perceived speediness, and subjective satisfaction of TTS-only and spearcon-enhanced Korean menus. We hypothesized that adding spearcons would improve performance (as has been found previously for English menus), and result in better navigation efficiency; and eventually users would no longer need TTS because spearcons can be informative enough after a short learning period.

METHOD

Participants

Twenty five Korean students at Georgia Tech (16 male, 9 female, mean age = 25.08 years) participated in this study for partial credit in psychology courses. Only native Korean speakers were recruited. They reported normal or corrected-to-normal hearing and vision. All signed informed consent forms and provided demographic information on age, gender, and language experience.

Apparatus and Equipment

Stimuli were presented using a 20″ iMac computer, running Max OSX 10.5.8 on a 2GHz Intel Core 2 Duo processor and 1GB of RAM. A software program written in

Table 1. The visual representation and English version of the auditory menu navigated by participants

	File	Edit	View	Favorites	Tools	Help
1	New tab	Cut	Toolbars	Add to favorites…	Delete browsing history…	Internet Explorer Help
2	Duplicate tab	Copy	Quick tabs	Add to Favorites bar	InPrivate browsing	What's New in Internet Explorer 9
3	New window	Paste	Explorer bars	Add current tabs to favorites…	Tracking protection…	Online support
4	New session	Select all	Go to	Organize favorites	ActiveX Filtering	About Internet Explorer
5	Open	Find on this page	Stop	Favorites Bar	Diagnose connection problems…	
6	Edit		Refresh		Reopen last browsing session	
7	Save		Zoom		Add site to Start menu	
8	Save as…		Text size		View downloads	
9	Close tab		Encoding		Pop-up Blocker	
10	Page setup…		Style		SmartScreen Filter	
11	Print…		Caret browsing		Manage add-ons	
12	Print preview…		Source		Compatibility View	
13	Send		Security report		Compatibility View settings	
14	Import and export…		International website address		Subscribe to this feed…	
15	Properties		Webpage privacy policy…		Feed discovery	
16	Work offline		Full screen		Windows Update	
17	Exit				F12 developer tools	
18					Internet options	

JAVA, using the APWidgets library (Davison & Walker, 2008), was used to run the experiment. It included randomization, response collection, and data recording. Participants listened to auditory stimuli using Sennheiser HD 202 headphones; at the beginning they were given the opportunity to adjust the fit and volume for personal comfort. The computer monitor was placed on a table 40cm in front of the seated participants.

Menu Structure

A two-dimensional menu from Microsoft Internet Explorer 9 (IE9) was used for the experiment (see Table 1). For the sake of practicality, we did not create a new random menu list but used an existing one. The IE9 menu had six top-level categories: File, Edit, View, Favorites, Tools, and Help, and each of them contained 4 to 18 items. The total number of items in the menu was 65.

Auditory Stimuli

There were two different types of auditory cues: TTS-only and spearcons + TTS. All of the menu item text labels were converted to speech using the "Yumi" (Female) voice from NeoSpeech TTS engine (http://neospeech.com/). Each word or text phrase was submitted separately to the TTS program online, and the resulting .wav file was saved for incorporation into the experiment. Spearcons were produced by speeding up the TTS phrase with the spearcons algorithm in MATLAB (Walker et al., 2006). It compressed each phrase logarithmically, so long phrases get a greater compression. The pitch of the sounds was maintained, and the spearc

were still related to the original source TTS sounds. Each spearcon + TTS stimulus was created by using GoldWave (http://www.goldwave.com/) software to insert 150ms interval (which is shorter than usual, c.f., Walker et al., 2006) between the spearcon and the TTS phrase.

Experimental task and procedure

There were two within-subjects conditions: TTS-only and spearcon + TTS. The task was to find specific menu items within the 2-dimensional menu hierarchy with the help of different auditory displays. Each sound cue was played accordingly as a participant navigated the menu structure, and no visual cues related to menu structure were provided (see Figure 1). Participants were instructed to find the target item as quickly as possible while still being accurate. Participants were allowed to use only the keyboard to navigate, with no pointing device such as a mouse; this is similar to what blind users normally use, and what is typically found on many mobile devices. Participants could navigate the menu with the four cursor keys, and when they found a target, they pressed the Enter key to confirm their selection. Each sound was preemptive so that when a participant recognized the current menu item was not the target, he or she could navigate to the next item without waiting.

Figure 1. Screen capture of experiment program

Figure 1 shows the screen that participants saw during the experiment. It is a full-screened application with white background with nothing but a target name. On each trial, a target item such as "오프라인으로 작업" (which is "Work offline" in English) was displayed on the bottom of the screen. In order to eliminate target perception (reading) time, and to allow us to measure navigation time only, participants were asked to press 'option + space bar' to signal the start of their navigation and to activate the auditory menu. The menu navigation timer started with the first arrow-key press to move within in the menu, and stopped with the Enter-key press that indicated selection of the target item. Following a trial, the next target item was immediately shown and participants repeated the same task with the new item (again, starting the trial at their own pace).

We did not force participants to find the correct target (i.e., the trial ended whenever they pressed the Enter key to select an item). We let them make mistakes but logged those errors along with navigation time. Also, the menu list wrapped around at the top and the bottom of the list. For example, if the last menu item was highlighted, pressing the down-cursor key would highlight the top menu item. Also, pressing the left- or right-cursor key during the navigation would highlight the title of the adjacent menu. Selecting top-level menu items as a target was prohibited.

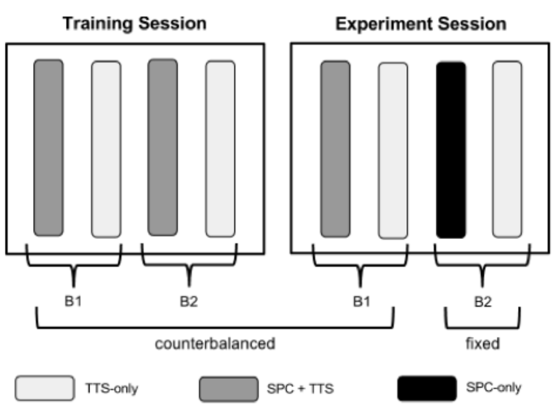

Figure 2. Procedure of the experiment

As shown in Figure 2, the Training session had two blocks, each of which had TTS-only and spearcons + TTS conditions. Each auditory cue condition had 25 trials, for a total of 100 trials for training. After each condition, a screen indicating the end of a block was shown to participants and they preceded to a new block. The order of the presentation of different auditory cues was counterbalanced across participants. Since the purpose of our experiment was evaluating each auditory display in a manner that was similar to real-world situations, we used training blocks as learning opportunities for participants to become familiar with the IE9 menu structure. Thus, in this experiment, the menu items were not reordered randomly (compare to Palladino & Walker, 2007, 2008).

The Experiment session also consisted of 2 blocks. The first block consisted of TTS-only and spearcons + TTS (the order was counterbalanced across participants), followed by the second block that consisted of spearcons-only (which was a main interest in this study) and then TTS-only. Although using spearcons-only after some experience with spearcons + TTS seems to be a natural strategy in a practical setting, we intentionally added the TTS-only condition at the end of the second block in order to show that plausible benefits of the spearcons-only design does not come only from the learning effect. After the experiment, participants filled out a short questionnaire. In addition to the response to perceived speediness, six-point Likert-type scales were used to assess their perceived performance (functional helpfulness, and effectiveness), and preference (likable, fun, and annoying) (Jeon & Walker, 2009) on each type of auditory stimuli. Then, they were asked to provide any comments about the study.

RESULTS

Errors in the tasks arising from selecting incorrect items were very rare, so we focus here on navigation time of correct trials. Figure 3 shows mean navigation time to the target item, across blocks and auditory cue conditions. Results were analyzed with a repeated measures analysis of variance (ANOVA). There was no main effect of auditory cue type, but there was a main effect of block, $F(1, 24) = 21.01$, $p < .001$, which means learning effects from Block 1 to Block 2. More importantly, there was an interaction between auditory cue type and block, $F(1, 24) = 6.90$, $p < .05$. This captures the fact that in Block 1 spearcons + TTS was no different from TTS alone, whereas in Block 2 spearcons-only was even faster than TTS.

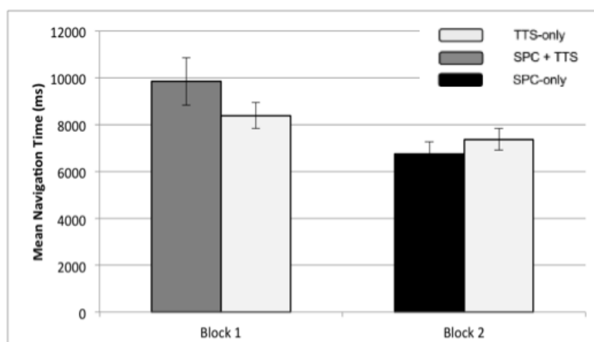

Figure 3. Interaction effect of auditory cue type and block

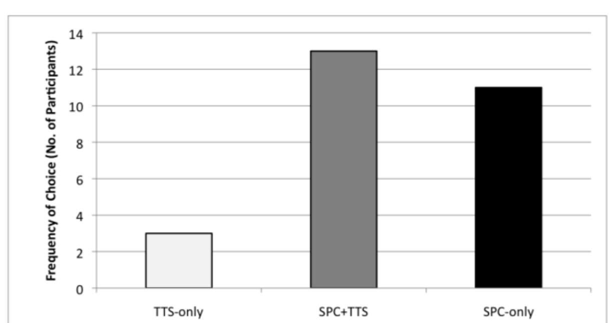

Figure 4. Frequencies of the perceived speediness

For the perceived speediness of the auditory cue types (Figure 4), participants clearly felt that the spearcons + TTS (N=13) and the spearcons-only (N=11) were faster than TTS-only (N=3) (two participants chose two types). Actual frequencies of the perceived speediness were significantly different from the null-case in which all frequencies are equal, $\chi^2(2, 27) = 6.22$, $p < .05$.

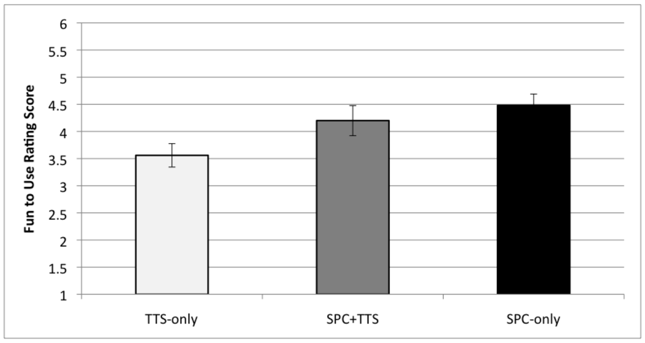

Figure 5. Fun-to-Use Rating Result

In the subjective rating data, the 'fun to use' scale showed a reliable difference among the conditions, $F(2, 48) = 4.86$, $p < .05$. Participants rated that adding spearcons to TTS ($p = .065$) or spearcon-only cues ($p = .01$) were more fun to use than plain TTS (see Figure 5). Moreover, there was no significant difference in the annoyance scale among cue types.

DISCUSSION

Palladino and Walker (2008) showed that in English, spearcons-enhanced auditory menus could significantly improve menu navigation performance compared to basic TTS-only menus in two-dimensional menus. Here, in Korean, we also showed that spearcon cues (especially, spearcons-only cues) can improve auditory menu navigation compared to TTS-only menus. We observed that after some practice, participants showed better performance in the spearcons-only condition than in the TTS-only condition, even though the last TTS-only condition always came later in the experimental order (and thus should have had the benefit of more exposure). High potential is shown for the use of spearcons in auditory menus because it can enhance navigation efficiency regardless of whether the spearcons are comprehensible as speech, or not. Even though the difference between the spearcons-only and the TTS-only did not reach conventional levels of statistical significance in the second block, we can expect that the tendency (lower navigation time in the spearcons-only) will persist, and would reach statistical significance with more blocks. Therefore, we might be closer to the generalization of the benefits of spearcons across languages. Moreover, this is exactly what we envisioned for the practical usage of spearcons. When users get familiar with the menu structure, spearcons-only cues can enhance navigation performance in auditory menus. We also expect that visually impaired users can benefit more from this spearcons-only menu because many of them are already familiar with the use of sped-up speech (though, note that spearcons are not the same as fast-talking TTS engines).

Anecdotally, there are potential reasons why prepending spearcons to TTS did not yield better performance but spearcon alone did. For the spearcons + TTS condition, participants got to know that spearcons would be followed by TTS. This may have made participants listen through the spearcon and the interval to hear the TTS phrase, rather than just passing by and concentrating on the spearcon itself. To support this, some participants reported, "Although I got spearcon cues, I still wanted to hear the following TTS to double check".

It is important to find that even though spearcons are a relatively a new type of auditory cue for these Korean speakers, participants did not consider it to be unusual or annoying. Also, participants thought spearcons were fun to use, which is an important factor for embedding them into the mobile device. In addition, it is promising that spearcon-enhanced menus obtained significantly higher ratings on perceived speediness, which indicates that participants' perceptions of the utility of spearcons matched reality. Planned research includes the replication of this study with other languages, with other auditory cues, and with visually impaired users.

REFERENCES

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

Brewster, S., Raty, V. P., & Kortekangas, A. (1996). Earcons as a method of providing navigational cues in a menu hierarchy. *Proceedings of HCI'96*, pp. 167-183.

Davison, B. K., & Walker, B. N. (2008). AudioPlusWidgets: Bringing sound to software widgets and interface components. *Proceedings of ICAD2008,* Paris, France.

Frauenberger, C. and Stockman, T., 2006. Patterns in Auditory Menu Design. *Proceedings of ICAD2006*, London, UK.

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

Jeon, M., & Walker, B. N. (2009). "Spindex": Accelerated initial speech sounds improve navigation performance in auditory menus, *Proceedings of HFES2009,* San Antonio, TX, pp. 1081-1085.

Jeon, M., & Walker, B. N. (2010). Spindex (speech index) improves auditory menu acceptance and navigation performance, *ACM Trans.Access.Comput. 3*(3), 10:1-26.

Palladino, D. K., & Walker, B. N. (2007). Learning Rates for Auditory Menus Enhanced with Spearcons versus Earcons. *Proceedings of ICAD2007,* pp. 274-279.

Palladino, D., & Walker, B. N. (2008). Navigation efficiency of two dimensional auditory menus using spearcon enhancements. *Proceedings of HFES2008,* NY (22-26 September). pp. 1262-1266.

Walker, B. N., Nance, A., & Lindsay, J. (2006). Spearcons: Speech-based earcons improve navigation performance in auditory menus. *Proceedings of ICAD2006,* London.

Walker, B. N., & Kogan, A. (2009). Spearcons enhance performance and preference for auditory menus on a mobile phone. *Proceedings of HCI International 2009*, San Diego, CA, USA, (19-24 July).

Wersényi, Gy. (2010). Auditory representations of a Graphical User Interface for a better Human-Computer Interaction. *Lecture Notes in ComputerScience (LNCS) 5954*, Springer Verlag, Berlin, 2010. pp. 80-102.