Spearcons (Speech-Based Earcons) Improve Navigation Performance in Advanced Auditory Menus

Bruce N. Walker, Jeffrey Lindsay, Amanda Nance, Yoko Nakano, Dianne K. Palladino, Tilman Dingler and Myounghoon Jeon

Human Factors: The Journal of the Human Factors and Ergonomics Society 2013 55: 157 originally published online 2 July 2012
DOI: 10.1177/0018720812450587

Spearcons (Speech-Based Earcons) Improve Navigation Performance in Advanced Auditory Menus

Bruce N. Walker, **Jeffrey Lindsay**, **Amanda Nance**, **Yoko Nakano**, **Dianne K. Palladino**, **Tilman Dingler**, and **Myounghoon Jeon**, Georgia Institute of Technology, Atlanta, Georgia

Objective: The goal of this project is to evaluate a new auditory cue, which the authors call *spearcons*, in comparison to other auditory cues with the aim of improving auditory menu navigation.

Background: With the shrinking displays of mobile devices and increasing technology use by visually impaired users, it becomes important to improve usability of non-graphical user interface (GUI) interfaces such as auditory menus. Using nonspeech sounds called *auditory icons* (i.e., representative real sounds of objects or events) or *earcons* (i.e., brief musical melody patterns) has been proposed to enhance menu navigation. To compensate for the weaknesses of traditional nonspeech auditory cues, the authors developed spearcons by speeding up a spoken phrase, even to the point where it is no longer recognized as speech.

Method: The authors conducted five empirical experiments. In Experiments 1 and 2, they measured menu navigation efficiency and accuracy among cues. In Experiments 3 and 4, they evaluated learning rate of cues and speech itself. In Experiment 5, they assessed spearcon enhancements compared to plain TTS (text to speech: speak out written menu items) in a two-dimensional auditory menu.

Results: Spearcons outperformed traditional and newer hybrid auditory cues in navigation efficiency, accuracy, and learning rate. Moreover, spearcons showed comparable learnability as normal speech and led to better performance than speech-only auditory cues in two-dimensional menu navigation.

Conclusion: These results show that spearcons can be more effective than previous auditory cues in menu-based interfaces.

Application: Spearcons have broadened the taxonomy of nonspeech auditory cues. Users can benefit from the application of spearcons in real devices.

Keywords: auditory menus, spearcons, auditory icons, earcons

Address correspondence to Bruce N. Walker, School of Psychology, Georgia Institute of Technology, 654 Cherry Street, Atlanta, GA 30332-0170, e-mail: bruce.walker@psych.gatech.edu.

HUMAN FACTORS
Vol. 55, No. 1, February 2013, pp. 157-182
DOI:10.1177/0018720812450587

INTRODUCTION

With visual displays shrinking or disappearing because of mobile and ubiquitous computing applications, and with the increasing use of technology by users who cannot look at or cannot see a traditional visual interface, it is important to identify methods or techniques that can improve the usability of non–graphical user interface (GUI) interfaces (e.g., Edwards, 1989; Gaver, 1989; Mynatt & Edwards, 1992; Raman, 1997). Often, nonvisual interfaces are implemented via a menu structure. Although considerable research has begun to lead to a visual menu design theory (e.g., Norman, 1991; Shneiderman, 1998, chap. 7) and to improve it (e.g., Bederson, 2000; Findlater & McGrenere, 2004; Sears & Shneiderman, 1994), there are still many open questions when it comes to nonvisual menus. The foundation of auditory menus is text to speech (TTS), but TTS-only menus are slow and limited. Accordingly, nonspeech audio cues including auditory icons (Gaver, 1986) and earcons (Blattner, Sumikawa, & Greenberg, 1989) have been suggested as ways to improve TTS-only interfaces. Although these are generally promising, there are shortcomings to the use of these enhancements, which may be resolved with the introduction of novel methods of creating auditory cues, such as spearcons (described in detail later) and the spindex (Jeon & Walker, 2011). In the current article, we focus on the potential benefits of spearcons and then present a systematic empirical evaluation of their effectiveness compared to auditory icons, to earcons, and to spoken menu items with no added auditory cues. This new technique is designed to help improve performance and usability of auditory menu-based interfaces as well as to make many interfaces more accessible to a broader group of users, in a wider range of applications and situations.

Auditory Menus

In applications as varied as telephone-based reservation systems, mobile phone operating systems, and desktop computing environments, presenting menu options via sound can greatly enhance the range of uses and users. In auditory menus, menu items are generally converted from text labels into spoken phrases using automated speech synthesis, or TTS software. Often a user navigates through an auditory menu by pressing "up" and "down" navigation keys and listening to the resulting TTS phrases. When the listener hears the desired menu item, a "select" or "enter" button (or sometimes a spoken command) is used to choose that item.

Because of the transient nature of sounds, there are important usability challenges inherent in auditory menus. Since it takes some time to listen to each menu item, quick and efficient movement through a menu structure can be difficult. Furthermore, as one moves about in a menu hierarchy, it can be difficult to maintain an awareness of which menu or submenu is currently active. Finally, since there is considerable memory load for auditory interfaces in general, learning an auditory menu structure—which generally enhances usability—can be difficult. Fairly recently, Zhao, Dragicevic, Chignell, Balakrishnan, and Baudisch (2007) introduced the earPod, in which users can benefit from the motor memory in addition to auditory cues by sliding their thumb on the circular touchpad. However, it requires a totally new device design and could not easily be incorporated into the existing interface.

To overcome these challenges of auditory menus, auditory researchers have developed some auditory menu enhancement techniques that are either menu *item*-level approaches or menu *structure*-level approaches. At the item level, every single menu item has a one-to-one mapping between sound and meaning, and thus "what" an item is, is important. In contrast, at the menu structure level, the focus is how to easily know approximately "where" the item is in the entire menu structure. Auditory icons (Gaver, 1986) are a representative item-level approach to enhancing auditory menus. Earcons (Blattner et al., 1989) are often suggested as a structure-level enhancement. In addition to earcons, auditory scroll bars (Yalla & Walker, 2008) also address the structure-level aspect of auditory menu usability. Our new sound cue, spearcons, can be categorized as an item-level approach to enhancing TTS menus but may also have the potential to improve the menu structure level like earcons in some ways.

The enhancements discussed here are typically accomplished by prepending a brief sound called a cue (i.e., an earcon, auditory icon, or spearcon) to the TTS phrase. As soon as the user navigates to a menu item, he or she hears the cue, and then the TTS phrase. The user can either select the current item or move to the next item, without necessarily hearing all (or, in some cases, any) of the TTS phrase. That is, if the cue sound is sufficiently informative, then the user need not listen to the TTS phrase. That clearly can lead to faster navigation. Therefore, our focus is how to make the cues sufficiently informative while keeping cues easy to learn.

The Improvements of Speech Menus and the Use of Sped-Up Speech

There have been several attempts to improve speech interfaces (Asakawa & Itoh, 1998; Morley, Petrie, O'Neill, & McNally, 1998; Pitt & Edwards, 1996; Thatcher, 1994), but most of them aim to help specifically visually impaired users. Certainly visually impaired populations may benefit most from speech interfaces, but sighted people can also benefit from them, as discussed before. Furthermore, most of the studies just cited address more qualitative and subjective data than objective and quantitative performance (e.g., the preference about the application of different voice gender). Thus, more systematical research is needed.

A more performance-directed enhancement (i.e., focusing more on navigation speed rather than intelligibility or aesthetics) in speech menu systems is to use sped-up speech, which is generally used in screen readers by visually impaired users. In fact, research showed promising results for the use of sped-up speech. For example, Asakawa, Takagi, Ino, and Ifukube (2003) showed that experienced blind users could listen to spoken material at a speech rate 1.6 times faster than the highest rate of the tested TTS engine. More recent research also showed that blind people might learn to understand synthesized speech at speaking rates up to 25 syllables per second, exceeding by far the maximum performance level of sighted people (Moos & Trouvain, 2007). However, whereas the use of sped-up speech can certainly improve accessibility for "advanced" visually impaired users, it is doubtful whether it is so useful for novices (including visually impaired users as well as sighted users who never learn and get familiar with that specific speech presentation type). Since sped-up speech speeds up every part of an auditory interface, the use of sped-up speech seems to quickly go beyond novices' cognitive capacity.

The Use of Auditory Icons and Earcons

Auditory icons (Gaver, 1986) are audio representations of objects, functions, and events. They are caricatures of naturally occurring sound-producing events such as bumps, scrapes, or even files "landing in" trash bins. As caricatures, auditory icons capture an event's essential features, by presenting a representative sound for the objects involved. Auditory icons can represent various objects or events in electronic devices more clearly than some other auditory cues because the relation between a source of sound and a source of data is generally quite natural. For example, a typing sound can represent a typewriter or typing, or even printing. Thus, auditory icons typically require little training and are easily learned. Adopting these advantages, Gaver (1989) created an auditory icon-enhanced desktop. Also, some researchers have attempted with mixed success to convert entire GUIs to nonvisual interfaces using auditory icons (e.g., Mynatt, 1997; Mynatt & Weber, 1994).

One alternative to auditory icons is earcons (Blattner et al., 1989). Earcons are brief musical melodies consisting of a few notes whose timbre, register, and tempo are manipulated systematically, to build up a "family of sounds" whose attributes reflect the structure of a hierarchy of information (Brewster, Wright, & Edwards, 1993). Using earcons has often been proposed as a method to add context to a menu in a user interface, helping users maintain awareness of where in the menu hierarchy they are currently located. Earcons have been applied to various menu systems ranging from GUI applications (Brewster, Raty, & Kortekangas, 1996), to mobile phones (LePlâtre & Brewster, 1998), to telephone-based auditory interfaces (Brewster, 1997, 1998). Menus in GUIs may also be improved by adding earcons to help prevent the user from selecting the wrong menu item, or from "slipping off" a chosen item (Brewster & Crease, 1999). In addition, earcons have been proposed as a way to help speed up a speech-based interface, including those designed for visually impaired users (e.g., Karshmer, Brawner, & Reiswig, 1994), as well as those intended for general usage such as in-vehicle displays (e.g., Vargas & Anderson, 2003). In these applications, the sound is meant to help the users know not just where they are in the menu hierarchy but also what the content of a menu item is (also see Wolf, Koved, & Kunzinger, 1995). Absar and Guastavino (2008) provide a recent overview of the use of auditory icons and earcons.

Issues With Auditory Icons and Earcons

When using either auditory icons or earcons in an interface, there are some important issues such as "ease of sound creation" and "flexibility of the auditory menu interface." Because auditory icons can have a direct mapping between the sounds and the menu items they represent, this can reduce learning or training time. On the other hand, auditory icons are sometimes difficult to create for many menu items, specifically those in computer interfaces that have no real sound (e.g., "connect to server" or "export file"; see Palladino & Walker, 2008a). As a result, there have been few systematic uses of auditory icons specifically in auditory menus. In terms of sound creation, earcons are likely to need a sound designer to create aesthetic sounds. Applying arbitrary mappings between musical notes and menu items, with no standard set of earcons, also leads to the need for initial training. In addition, there may be very limited transfer of training when moving between various systems employing different earcon "languages."

From a systems engineering perspective, the flexibility of menus that use either earcons or auditory icons is *brittle*, in that a change to either the menu hierarchy or menu items is not well supported by the sounds. If a menu or menu item needs to be added, then each new auditory icon needs to be created *manually* (assuming an iconic sound can be found for the new item). This need for manual intervention is clearly a problem for dynamic systems. The hierarchical earcon approach may handle the addition of menu items automatically, so long as the item is added *after* the existing items. For example, adding an item to the bottom of a menu would mean that the next timbre or tempo or pitch from a preset list could be used to create the earcon appropriately. This requires that the method for creating earcons anticipates a great enough variety in menu items to handle the menu growth. This can be hard to predict, especially for systems that have varied usage or long life expectancies. More problematic is when a new menu item is inserted in the middle of a menu. For example, if the first item in a file list starts with "C," it is likely that items will subsequently be added ahead of it in the list (i.e., as soon as a file whose name starts with "B" is created). Menus enhanced with earcons do not handle this situation very well, nor do they handle the related challenge of re-sorting or reordering menus (as is often done in "intelligent" menus that bubble the most commonly selected items toward the top). Either the hierarchical order of the earcons must be rearranged, which diminishes their role in providing context, or else the learned mappings for every earcon below the new menu item will need to be relearned.

To summarize these issues, Figure 1 presents the dimensions of "ease of sound creation" and "flexibility of the auditory menu interface." Neither earcons nor auditory icons rate highly in both dimensions. An optimal solution, then, would be sounds that (a) can be simply and automatically generated, (b) provide less arbitrary mappings than earcons, (c) cover a wider range of menu content than auditory icons, and (d) are flexible enough to support rearranging, re-sorting, interposition, and deletion of menu items. If such sounds could also increase the speed and/or accuracy of menu selections, they would be even more useful.

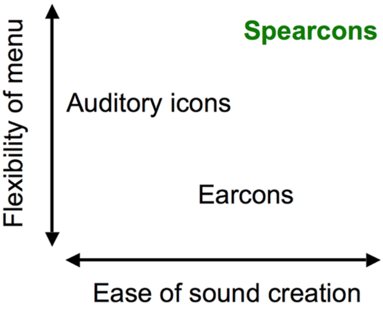


Figure 1. Relative position of auditory cue types along two axes important in menu effectiveness and usability. In theory, spearcons should be better than previous auditory cue types in terms of both the ease of sound creation and the flexibility of the resulting menu structure.

Spearcons: Speech-Based Earcons

Spearcons in auditory menus are brief audio cues that can play similar roles as auditory icons and earcons, but presumably in a more effective manner, overall. Spearcons are created automatically by converting the text of a menu item (e.g., "Export File") to speech via TTS and then speeding up the resulting audio clip (without changing pitch), even to the point where it is no longer comprehensible as speech. Spearcons are unique to the specific menu item, just as with auditory icons, though the uniqueness is acoustic, and not semantic or metaphorical. At the same time, the similarities in menu item content cause the spearcons to form families of sounds. For example, the spearcons for "Save," "Save As," and "Save As Web Page" are all unique, including being of different lengths. However, they are acoustically similar at the beginning of the sounds, which allows them to be grouped together (even though they are not comprehensible as any particular words). The different lengths help the listener learn the mappings and provide a "guide to the ear" while scanning down through a menu, just as the ragged right edge of items in a visual menu aids in visual search.

Since the mapping between spearcons and their menu item is nonarbitrary, there should be less learning required than would be the case for a purely arbitrary mapping. Moreover, as discussed at the outset, spearcons are prepended to the normal TTS phrase. Thus, users can take their time to become familiar with the use of the system and gradually take advantage of spearcons. This smooth transition is a big distinction between spearcons and sped-up speech systems in which there are just dichotomous states: normal speech and speeding up everything.

The menus resulting from the use of spearcons can be rearranged, be sorted, and have items inserted or deleted, without changing the mapping of the various sounds to menu items. Spearcons can be created algorithmically, so they can be created dynamically, and can represent any possible concept. Thus, spearcons should support more "intelligent," flexible, automated, nonbrittle menu structures. Now, it should be said that in menus that never change and where navigation is particularly important (e.g., particularly complex menus), spearcons may not be as effective at communicating their location as hierarchical earcons. However, spearcons would still provide more direct mappings between sound and menu item than earcons and cover more content domains, more flexibly, than auditory icons.

To evaluate this theoretical assessment using real user data, we conducted a series of five experiments comparing menu navigation performance using spearcons to traditional cues such as auditory icons and earcons. In Experiments 1 and 2, we measured menu navigation efficiency and accuracy among cues. In Experiments 3 and 4, we evaluated learning rate of cues and speech itself. In Experiment 5, we assessed spearcon enhancements compared to plain TTS in a two-dimensional auditory menu.

EXPERIMENT 1

In Experiment 1, we focused on assessments of menu navigation time and accuracy rate. Based on the characteristics of auditory cues described before, we hypothesized that spearcons would outperform other auditory cues in terms of mean time to target and mean accuracy. To test these hypotheses, we conducted the first empirical experiment with four different auditory cue types (TTS only; earcons + TTS; auditory icons + TTS; and spearcons + TTS).

Method

Participants. Experiment 1 involved nine undergraduate students (4 male, 5 female, age range = 19–21) who reported normal or corrected to normal hearing and vision and who participated for partial credit in psychology courses.

Apparatus and equipment. A software program written in E-Prime (Psychological Software Tools, n.d.), running on a Dell Dimension 4300S PC with Windows XP, controlled the experiment, including randomization, response collection, and data recording. Listeners sat in a sound-attenuated testing room and wore Sony MDR-7506 headphones, adjusted for fit and comfort.

Menu structure. The menu structure chosen for Experiment 1 is presented in Table 1. In developing this menu, it was important not to bias the study against any of the audio cue methods. For that reason, the menu includes only items for which reasonable auditory icons could be produced. This precluded a computer-like menu (File, Edit, View, etc.) since auditory icons cannot be reliably created for items such as "Select Table." A computer menu was also avoided because that would necessarily be closely tied to a particular kind of interface (e.g., a desktop GUI, a mobile phone), which would result in confounding variables relating to previously learned menu orders. This is particularly important in the present experiments, in which it was necessary to be able to reorder the menus and menu items without prior learning causing differential carryover effects. That is, it was important to assess the effectiveness of the sound cues themselves, and not the participants' familiarity with a particular menu hierarchy. Thus, finally, a menu structure with animals, nature, objects, instruments, and people sounds was developed (refer to Table 1).

Auditory stimuli: TTS phrases. All of the menu item text labels were converted to speech using Cepstral (n.d.) TTS, except the word *camera*, which was produced using AT&T Research Labs (n.d.) TTS Demo program. This exception was made because the Cepstral version of that word was rated as unacceptable during pilot

TABLE 1: Menu Structure Used for Experiments 1, 2, and 3.

Animals	Nature	Objects	Instruments	People Sounds
Bird	Wind	Camera	Flute	Sneeze
Dog	Ocean	Typewriter	Trumpet	Cough
Horse	Lightning	Phone	Piano	Laughing
Elephant	Rain	Car	Marimba	Snoring
Cow	Fire	Siren	Violin	Clapping

testing. The speech phrases lasted on average 0.57 s (range = 0.29–0.98 s).

Auditory stimuli: Earcons. The earcon design was based on Brewster et al. (1996). For each menu item, hierarchical earcons were created using Apple (2007) GarageBand MIDI-based software. On the top level of the menus, the earcons included a continuous tone with varying timbre (instrument), including a pop organ, church bells, and a grand piano; these instruments are built into GarageBand. Each item within a menu used the same continuous tone as its parent. Items within a menu were distinguished by adding different percussion sounds, such as bongo drums or a cymbal crash (also from GarageBand). The earcons lasted on average 1.26 s (range = 0.31–1.67 s).

Auditory stimuli: Auditory icons. Sounds were identified from sound effects libraries and online resources. The sounds were as directly representative of the menu item as possible. For example, the click of a camera shutter represented "camera"; a neigh sound represented "horse." The sounds were manipulated by hand to be brief and still recognizable. Pilot testing ensured that all of the sounds were identifiable as the intended item. The auditory icons averaged 1.37 s (range = 0.47–2.73 s). Note that for the auditory icon and spearcon conditions, the category titles (e.g., "Animals") were not assigned audio cues—only TTS phrases, as described earlier.

Auditory stimuli: Spearcons. The TTS phrases were sped up using a pitch-constant time compression to ensure that they were generally not recognizable as speech sounds (though this is not strictly necessary). In this article, all of this time compression was accomplished by running TTS files through a SOLA (synchronized overlap add method) algorithm (Hejna, 1990; Roucos & Wilgus, 1985), which produces the best-quality speech for a computationally efficient time domain technique. By varying time scale options, we can directly specify the output length of the spearcons or specify in-to-out ratio with which the target length will be determined. TTS phrases can be compressed *linearly* (discussed earlier) or *logarithmically*, such that the longer words and phrases were compressed to a relatively greater extent than those of shorter words and phrases. Therefore, spearcons are not simply "fast talking" menu items; they are distinct and unique sounds, albeit acoustically related to the original speech item. They are analogous to a fingerprint—a unique identifier that is only part of the information contained in the original.

For Experiment 1 (and for Experiment 2), we used linear compression that resulted in around 40% to 50% the length of the original speech sounds. Spearcons averaged 0.28 s (range = 0.14–0.46 s).

Combined audio cues and TTS phrases. All of the sounds were converted to WAV files (22.1 kHz, 8 bit) for playback through the E-Prime experiment control program. For the three listening conditions where there was an auditory cue played before the TTS phrase (earcon, auditory icon, and spearcon conditions), the audio cue and TTS segment were added together into a single file for ease of manipulation by E-Prime. For example, one file contained the auditory icon for sneeze, plus the TTS phrase "sneeze," separated by a brief silence. This was similar to the approach by Vargas and Anderson (2003). For the "speech only" condition, the TTS phrase was played without any auditory cue in advance, as is typical in many TTS menus, such as in the JAWS screen reader software (Freedom Scientific, n.d.). The overall sound files averaged 1.66 s (range = 0.57–3.56 s).

Procedure

Task and conditions. The task was to find specific menu items within the menu hierarchy. On each trial, a target was displayed on the screen, such as, "Find *Dog* in the *Animals* menu." This text appeared on the screen until a target was selected to avoid any effects of a participant's memory for the target item. The menus, themselves, did not have any visual representation—only audio as described earlier.

The W, A, S, and D keys on the keyboard were used to navigate the menus (e.g., W to go up, A to go left), and the J key was used to select a menu item. When users moved onto a menu item, the auditory representation (e.g., an earcon followed by the TTS phrase) began to play. Each sound was interruptible such that a participant could navigate to the next menu item as soon as he or she recognized that the current one was not the target. Sounds representing the initial item would stop, and the new item sounds would start immediately.

Menus "wrapped," so that navigating "down" a menu from the bottom item would take a participant to the top item in that menu. Moving left or right from a menu title or menu item took the participant to the top of the adjacent menu, as is typical in software menu structures. Once a participant selected an item, visual feedback on the screen indicated whether the selection was correct. Participants were instructed to find the target as quickly as possible while still being accurate. This would be optimized by navigating based on the audio cues whenever possible (i.e., not waiting for the TTS phrase if it was not required). Listeners were also encouraged to avoid passing by the correct item and going back to it. These two instructions were designed to move the listener through the menu as efficiently as possible, pausing only long enough on a menu to determine if it was the target for that trial. On each trial the dependent variables of total time to target and accuracy (correct or incorrect) were recorded. Selecting top-level menu names was possible, but such a selection was considered incorrect even if the selected menu contained the target item.

After each trial in the block, the menus were reordered randomly, and the items within each menu were rearranged randomly to avoid simple memorization of the location of the menus and items. This was to ensure that listeners were using the sounds to navigate rather than memorizing the menus. This would be typical for new users of a system, or for systems that dynamically rearrange items. The audio cue associated with a given menu item moved with the menu item when it was rearranged. Participants completed 25 trials in a block, locating each menu item once. Each block was repeated twice more for a total of three blocks of the same type of audio cues in a *set* of blocks.

There were four listening conditions: TTS only; earcons + TTS; auditory icons + TTS; and spearcons + TTS. Each person performed the task with each type of auditory stimuli for one complete set. This resulted in a total of four sets (i.e., 12 blocks, or 300 trials) for each participant. The order of sets in this within-subjects design was counterbalanced using a Latin square.

Training. At the beginning of each set in the experiment, participants were taught the meaning of each audio cue that would be used in that condition. During this training period, the speech version of the menu name or item was played once, followed by the matching audio cue, followed by the speech version again. These were grouped by menu so that, for example, all animal items were played immediately following the animal menu name. In the TTS condition, each menu name or item was simply played twice in a row.

Results of Experiment 1

For navigation time analysis, we included only correct responses in all experiments, as is typical. Figure 2 presents the mean time to target (in seconds) for each audio cue type, split out by the three blocks in each condition for Experiment 1. Table 2 summarizes both time to target and accuracy results for Experiment 1, collapsing across blocks for simplicity. Considering both time to target and accuracy together, a 4 (auditory cue type) × 3 (block) repeated measures multivariate analysis of variance (MANOVA) revealed that there was a

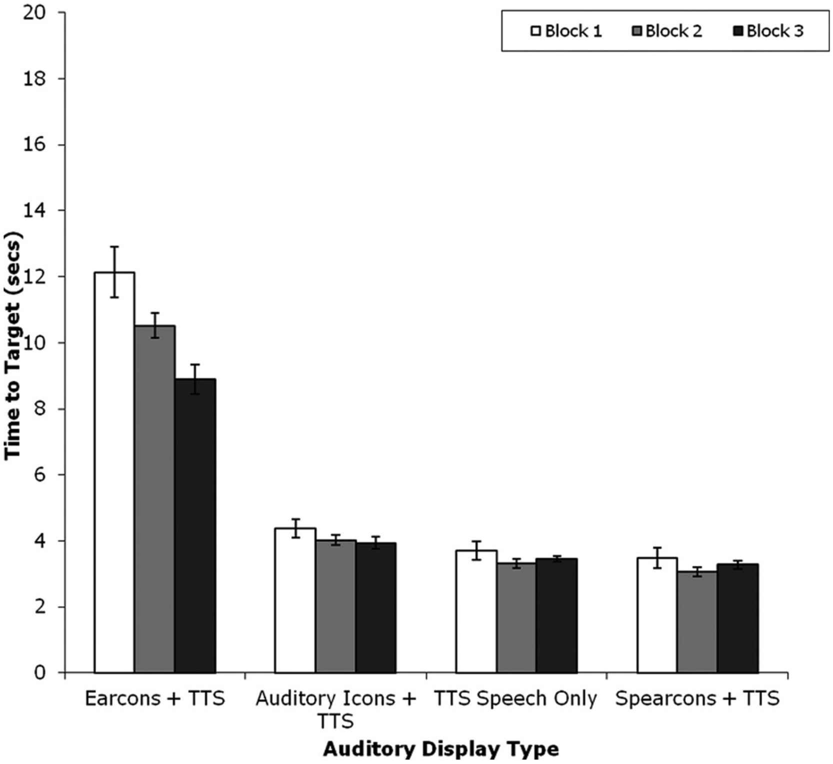


Figure 2. Mean time to target for each type of auditory display, for each block within each condition for Experiment 1. Note the practice effect and the relatively poor performance of hierarchical earcons. The TTS-only and spearcons + TTS conditions were statistically faster than both auditory icons and earcons. Error bars indicate standard error of the mean. TTS = text to speech.

TABLE 2: Overall Mean Time to Target and Mean Accuracy for Each Type of Audio Cue, Collapsed Across Block for Experiment 1

Type of Audio Cue	Time to Target (s)		Accuracy (%)	
	M	*SD*	*M*	*SD*
Spearcons + TTS phrase	3.28	0.52	98.1	1.5
TTS phrase only	3.49	0.49	97.6	2.0
Auditory icons + TTS phrase	4.12	0.59	94.7	3.5
Earcons + TTS phrase	10.52	11.87	94.2	5.4

Note. TTS = text to speech. Results are sorted by increasing time to target and decreasing accuracy. Spearcons were both faster and more accurate than auditory icons and hierarchical earcons.

significant difference between auditory cue types, $F(3, 6) = 40.20$, $p = .006$, Wilks's Lambda = .012, and between blocks, $F(5, 4) = 12.92$, $p = .008$, Wilks's Lambda = .088, but there was no interaction between auditory cue type and block. Because there was no trade-off between the two dependent variables (i.e., speed and accuracy), we conducted separate repeated measures ANOVAs for each dependent measure.

The separate ANOVA revealed that time to target (in seconds) was significantly different between conditions, $F(3, 24) = 177.14$, $p < .001$, $\eta_p^2 = .96$ (see Table 2). Pairwise comparisons showed that hierarchical earcons were the slowest auditory cue (*p*s < .001) followed by auditory icons. Spearcons were faster than the other two cue types (*p*s < .05). Although spearcons were also numerically faster than TTS only

(3.28 s vs. 3.49 s, respectively), this difference did not reach statistical significance ($p = .32$) in Experiment 1. The separate ANOVA for accuracy also found significantly different results between conditions, $F(3, 24) = 3.73$, $p = .025$, $\eta_p^2 = .32$, with the same pattern of results as for time to target (see Table 2).

The practice effect that is evident in Figure 2 is statistically reliable, such that participants generally got faster across the blocks in a condition, $F(2, 24) = 19.17$, $p < .001$, $\eta_p^2 = .71$. There was no change in accuracy across blocks, $F(2, 24) = 0.14$, $p = .87$, $\eta_p^2 = .02$, indicating a pure speedup, with no speed–accuracy trade-off again. The fastest earcon block (Block 3) was still much slower than the slowest auditory icon blocks (Block 1; $p = .001$). Anecdotally, a couple of participants noted that using the hierarchical earcons was particularly difficult, even after completing the training and experimental trials.

Discussion of Experiment 1

Earcons and auditory icons (particularly the former) have been proposed as beneficial additions to auditory menu items. The addition of such audio cues is sometimes proposed to speed up overall performance. More often, earcons and auditory icons are suggested to help provide navigational context and help prevent choosing the wrong item, or "slipping off" of the intended item. In Experiment 1, both earcons and auditory icons resulted in slower and less accurate performance than the TTS-only condition. This would argue against their usage in a speech-based menu system, at least as far as search performance is concerned. This is not too surprising, since the addition of a 1- or 2-s sound before each menu item would seem likely to slow down the user. This is particularly true with the earcons, since their hierarchical structure requires a user to listen to most or all of the tune before the exact mapping can be determined. On the other hand, the use of spearcons—speech-based earcons—led to performance that was *at least* as fast and accurate as speech alone, despite the prepended sound. It also seems likely that spearcons could gain in performance with greater familiarity. Spearcons were also clearly faster and more accurate than either earcons or auditory icons.

Although the performance levels are important on their own, the use of spearcons should also lead to auditory menu structures that are more flexible. Spearcon-enhanced menus can be re-sorted, and can have items added or deleted dynamically, without disrupting the mappings between sounds and menu items that users will have begun to learn. This supports advanced menu techniques such as bubbling to the top of a menu the most frequently chosen item, or the item most likely to be chosen in a given context. As discussed before, such "intelligent" and dynamic menus are not well supported by earcons, and auditory icons are of limited practical utility in modern computing systems where many concepts have no natural sound associated with them. Spearcons enable interfaces to evolve, as well. That is, new functionality can be easily added, without having to extend the audio design, which increases the life of the product without changing the interface paradigm.

EXPERIMENT 2

Experiments 1 and 2 were nearly identical, with the exception of small but important differences in the stimuli structure (described in detail later). The near replication of Experiment 1 in Experiment 2 was important to study the stability of the results as well as to allow for a more precise quantitative analysis of user interaction than was possible from the stimuli in Experiment 1.

Method

Participants. Experiment 2 had 11 undergraduates (6 male, 5 female, age range = 18–20), who reported normal or corrected to normal hearing and vision and participated for course credit. None had participated in Experiment 1.

Stimuli. In Experiment 1, the duration of the silence between the audio cue and the TTS was approximately, but not always exactly, the same length (about 250 ms). This slight inexactness made some advanced analyses difficult, so in Experiment 2 the duration of the silence was made to be identical for all stimuli (*exactly* 250 ms). This slight but important change was made so that it could be accurately determined if participants were responding after only hearing the

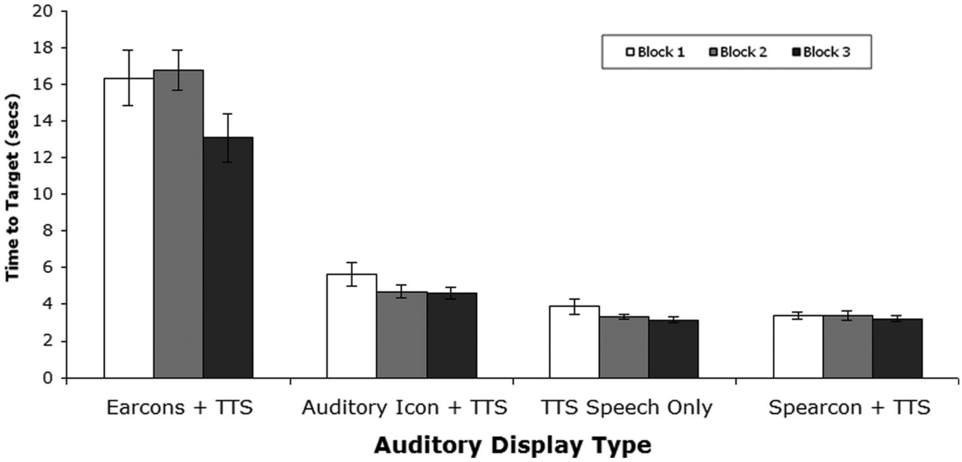


Figure 3. Mean time to target for each type of auditory display, for each block within condition for Experiment 2. Note the replication of the results from Experiment 1. Error bars indicate standard error of the mean.

audio cue or if they were also listening to some of the TTS segment before making their response.

Apparatus and procedure. The apparatus and all the experimental procedure, including task, conditions, and training, were the same as in Experiment 1.

Results of Experiment 2

Figure 3 shows the mean time to target (in seconds) for each audio cue type, split out by block for each condition in Experiment 2. Figure 4 shows the mean accuracy in the same configuration. Again a 4 (auditory cue type) × 3 (block) repeated measures MANOVA showed a significant difference between auditory cue types, $F(6, 5) = 40.04$, $p < .001$, Wilks's Lambda = .020, and between blocks, $F(4, 7) = 13.61$, $p = .002$, Wilks's Lambda = .114, but there was no interaction between auditory cue type and block. Because there was no interaction or trade-off between two dependent variables, we conducted separate repeated measures ANOVAs for each dependent measure.

As in Experiment 1, the separate ANOVA showed that time to target was significantly different between conditions, $F(3, 30) = 95.68$, $p < .001$, $\eta_p^2 = .91$. Pairwise comparisons revealed that all the auditory cues differed significantly from each other in time to target except for spearcons and TTS. Hierarchical earcons were significantly slower than auditory icons ($p < .001$), TTS ($p < .001$), and spearcons ($p < .001$). Auditory icons were significantly slower than TTS ($p = .001$) and spearcons ($p = .008$). Accuracy between the auditory cues was also significantly different, $F(3, 30) = 5.22$, $p = .04$, $\eta_p^2 = .34$. Pairwise comparisons showed auditory icons to be significantly less accurate than TTS ($p = .046$) and spearcons ($p = .041$). Similarly, hierarchical earcons were significantly less accurate than TTS ($p = .040$) and spearcons ($p = .038$). There was no significant difference in accuracy between hierarchical earcons and auditory icons or between TTS and spearcons.

The refined stimuli in Experiment 2 allowed us to conduct a more detailed analysis of whether participants made their judgments based on listening to just the prepended sound or whether they also listened to the TTS phrase. Thus, Table 3 shows the mean percentage of times participants listened to a portion of the TTS speech for each auditory cue, along with the corresponding standard errors of the mean. A repeated measures ANOVA conducted on this measure showed a significant difference between auditory cue types, $F(2, 20) = 144.654$, $p < .001$, $\eta_p^2 = .94$. A pairwise comparison revealed that participants listened to the TTS phrase significantly more

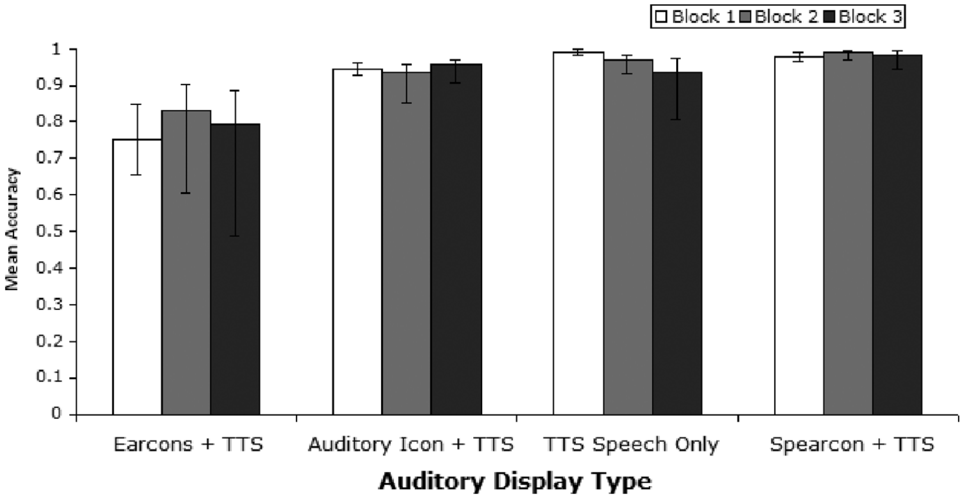


Figure 4. Mean accuracy for each type of auditory display, for each block within condition for Experiment 2. Note the relatively poor accuracy for hierarchical earcons, and the near-ceiling performance for spearcons. Error bars indicate standard error of the mean.

when using hierarchical earcons than when using auditory icons ($p < .001$) or spearcons ($p < .001$), and they listened to TTS significantly more when using auditory icons compared to spearcons ($p = .032$). It is important to note that these data reflect every auditory cue of a given type the participants listened to (i.e., when performing a single trial during a block using auditory icons a participant would listen to multiple icons per trial while traversing the menu), and not just a measure per trial.

Discussion of Experiment 2

In Experiment 2, we replicated Experiment 1 with more systematically designed stimuli and obtained very similar results. In terms of menu navigation efficiency and accuracy, traditional auditory icons and earcons showed significantly degraded performance compared to spearcons. Because there was no trade-off between two dependent measures, we can say that spearcons can enhance auditory menu navigation speed and accuracy.

One comment that could be made about spearcons is that perhaps these cues lead to faster performance simply because they are shorter than earcons and auditory icons. This is partially true, but that is simply a structural benefit of spearcons. The musical structure of earcons, and the acoustic realities of auditory icons, essentially "forces" them to be longer, so spearcons have an advantage from the outset, which is reflected in the performance results here. In addition, the detailed analysis in Experiment 2 clearly shows that participants listened to TTS almost half the time they were using earcons, while doing so less than 1% of the time when using auditory icons and spearcons. This demonstrates that performance is not dependent only on the length of the auditory cue since auditory icons in this study were longer, on average, than earcons, yet they led to considerably better performance. In any case, none of this discussion about the length of the sounds diminishes the fact that spearcons also lead to better accuracy than auditory icons or earcons.

EXPERIMENT 3

Although Experiments 1 and 2 addressed the issue of speed and accuracy in menu navigation, and showed that spearcons outperform auditory icons and earcons, it still remains unclear how learning rates vary for menu items enhanced with different types of sounds. Therefore, in Experiments 3 and 4, we assessed

TABLE 3: Mean Percentage of Times Participants Listened to TTS Speech Phrase for Each Auditory Cue Type for Experiment 2

Type of Auditory Cue	*M* (%)	*SE* (%)
Spearcons	0.11	0.06
Auditory icons	0.64	0.20
Earcons	49.68	4.15

Note. TTS = text to speech. Results are sorted by increasing percentage of times listening to speech. Speech was listened to significantly less often when using spearcons than when using auditory icons or hierarchical earcons.

learning rates for spearcons compared to other auditory cue types.

If spearcons are more easily learned, they will decrease frustration for the user and increase usability, and this interface enhancement will be more likely to be adopted by device manufacturers. As an initial assessment of learning rates, in Experiment 3 we examined the average number of trials needed for a user to learn menus of words presented with cues that were either spearcons or earcons. Earcons have required quite short learning sessions (e.g., Brewster et al., 1996). Therefore, in Experiment 3 we initially compared the learning rate of spearcons with only earcons; in Experiment 4 we tried to extend the comparison range to more general auditory cues including earcons, auditory icons, and a couple of hybrids of the existing ones.

Method

Participants. For extra credit in psychology courses, 24 undergraduate students (9 male, 15 female, mean age = 19.9) with normal or corrected to normal hearing and vision participated in Experiment 3. Participants were also required to be native English speakers. Five of these participants, plus an additional six participants, also participated in a brief follow-up experiment of spearcon comprehension. The age range and gender composition of these additional six participants is included in those mentioned earlier. Finally, three additional participants attempted the primary experiment but were unable to complete the task within the 2-hr maximum time limit. Data from these individuals were not included in any of the analyses or in the demographic information listed earlier.

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

Menu structures and word lists. The key research question was whether listeners could learn to associate cue sounds with TTS phrases and whether the rate of learning would differ for earcons and spearcons. Thus, participants were required to learn sound–word pair associations for two different types of lists.

Noun list. The noun list was the same as that used in Experiments 1 and 2. This list was used to study performance with brief, single-word menu items that were related within a menu (e.g., all animals) but not necessarily across menus. The identical words were used in an effort to extend the previous experiments.

Cell phone list. The cell phone list was added to begin to study performance in actual menu structures found in technology. This list involved words that were taken from menus found in the interface for the Nokia N91 mobile phone (http://www.nokia.com/nseries/index.html?loc = inside,main_n91). As can be seen in Table 4, these words and phrases tended to be relatively longer and also were obviously technological in context. As discussed previously, most of these items do not have natural sounds associated with them, so auditory icons were not a feasible cue type for this experiment and were not included here.

Auditory stimuli. The auditory stimuli included earcon or spearcon cues and TTS phrases, generated from the two word lists already described. During training, when listeners were learning the pairings of cues to TTS phrases, the TTS was followed by the cue sound.

Text to speech. All TTS phrases of the word lists were created specifically for this experiment using the AT&T Labs, Inc. TTS Demo program. Each word or text phrase was submitted separately to the TTS demo program via an

TABLE 4: Menu Structure Used for the Cell Phone List Condition for Experiment 3

Text Message	Messaging	Image Settings	Settings	Calendar
Add recipient	New message	Image quality	Multimedia message	Open
Insert	Inbox	Show captured image	Email	Month view
Sending options	Mailbox	Image resolution	Service message	To do view
Message details	My folders	Default image name	Cell broadcast	Go to date
Help	Drafts	Memory in use	Other	New entry

Note. Items were taken from existing menus on Nokia N91 mobile phones.

online form, and the resulting .WAV file was saved for incorporation into the experiment.

Earcons. As discussed, the noun list words (see Table 1) came from Experiments 1 and 2. The original 30 earcons from Experiments 1 and 2 were used again here as cues for the noun list.

For the cell phone list (see Table 4), 30 new hierarchical earcon cues were created using Audacity software. Each menu (i.e., column in Table 4) was represented with sounds of a particular timbre. Within each menu category (column), each earcon started with a continuous tone of a unique timbre, followed by a percussive element that represented each item (row) in that category. In other words, the top item in each column in the menu structure was represented by the unique tone representing that column alone, and each of that column's subsequent row earcons comprised that same tone, followed by a unique percussive element that was the same for every item in that row.

Earcons used in the noun list were an average of 1.26 s in length, and those used in the cell phone list were on average 1.77 s long.

Spearcons. The spearcons in this study were created by compressing the TTS phrases that were generated from the word lists. In Experiments 1 and 2, TTS items were compressed linearly by approximately 40% to 50%, while maintaining original pitch. That is, each spearcon was around half the length of the original TTS phrase. Although it is a simple algorithm, our experience has shown that this approach can result in very short (one word) phrases being cut down too much (making them into "clicks," in some cases). In contrast, longer phrases remain too long. Therefore, for Experiments 3, 4, and 5, TTS phrases were compressed *logarithmically*, still maintaining constant pitch. By logarithmical compression, the longer words and phrases were compressed to a relatively greater extent than those of shorter words and phrases. This type of compression also decreased the amount of variation in the length of the average spearcon because the length of the file will be inversely proportional to the amount of compression applied to the file.

Spearcons used in the noun list were an average of 0.28 s in length, and those used in the cell phone list were on average 0.34 s long.

Procedure: Main experiment. The participants were trained on the entire list of 30 words in a particular list type condition by presenting each TTS phrase just before its associated cue sound (earcon or spearcon). During this training phase, the TTS words were presented in menu order (top to bottom, left to right). After listening to all 30 TTS + cue pairs, participants were tested on their knowledge of the words that were presented. Each auditory cue was presented in random order, and after each a screen was presented displaying all of the words that were paired with sounds during the training in the grids illustrated in Tables 1 and 4. The participant was instructed to click the menu item that corresponded to the cue sound that was just played to him or her. Feedback was provided indicating a correct or incorrect answer on each trial. If the answer was incorrect, the participant was played the correct TTS + cue pair to reinforce learning. The number of correct and incorrect answers was recorded. When all 30 words had been tested, if any responses were incorrect, the participant was "retrained" on all 30 words and retested. This process continued until the participant received a perfect score on the test for that list. Next, the participant was presented with the same training process, but for the other list type. The procedure for the second list type was the same as for the first. The order of list presentation to the participant was counterbalanced.

After the testing process was complete, participants filled out a demographic questionnaire about age, ethnicity, and musical experience. They also completed a separate questionnaire pertaining to their experience with the experiment such as how long it took them to recognize the sound patterns and how difficult they considered the task to be on a 6-point Likert-type scale.

Procedure: Follow-up spearcon analysis experiment. Spearcons are always made from speech sounds. Most spearcons are heard by listeners to be nonspeech squeaks and chirps. However, some spearcons are heard by some listeners as very fast words (that is, after all, what they are made from). It is important to remember that it does not matter whether a given spearcon is heard as speech or nonspeech, but it is still interesting to examine the details of this new audio cue type. To this end, an additional exploratory study was completed in conjunction with the main experiment. After completing the main experiment, five participants assigned to the spearcon condition were also asked to complete a recall test of the spearcons they had just learned in the main experiment. For this, a program in Macromedia Director played each of the 60 spearcons (but not the TTS phrase) from the main experiment one at a time randomly to the participant. After each spearcon was played, the participants were asked to type in a field what word or phrase they thought the spearcon represented. We also asked six naïve users (new individuals who had had no exposure to the main experiment in any way) to complete this same follow-up experiment. These six naïve listeners would presumably allow us to determine which spearcons were more immediately "recognizable" as spoken words. Note that all participants were informed on an introduction screen that spearcons were compressed speech to control for any possible misinterpretation of the origin of the sounds.

Results of Experiment 3

Main experiment of learning rates. A 2 × 2 mixed design repeated measures ANOVA was completed on the number of training blocks required for 100% accuracy on the recall test. The first independent variable was a between-subjects measure of cue type (earcons vs. spearcons), and the second independent variable was a within-subjects manipulation of list type (noun list vs. cell phone list). The means and standard deviations of numbers of trial blocks for each of the four conditions are shown in Table 5 and illustrated in Figure 5. Overall, spearcons led to faster learning than earcons, as supported by the main effect of cue type, $F(1, 22) = 42.115$, $p < .001$, $\eta_p^2 = .66$. It is also relevant to mention that the three individuals who were unable to complete the experiment in the time allowed (2 hr), and whose data are not included in the results reported here, were all assigned to the earcons group. This suggests that even larger differences would have been found between earcons and spearcons if those data had been included.

Overall, the noun list was easier to learn than the cell phone list, as evidenced by the main effect of list type, $F(1, 22) = 7.086$, $p = .014$, $\eta_p^2 = .24$. This main effect was moderated by a significant interaction of cue type and list type, in which the noun list was learned more easily than the cell phone list for the earcon cues (Figure 5, white bars), but there was no difference in list type learning in the spearcons condition (Figure 5, gray bars). Interpreting this interaction is difficult with the results available here because it may be attributed to a floor effect apparent for results in the spearcons condition.

Debriefing and follow-up study results. Debriefing questions included a 6-point Likert-type scale (1 = *very difficult*, 6 = *very easy*) on which participants were requested to rate the difficulty of the task they had completed. Participants reported that the earcons task (M = 2.91, SD = 0.831) was significantly more difficult than the same task using spearcons (M = 5.25, SD = 0.452), $t(21) = –8.492$, $p < .001$.

TABLE 5: Number of Training Blocks Necessary to Obtain a Perfect Recall Score, for Each of the Four Conditions for Experiment 3

Condition	*M*	*SD*
Spearcons: Cell phone list	1.08	0.28
Spearcons: Noun list	1.08	0.28
Earcons: Cell phone list	6.55	3.30
Earcons: Noun list	4.55	2.25

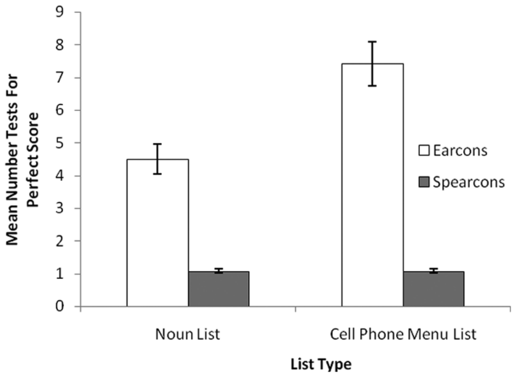


Figure 5. Mean number of trials necessary for participants to obtain perfect score on sound recall for both earcons and spearcons for noun and cell phone lists for Experiment 3. Error bars indicate standard error of the mean.

Finally, the spearcons analysis of the follow-up experiment data revealed that the training that the participants received on the word–spearcons associations in these lists led to greater comprehension. Out of a possible 60 points, the mean performance of individuals who had completed the spearcons condition in the main experiment before the spearcons recall test (M = 59.0, SD = 1.732) was significantly better than that for naïve users (M = 38.50, SD = 3.782), $t(9) = –11.115$, $p < .001$). No significant main effect was found for list type in the follow-up experiment.

Discussion of Experiment 3

The difference in means between auditory cue types was as expected, as spearcons clearly outpaced earcons in learning rates. The effect of list type, however, was not as expected. Since earcons do not provide cues to the word itself and need to be trained for associations to items on a menu to exist, it was not expected that the actual words included in a menu would make a difference. In contrast, the spearcons conditions were expected to lead to a significant difference between the two list types, mainly because of the increased contextual information provided by spearcons because they are created directly from the word that they represent.

The earcons condition with the noun list showed faster learning rate than that with the cell phone list because the nature of the earcons used in the noun list might be inherently easier to remember because of the particular sounds used. The lack of significant difference in list type for the spearcons condition may also have been a result of the floor effect apparent in the results. If the learning rates had not turned out as fast on average, we may very well have seen more variability in the spearcons condition, and the interaction might not have been significant. In general, however, these results, combined with the participants' perceptions that learning the spearcons task was significantly easier than for the same task with earcons and the findings that spearcons used in this study indeed were more recognizable on the whole after training, all provide strong empirical evidence of the superior nature of spearcons for use in auditory menus. It is feasible that the time to reach a menu item, once learned, will be much less with menus using spearcons than earcons, and therefore spearcons will provide a faster, less frustrating user experience.

EXPERIMENT 4

In Experiment 4, we sought to extend the results of Experiment 3 to generalize the benefit of spearcons in learnability. Accordingly, more diverse natural environmental sounds were involved in Experiment 4.

Awareness of features and objects in the world around us is vital in many aspects of life. Their importance affects all aspects of life, ranging from our safety and ability to travel to helping determine our comfort and productivity levels. Landmarks are crucial to navigation, aiding individuals to determine where they are and to plot a course toward a desired destination. Failure to avoid an object as a driver or pedestrian could spell disaster. We often rely on vision to make salient these aspects of our environment, but sometimes this is not preferable, or even possible. In these instances, auditory cues can be an effective alternative and have been incorporated into the SWAN navigation system (Wilson, Walker, Lindsay, Cambias, & Dellaert, 2007).

When devising an auditory display scheme for environmental features and objects, one key consideration must be how learnable the scheme is. In some situations, users might not interact with a display that is difficult to learn enough to understand it well. Even in usage scenarios where extended learning time did exist, users might not wish to invest the time in doing so. In light of this, Experiment 4 was designed to investigate the relative learnability of different auditory displays of environmental features surrounding a listener. To this end, in addition to traditional auditory cue types such as earcons and auditory icons, Experiment 4 included and examined more novel approaches such as certain combinations of auditory icons and earcons.

Method

Participants. For extra credit in psychology courses, 39 undergraduate students (25 male, 14 female, mean age = 20; auditory icons n = 6, earcons n = 6, earcon–icon hybrids n = 7, sized hybrids n = 7, TTS n = 6, spearcons n = 7) with normal or corrected to normal hearing and vision participated in Experiment 4.

Apparatus and equipment. A computer running Windows XP, with an external Creative Soundblaster Extigy sound card, was used for sound production. Participants listened to auditory stimuli using Sennheiser HD 202 headphones. The software used in this experiment was created for that purpose, using a Flash-based front end for the experiment interface and a Java-based server applet for data logging.

Menu structure. A total of 18 common environmental features were selected from the area outside a campus building. The features chosen are common in many urban environments and not (with one exception) unique to the location they were drawn from. Each feature was then classified into a high level category and a size category and by whether it directly produces a sound or not (see Table 6 for a list of the features as well as their classifications). Two of the features, stairs and curb cuts, have both an up and a down version, for a total of 20 features.

There were six high level categories, which were chosen based on the perspective of a visually impaired pedestrian: building/area, intersection helpers, obstacles, plants, usable objects, and landmarks. Buildings indicated large structures that an individual could enter. Intersection helpers were features that are useful when attempting to cross the street at an intersection. Features that would not be used and that would need to be avoided by a visually impaired pedestrian were classified as obstacles. All vegetation was classified as plants. Features in the environment a visually impaired pedestrian might need to interact with were designated as useful objects. The landmarks category comprised distinctive features that could aid in navigation. The “landmark” feature in this category referred to a unique historical site on campus. The category classifications of direct sound and size are self-evident. Six sounds were then constructed for each feature, one for each auditory cue design to be tested: auditory icons, earcons, TTS, spearcons, earcon–icon hybrids, and sized hybrids. The sounds ranged in duration from approximately 0.25 s to 4 s.

Auditory stimuli: Auditory icons. In building the auditory icons, the initial focus was the object and its natural sound. Since most of the identified objects, such as streetlights or crosswalks, do not emit any kind of natural sounds, an indirect auditory representation was needed. As an example, a tree is represented by the sound of the wind going through the leaves mixed with the sound of bending wood. In some cases there were no natural sounds that could be used as a representation (e.g., a crosswalk or a street light). In these cases musical instruments or the sound of the materials these objects were made of were used. The sounds were gathered from a comprehensive sound effects library. In most cases, various sound files were mixed together to achieve the desired icon. Hints for category allocation are not included into the auditory icon sounds. Thus, each sound stands for a specific object and comprises neither a category

TABLE 6: Environmental Features Used in Experiment 4, as Well as Their Classification by Category, Sound Production, and Size

Feature	Category	Direct Sound	Size
Public building	Building	No	Large/huge
Pedestrian light	Intersection aids	No	Medium
Crosswalk		No	Large
Curb cut (up and down)		No	Medium
Street light/sign	Obstacles	No	Medium
Fire hydrant		No	Small
Parking meter		No	Small
Road work		Yes	Large
Tree	Plants	No	Medium
Bush		No	Small
Bench	Usable objects	No	Medium
Public phone		Yes	Medium
Emergency phone		Yes	Medium
Garbage can		No	Medium
Stairs (up and down)		No	Medium
Bus stop		No	Medium
Fountain	Landmarks	Yes	Large
Landmark		No	Medium/large

teaser nor a size allocation. An auditory icon is simply the most natural representation of an object we could create. They are mostly short and straightforward and without additional object information.

Auditory stimuli: Earcons. As mentioned previously, earcons are musical patterns that can be decomposed into five dimensions: rhythm, pitch, timbre, register, and dynamics. Because of their characteristics to build hierarchies, in the design of the earcons, we included the object categorizations. Thus, each earcon started with an opening sound that represented the category to which the sound belonged. We used distinctive instruments for each object category:

- Buildings: Whirly keyboard
- Intersection helpers: Dings and dongs, mallets
- Obstacles: Grand piano
- Plants: Drums and percussion sounds
- Practical objects: Flute
- Landmarks: Organ

After the category sound, the actual object sound began. Each object was represented by a unique melody or rhythm. Since the chosen instruments and melodies were more or less arbitrary, we tried to choose the instruments to be an appropriate representation of the according category. For example, plants were assigned naturalistic percussion sounds like wood blocks. Natural mappings were also considered when designing the single melody. For example, two feature sounds were used for stairs. The melody displays the direction of the stairs in terms of an increasing or decreasing melody. Apple's (2007) GarageBand software was used to compose the teasers as well as the melody sounds.

Auditory stimuli: Earcon–icon hybrids. Because earcons are more or less arbitrary, their learnability often suffers. On the other hand, each auditory icon is distinct and bears no categorical resemblance to other related icons. To use the strengths of each to overcome the weaknesses of the other, earcon–icon hybrids were developed by combining the opening sound of each object category from the earcon and the auditory icon of a specific object. Thus, each feature consisted of an opening sound according to the category it belonged to, plus a unique icon sound.

Auditory stimuli: Sized hybrids. To give an impression of the size of an object, a sound layer containing size information was added to the earcon–icon hybrid sounds. A size classification with four steps was introduced: small, medium, large, and huge. For each size category, a unique melody was composed differing in pitch and duration. For instance, the sound representing huge objects was low pitched and long; in contrast, for small features a short and high-pitched two-note melody was used. Because the category teaser and the object sound are sequentially arranged, we considered adding the size sound at the end of the icon sound. However, to keep the sounds shorter, the size sound was played in parallel with the actual auditory icon. The size sounds were designed using frequencies such that they would not interfere with the actual object sound. The resulting sounds were checked to ensure that no masking effects took place.

Auditory stimuli: TTS. The same online AT&T Labs, Inc. TTS Demo program used in Experiment 3 was used to create the entire set of speech-based feature sounds.

Auditory stimuli: Spearcons. To create the spearcons, the speech stimuli were compressed using the same logarithmic algorithm coded in MATLAB, as described in Experiment 3.

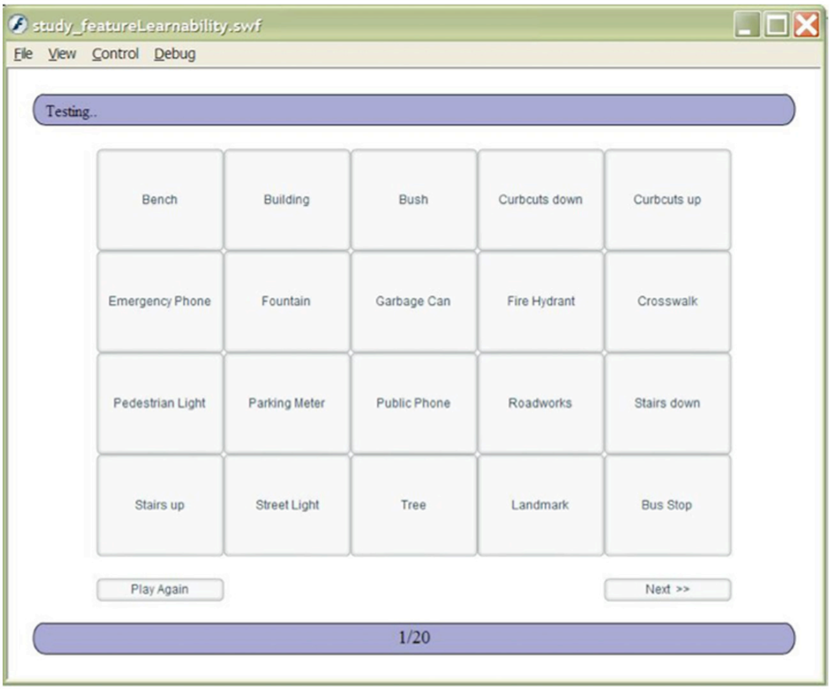
Procedure. After participants' informed consent was obtained, their demographics were recorded and they were randomly assigned to one of the six sound conditions. Participants were given instructions and then began the experiment.

In the training phase of the experiment, participants were shown a single target word (e.g., *bench*) and the sound associated with that environmental feature was played once. Participants would then advance the program to see the next feature and hear its associated sound. After being trained on all 20 stimuli, the testing phase would begin. Participants were presented with a grid containing all of the features presented in the training phase (see Figure 6). A sound from the training phase was then played, and participants were asked to select the environmental feature associated with that sound from the grid by clicking on it with the mouse. Participants were given the option to listen to a sound as often as they liked before making a selection by clicking a “Play Again” button. Once they had made their final selection, they clicked on the “Next” button and the next sound was played. At the end of the testing phase, after having been presented with all 20 stimuli, participants were shown their performance (e.g., 12/20). If a participant had not answered all 20 items correctly, the training phase was started again, after which another testing phase occurred. This process was repeated until a participant had successfully identified all 20 features correctly in a single testing phase. All answers given by participants were recorded by the software, which also noted the aggregate percentage correct of a given participant across all testing phases as well as how many training cycles were required to reach perfect performance.

Results of Experiment 4

The independent variable of sound type was analyzed with respect to the dependent variables of (a) the number of training cycles required to reach 100% accuracy and (b) the aggregate percentage accuracy of a participant across all testing cycles. A repeated measures MANOVA found a significant effect of sound type, $F(10, 64) = 9.66$, $p < .001$, Wilks's Lambda = .159, and both dependent measures similarly contributed to the significant effect. Subsequent repeated measures ANOVAs showed a significant effect of sound type for both the number of training cycles, $F(5, 33) = 10.77$, $p < .001$, and aggregate percentage accuracy, $F(5, 33) = 20.15$, $p < .001$.

In terms of the number of training cycles necessary to achieve 100% accuracy, the spearcons and TTS sound types clearly required the smallest number of cycles ($M = 1.14$, $SD = 0.378$ and $M = 1.14$, $SD = 0.378$, respectively), which can be seen in Figure 7. Pairwise comparisons confirmed both sound types to require significantly fewer trials compared to all other sound types. Earcons required the largest number of cycles ($M = 8.50$, $SD = 4.087$). Pairwise



Bench	Building	Bush	Curbcuts down	Curbcuts up
Emergency Phone	Fountain	Garbage Can	Fire Hydrant	Crosswalk
Pedestrian Light	Parking Meter	Public Phone	Roadworks	Stairs down
Stairs up	Street Light	Tree	Landmark	Bus Stop

Figure 6. The grid that participants used to select an answer during the testing phase for Experiment 4. Clicking the "Play Again" button in the lower-left corner allowed them to hear a sound as many times as they liked. The "Next" button in the lower-right corner indicated their answer choice was final.

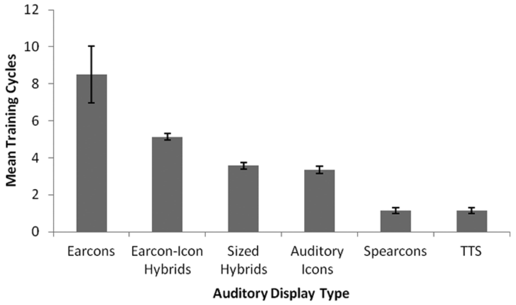


Figure 7. Mean number of training cycles needed to reach 100% accuracy in a testing phase for Experiment 4. Error bars indicate standard error of the mean.

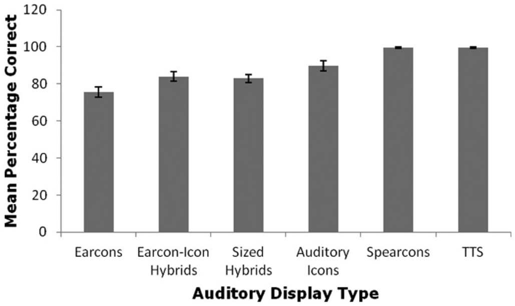


Figure 8. Mean percentage accuracy of participants with each type of auditory display across all trials for Experiment 4. Error bars indicate standard error of the mean.

comparisons determined this to be significantly more than all other sound types except for earcon–icon hybrids. The inclusion of a "size" attribute to the sounds led to no statistically significantly different performance between earcon–icon hybrids and sized hybrids.

The aggregate percentage accuracy also showed spearcons and TTS to be identical to each other ($M = 99.64\%$, $SD = 0.945$ and $M = 99.64\%$, $SD = 0.945$, respectively), which can be seen in Figure 8. Pairwise comparisons revealed both spearcons and TTS to have a significantly higher aggregate accuracy compared to the other sound types. Earcons, on the other hand, had a significantly worse aggregate accuracy than any other sound type except for sized hybrids as indicated by pairwise comparisons. No statistically significant difference was found between the earcon–icon hybrids and the sized hybrids.

Discussion of Experiment 4

The principal finding of this experiment is that spearcons are as easy to learn as TTS, for environmental feature sounds. Performance with both dependent measures was identical, with almost no errors across any trials and very few participants taking more than one cycle to identify all the feature sounds correctly. This seems to indicate that spearcons, like TTS, require virtually no learning to comprehend. In addition, spearcons are faster than TTS and hence do not occupy as much of the display time as speech. Spearcons are also not actual speech, which presumably allows the verbal channel to be left unimpeded while they are being used (see, e.g., Wickens, 2002). Taking into account all of these advantages, it is clear that spearcons are a distinct and useful auditory display technique. Even though each condition of this experiment had a relatively small number of participants (6 or 7), the results were promising and consistent with the previous results. Recruiting more participants would be expected to yield more statistical power, but not likely to change the overall conclusions.

Another interesting finding is in the results of the two novel sound types, earcon–icon hybrids and sized hybrids. In terms of both dependent measures, combining earcons and auditory icons led to better performance than earcons alone. This increased learning performance is likely a result of the familiarity that the auditory icons lend to the sounds. However, both earcon–icon hybrids and sized hybrids showed worse learning performance than auditory icons alone. This is possibly a result of the fact that these two new sound types are much more complex than the auditory icons and therefore possibly more difficult to learn. Although these two sound types do allow the hierarchical structuring of auditory icons, the overshadowing performance of spearcons and TTS makes those far more appealing options when interface learnability is a concern.

In brief, spearcons have once again proven to be comparable to speech with respect to learnability. At the same time, they are different enough to leave the speech channel open and are briefer and therefore occupy less display time. This reinforces their potential as an excellent auditory display methodology. Moreover, although fusing auditory icons and earcons does allow for a combination of some of their strengths, it also dilutes the learnability of the auditory icons, which is one of their principal advantages.

EXPERIMENT 5

In Experiments 1 and 2, both spearcons and TTS led to faster and more accurate menu navigation than auditory icons and hierarchical earcons. In Experiments 3 and 4, spearcons also showed a better learning rate than other traditional or newer types of auditory cues, and the learnability of spearcons is comparable to speech.

Here, we need to look back at the aim of adding this class of nonspeech sounds to speech menus. The goal of these cues is to improve performance of speech interfaces. Therefore, it is not enough to show that spearcons facilitate learnability and efficiency of auditory interfaces as much as plain speech does. If adding spearcons does not outperform speech-only menus, we do not need to add spearcons to the real device no matter how easy the implementation of spearcons is.

Experiment 5 was designed to determine if navigation efficiency would be enhanced when using prepended spearcons on realistic two-dimensional auditory menus compared to plain TTS. This experiment's hypothesis was that the speed of navigation would be faster when the menu items were prepended with spearcons than when using only TTS, even though the spearcon-enhanced cues were longer than plain TTS.

Method

Participants. A total of 28 undergraduates (9 male, 19 female, mean age = 19.14) with normal or corrected to normal hearing and vision participated for extra credit in psychology

TABLE 7: Visual Representation of the Auditory Menu Navigated by Participants in Each Condition for Experiment 5

	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	

Note. The left column shows the level number corresponding to each item in the row of the menu.

courses. English was the native language of all participants.

Apparatus and equipment. Participants were tested with a computer program written with Macromedia Director MX and Lingo on the Windows XP computer used in Experiment 4, listening through Sennheiser HD 202 headphones.

Menu structure. 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, and 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 7 lists the items included in the menu structure.

Auditory stimuli: TTS and Spearcons. TTS files were generated for all 44 items using the same AT&T Labs TTS Demo program and were converted to spearcons using the same logarithmic algorithm from previous experiments. For the Spearcon + TTS condition, the spearcons were prepended to the TTS with 250 ms between the two sounds. No visual menu was needed for this experiment, except for the screens that provided instructions to, and collected information from, the participants.

Procedure. A between-subjects design with two conditions was used. The independent variable was sound type (TTS only, Spearcon + TTS), and the dependent variable was average time in milliseconds to select the requested target item. There were 14 participants in each condition.

Participants were presented with 10 blocks of 22 trials each. Two stimulus lists were created from the original 44 items, and each list was alternated throughout the 10 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 screen that explained how to navigate the auditory menu using the keyboard. They were instructed that their task was to find the target item as quickly as possible without sacrificing accuracy (e.g., “Find *Genres* on the *Music* menu”). Each trial in a block was followed immediately by the next trial, but the

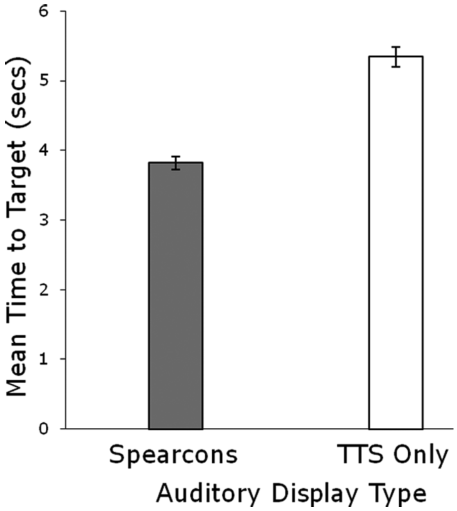


Figure 9. Mean time to target (ms) for navigating auditory menus with TTS-only versus TTS menu items with spearcon enhancements for Experiment 5. Participants in spearcons condition performed significantly better than those in the TTS-only condition. Error bars indicate standard error of the mean. TTS = text to speech.

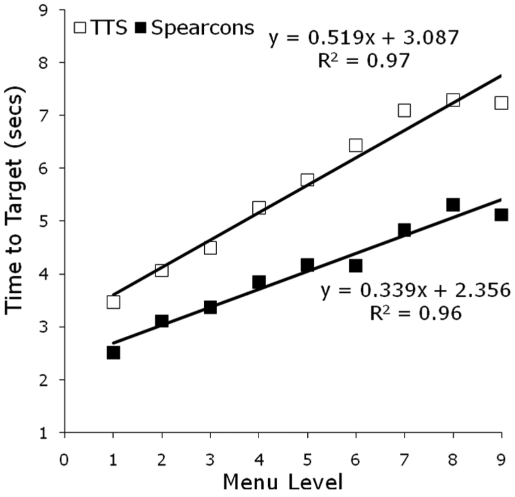


Figure 10. Mean time to target (ms) as a function of menu level for Experiment 5. Spearcons led to faster performance at all menu depths, and there was a lower per-item cost for spearcons-enhanced items as depth in the menu increased. Error bars indicate standard error of the mean.

participants could control the start of each new block. After completing the 10th block of trials, participants filled out a brief questionnaire and were debriefed.

Results of Experiment 5

Error trials, arising from incorrect item selection, were removed from analyses; this meant 1.10% trials (26 in Spearcons, 42 in TTS only) were eliminated. One outlier was also eliminated because of an extreme time to target. On 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 6,092 trials remained. Because there was no salient difference in accuracy just as in previous experiments, we focused here on the analysis of navigation time.

Figure 9 shows the mean time to target for each condition. 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 = 3.82 s, SD = 3.917) than for those in the TTS-only condition (M = 5.34 s, SD = 3918), $t(6089) = 17.89, p < .001$.

Because of the significant difference in navigation time between the two conditions, further analysis was performed based on the level of the item on the menu. The number on the left-hand side of the menu structure on Table 7 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 6 through 9, the number of menu categories having each of the levels decreased until Level 9, 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 10). The slope of the TTS condition (slope = 0.519) was significantly steeper than for the spearcons condition (slope = 0.339; $z = 5.064$, $p < .05$).

Discussion of Experiment 5

In Experiment 5, we found that spearcons improved navigation speed significantly when compared to plain TTS in the realistic auditory menu system. 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.

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 in 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 enhancements result in such improvements in navigational speed.

GENERAL DISCUSSION

As auditory menu-based interfaces become more important and more common, it is crucial to improve their usability, effectiveness, speed, and accuracy. In this article, to compensate for traditional auditory enhancements such as auditory icons and earcons, a newer menu-item-level enhancement technique called spearcons—speech based earcons—has been introduced and systematically evaluated. Spearcons led to significantly better navigation efficiency and accuracy than either auditory icons or earcons (Experiments 1 and 2). This performance benefit of spearcons comes from the lower per-item cost in menu navigation behavior. Also, spearcons demonstrated better learning rates than traditional auditory cues and newer hybrid ones (Experiments 3 and 4). Finally, we obtained the key result that adding spearcons led to better performance than the plain TTS menu in a realistic menu navigation, even though adding spearcons makes menu items longer (Experiment 5).

From a practical standpoint, the support for spearcons as a preferred auditory cue for menu enhancement is fourfold. First, spearcons are very easy to create, so it is feasible that they could be created on the fly, to increase ease of use in any language or application. Second, using spearcons does not restrict the structure of a menu system. Their use in a menu hierarchy can be as fluid as necessary because they do not require fixed indications of menu position. For this reason, they also can be considered a strong candidate for any imaginable menu system, not just for the standard hierarchical menu common in today's applications. Third, research demonstrated that spearcons are very easy to learn and thus will minimize frustration and training time for new users. Finally, spearcons are shorter in length than other traditional auditory cues. Moreover, despite their short length, because spearcons have reminiscences of the original words, users can listen to TTS speech phrase less than in other auditory displays (Experiment 2). Consequently, spearcons are poised to provide greater efficiency for users of electronic menus.

The use of small electronic devices is increasing and becoming more integrated into our lives on a daily basis. These devices are becoming essential not only for business use but also for communication and information seeking in countless occupations. It is essential that these devices be accessible to all who could benefit from them, including those who rely on auditory cues exclusively, such as the blind and those with temporarily obstructed vision, such as firefighters and soldiers. The ability to use these devices with minimum frustration and efficient rates of learning will stem directly from the characteristics of the auditory cues that are provided by these devices. Spearcons are clearly capable of fulfilling these needs. Thus, implementing spearcons in mobile device menus, in telephone-based interfaces for banks and airlines, and in screen-reader software such as JAWS could lead to a much richer and more effective user experience, with relatively little effort on the part of the developer.

The fact that spearcons are nonarbitrary (which has been discussed here as a benefit) might lead to one possible downside: Spearcons are language dependent, whereas earcons are not. That is, if an interface were translated from, say, English to Spanish, then the spearcons would be different in the two interfaces, whereas an earcon hierarchy would not be different. In some situations this could be problematic. On the other hand, the spearcons can be regenerated automatically, so there is no extra work involved in "internationalizing" an auditory menu with spearcons. Also, Spanish-based spearcons actually sound distinct from English-based spearcons, which is appropriate.

Planned research includes replicating this study with participants who are visually impaired or blind. These studies will provide data from the demographic of participants likely to benefit the most from enhanced navigational efficiency on auditory menus. Furthermore, we can assess the potential of spearcons by leveraging their advantages within speech recognition systems (e.g., Gardner-Bonneau, 1992; Polkosky & Lewis, 2001) and automotive user interfaces (e.g., Jeon, Davison, Nees, Wilson, & Walker, 2009; Vargas & Anderson, 2003). Also, diverse combinations of nonspeech sounds (e.g., auditory scroll bars + spearcons + TTS) can be examined.

In conclusion, the use of spearcons might allow modern menu interfaces to remain "intelligent," while still incorporating audio cues that are as flexible and dynamic as the interface itself. Spearcons enhance both the system effectiveness and the user's interaction with the system, which is an important joint outcome in the field of human–computer interaction, especially in novel and less well-studied interfaces such as auditory menus.

ACKNOWLEDGMENTS

This article represents a compilation of results collected over the past few years. As is typical in such a program of research, some of the results have been discussed in preliminary form at conferences over the course of the project; however, considerable new data, analyses, and discussion have been added. The data from Experiment 1 were discussed at ICAD 2006 (Walker, Nance, & Lindsay, 2006). The data for Experiment 2 are entirely novel and have never been presented or published before. It is a follow-up study of Experiment 1. Therefore, discussion for the combined Experiments 1 and 2 is also totally new for this article. The data from Experiment 3 were presented in abbreviated form at ICAD 2007 (Palladino & Walker, 2007), and Experiment 4 was presented in abbreviated form at ICAD 2008 (Dingler, Lindsay, & Walker, 2008). Finally, Experiment 5 was discussed in abbreviated form at HFES 2008 (Palladino & Walker, 2008b). Overall, detailed information on making spearcons and other auditory cues and much more substantial and integrative discussion were added to the present treatment. This is in addition to the totally new data contained in this article. Portions of this research were funded by a grant from the Rehabilitation Engineering Research Center for Wireless Technologies (WirelessRERC), funded by the National Institute on Disability and Rehabilitation Research (NIDRR) of the U.S. Department of Education (Grant H133E060061).

KEY POINTS

- Interfaces for electronic devices often have a menu structure.
- Auditory menus are useful when users cannot look at or cannot see a visual menu.
- Improving utility and usability of auditory menus remains a challenge.
- Spearcons are a novel class of interface sounds that can enhance auditory menus.
- A series of five experiments demonstrate that spearcon-enhanced auditory menus are faster and more accurate to use, easier to learn, more appropriate, and preferred over other kinds of auditory menus.

REFERENCES

Absar, R., & Guastavino, C. (2008). Usability of non-speech sounds in user interfaces. In Susini, P. & Warusfel, O., *Proceedings of the International Conference on Auditory Display (ICAD2008)*. Paris, France: IRCAM.

Apple. (2007). *GarageBand* [MIDI-based software]. Cupertino, CA: Author.

Asakawa, C., & Itoh, T. (1998). User interface of a home page reader. In *Proceedings of the Annual ACM Conference on Assistive Technologies (ASSETS98)* (pp. 149–156). Marina del Rey, CA.

Asakawa, C., Takagi, H., Ino, S., & Ifukube, T. (2003). Maximum listening speeds for the blind. In Brazil, E. & Shinn-Cunningham, B., *Proceedings of the International Conference on Auditory Display (ICAD2003)* (pp. 276–279). Boston, MA: Boston University Publications.

AT&T Research Labs. (n.d.). *AT&T text-to-speech demo*. Retrieved from http://www2.research.att.com/~ttsweb/tts/demo.php.

Bederson, B. B. (2000). Fisheye menus. In Ackerman, M. & Edwards, K., *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST'00)* (pp. 217–225). New York, NY: ACM.

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. (1997). Navigating telephone-based interfaces with earcons. In Thimbleby, H. W., O'Conaill, B., & Thomas, P., *Proceedings of the BCS HCI'97* (pp. 39–56). Bristol, UK.

Brewster, S. (1998). Using non-speech sounds to provide navigation cues. *ACM Transactions on Computer-Human Interaction, 5*, 224–259.

Brewster, S., & Crease, M. G. (1999). Correcting menu usability problems with sound. *Behaviour and Information Technology, 18*, 165–177.

Brewster, S., Raty, V.-P., & Kortekangas, A. (1996). Earcons as a method of providing navigational cues in a menu hierarchy. In Sasse, M. A., Cunningham, J., & Winder, R. L., *Proceedings of the BCS HCI'96* (pp. 169–183). London, UK: Imperial College.

Brewster, S., Wright, P. C., & Edwards, A. D. N. (1993). An evaluation of earcons for use in auditory human-computer interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI93)* (pp. 222–227). Amsterdam, Netherlands.

Cepstral Corp. (n.d.). *Cepstral text-to-speech*. Retrieved from http://www.cepstral.com

Dingler, T., Lindsay, J., & Walker, B. N. (2008). Learnability of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech. In Susini, P. & Warusfel, O., *Proceedings of the International Conference on Auditory Display (ICAD2008)*. Paris, France: IRCAM.

Edwards, A. D. N. (1989). Soundtrack: An auditory interface for blind users. *Human-Computer Interaction, 4*, 45–66.

Findlater, L., & McGrenere, J. (2004). A comparison of static, adaptive, and adaptable menus. In Dykstra-Erickson, E. & Tscheligi, M., *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'04)* (pp. 89–96). New York, NY: ACM.

Freedom Scientific. (n.d.). *JAWS for Windows*. Retrieved from http://www.freedomscientific.com/fs_products/software_jaws.asp

Gardner-Bonneau, D. J. (1992). Human factors problems in interactive voice response (IVR) applications: Do we need a guideline/ standard? In *Proceedings of the Human Factors Society 36th Annual Meeting (HFES1992)* (pp. 222–226). Santa Monica, CA: HFES. DOI: 10.1177/154193129203600101

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

Gaver, W. W. (1989). The SonicFinder, a prototype interface that uses auditory icons. *Human-Computer Interaction, 4*, 67–94.

Hejna, D. J., Jr. (1990). *Real-time time-scale modification of speech via the synchronized overlap-add algorithm* (Unpublished masters thesis). Massachusetts Institute of Technology, Cambridge, MA.

Jeon, M., Davison, B. K., Nees, M. A., Wilson, J., & Walker, B. N. (2009). Enhanced auditory menu cues improve dual task performance and are preferred with in-vehicle technologies. In Schmidt, A., Dey, A., Seder, T., Juhlin, O., & Kern, D., *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI09)* (pp. 91–98). New York, NY: ACM.

Jeon, M., & Walker, B. N. (2011). Spindex (Speech Index) improves auditory menu acceptance and navigation performance. *ACM Transactions on Accessible Computing, 3*, 10.

Karshmer, A., Brawner, P., & Reiswig, G. (1994). An experimental sound-based hierarchical menu navigation system for visually handicapped use of graphical user interfaces. In *Proceedings of the Annual ACM Conference on Assistive Technologies (ASSETS94)* (pp. 123–128). New York, NY: ACM.

LePlatre, G., & Brewster, S. (1998). Designing non-speech sounds to support navigation in mobile phone menus. In Cook, P. R., *Proceedings of the International Conference on Auditory Display (ICAD2000)* (pp. 190–199). Atlanta, GA: ICAD.

Moos, A., & Trouvain, J. (2007). Comprehension of ultra-fast speech—Blind vs. "normally hearing" persons. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 677–680). Saarbrücken, Germany: Saarland University.

Morley, S., Petrie, H., O'Neill, A. M., & McNally, P. (1998). Auditory navigation in hyperspace: Design and evaluation of a non-visual hypermedia system for blind users. In *Proceedings of the Annual ACM Conference on Assistive Technologies (ASSETS98)*. New York: ACM.

Mynatt, E. (1997). Transforming graphical interfaces into auditory interfaces for blind users. *Human-Computer Interaction, 12*, 7–45.

Mynatt, E., & Edwards, W. (1992). Mapping GUIs to auditory interfaces. In *Proceedings of the 5th Annual ACM Symposium on User Interface Software and Technology* (pp. 61–70). New York: ACM.

Mynatt, E., & Weber, G. (1994). Nonvisual presentation of graphical user interfaces: Contrasting two approaches. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI94)* (pp. 166–172). New York: ACM.

Norman, K. L. (1991). *The psychology of menu selection: Designing cognitive control of the human/computer interface.* Norwood, NJ: Ablex.

Palladino, D. K., & Walker, B. N. (2007). Learning rates for auditory menus enhanced with spearcons versus earcons. In Scavone, G. P., *Proceedings of the 13th International Conference on Auditory Display (ICAD2007)* (pp. 274–279). Montreal, Canada: ICAD.

Palladino, D. K., & Walker, B. N. (2008a). Efficiency of spearconenhanced navigation of one dimensional electronic menus. In Susini, P. & Warusfel, O., *Proceedings of the International Conference on Auditory Display (ICAD2008)*. Paris, France: IRCAM.

Palladino, D. K., & Walker, B. N. (2008b). Navigation efficiency of two dimensional auditory menus using spearcon enhancements. In *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society (HFES2008)* (pp. 1262–1266). Santa Monica, CA: HFES. doi: 10.1177/154193120805201823

Pitt, I. J., & Edwards, A. D. N. (1996). Improving the usability of speech-based interfaces for blind users. In *Proceedings of the Annual ACM Conference on Assistive Technologies (ASSETS96)* (pp. 124–130). New York, NY: ACM.

Polkosky, M. D., & Lewis, J. R. (2001). *The function of nonspeech audio in speech recognition applications: A review of the literature* (IBM Voice Systems Technical Report, TR 29.3405). West Palm Beach, FL: IBM.

Psychological Software Tools. (n.d.). *E-Prime*. Retrieved from http://www.pstnet.com

Raman, T. V. (1997). *Auditory user interfaces: Toward the speaking computer*. Boston, MA: Kluwer.

Roucos, S., & Wilgus, A. M. (1985). High quality time-scale modification for speech. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing* (pp. 493–496). New York, NY: IEEE. doi: 10.1109/ICASSP.1985.1168505

Sears, A., & Shneiderman, B. (1994). Split menus: Effectively using selection frequency to organize menus. *ACM Transactions on Computer-Human Interaction, 1*, 27–51.

Shneiderman, B. (1998). *Designing the user interface: Strategies for effective human-computer-interaction* (3rd ed.). Reading, MA: Addison-Wesley Longman.

Thatcher, J. (1994). Screen reader/2 access to OS/2 and the graphical user interface. In *Proceedings of the Annual ACM Conference on Assistive Technologies (ASSETS94)* (pp. 39–46). New York, NY: ACM.

Vargas, M. L. M., & Anderson, S. (2003). Combining speech and earcons to assist menu navigation. In Brazil, E. & Shinn-Cunningham, B., *Proceedings of the International Conference on Auditory Display (ICAD2003)* (pp. 38–46). Boston, MA: ICAD.

Walker, B. N., Nance, A., & Lindsay, J. (2006). Spearcons: Speechbased earcons improve navigation performance in auditory menus. In Stockman, T., Nickerson, L., Frauenberger, C., Edwards, A. D. N., & Brock, D., *Proceedings of the 12th International Conference on Auditory Display (ICAD2006)* (pp. 63-68). London, UK: ICAD.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*, 159–177.

Wilson, J., Walker, B. N., Lindsay, J., Cambias, C., & Dellaert, F. (2007). SWAN: System for wearable audio navigation. In *Proceedings of the 11th International Symposium on Wearable Computers (ISWC 2007)* (pp. 91–98). New York, NY: IEEE. doi: 10.1109/ISWC.2007.4373786

Wolf, C., Koved, L., & Kunzinger, E. (1995). Ubiquitous mail: Speech and graphical interfaces to an integrated voice/email mailbox. In Nordby, K., Helmersen, P. H., Gilmore, D. J., & Arnesen, S. A. *Proceedings of the IFIP Interact'95* (pp. 247–252). London: Chapman & Hall.

Yalla, P., & Walker, B. N. (2008). Advanced auditory menus: Design and evaluation of auditory scrollbars. In Harper, S. & Barreto, A., *Proceedings of the Annual ACM Conference on Assistive Technologies (ASSETS'08)* (pp. 105–112). New York, NY: ACM.

Zhao, S., Dragicevic, P., Chignell, M., Balakrishnan, R., & Baudisch, P. (2007). earPod: Eyes-free menu selection using touch input and reactive audio feedback. In Begole, B., Payne, S., Churchill, E., St. Amant, R., Gilmore, D., Rosson, M. B., *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI07)* (pp. 1395–1404). New York, NY: ACM.

Bruce N. Walker is associate professor in the School of Psychology and the School of Interactive Computing at the Georgia Institute of Technology. He received his PhD in psychology in 2001 from Rice University.

Jeffrey Lindsay received his MS in psychology from Georgia Tech and is a PhD candidate in the School of Psychology at Georgia Tech.

Amanda Nance received her MS in HCI from Georgia Tech and is senior usability analyst at Sage.

Yoko Nakano received her MS in HCI from Carnegie Mellon University and is user experience designer at Schematic.

Dianne K. Palladino received her BS in psychology from Georgia Tech and is a PhD candidate in the Social and Health Psychology program at Carnegie Mellon University.

Tilman Dingler is a graduate student of media computer science at the Ludwig Maximilians-University in Munich.

Myounghoon Jeon received his MS in psychology from Georgia Tech and is a PhD candidate in the School of Psychology at Georgia Tech.

Date received: January 18, 2011
Date accepted: April 18, 2012