Spindex (Speech Index) Improves Auditory Menu Acceptance and Navigation Performance

MYOUNGHOON JEON and BRUCE N. WALKER, Georgia Institute of Technology

Users interact with mobile devices through menus, which can include many items. Auditory menus have the potential to make those devices more accessible to a wide range of users. However, auditory menus are a relatively new concept, and there are few guidelines that describe how to design them. In this paper, we detail how visual menu concepts may be applied to auditory menus in order to help develop design guidelines. Specifically, we examine how to optimize the designs of a new contextual cue, called "spindex" (i.e., speech index). We developed and evaluated various design alternatives for spindex and iteratively refined the design with sighted users and visually impaired users. As a result, the "attenuated" spindex was the best in terms of preference as well as performance, across user groups. Nevertheless, sighted and visually impaired participants showed slightly different responses and feedback. Results are discussed in terms of acoustical theory, practical display design, and assistive technology design.

Categories and Subject Descriptors: H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Auditory (non-speech) feedback, evaluation/methodology, interaction styles (e.g., commands, menus, forms, direct manipulation), user-centered design, voice I/O*; J.4 [**Computer Application**]: Social and Behavioral Sciences—*Psychology*

General Terms: Design, Experimentation, Human Factors, Performance

Additional Key Words and Phrases: Auditory menus, spindex, assistive technology

ACM Reference Format:
Jeon, M. and Walker, B. N. 2011. Spindex (speech index) improves auditory menu acceptance and navigation performance. ACM Trans. Access. Comput. 3, 3, Article 10 (April 2011), 26 pages.
DOI = 10.1145/1952383.1952385 http://doi.acm.org/10.1145/1952383.1952385

1. INTRODUCTION

Although blindness remains a priority for accessibility researchers [Newell 2008], there are still relatively few widely deployed auditory interfaces. And, of course, not only visually impaired users, but sighted users as well can frequently benefit from eyes-free auditory displays, such as when dealing with small or nonexistent screens on mobile devices, especially when on the go (e.g., walking, cycling, driving, or with the device in a pocket). Often in the past, research on auditory interfaces has focused on speech interfaces involving desktop computer screen readers [Asakawa and Itoh 1998; Pitt and Edwards 1996; Thatcher 1994], audio HTML [James 1998; Morley et al. 1998], and online help systems [Kehoe and Pitt 2006]. There remain many questions regarding auditory displays and mobile devices.

Many modern electronic devices can have a very large menu structure (e.g., MP3 players with up to 30,000 songs). Speech (usually text-to-speech, TTS) is the most obvious means of providing users with audible feedback about menu navigation.

Author's address: B. N. Walker; email: bruce.walker@psych.gatech.edu.

DOI 10.1145/1952383.1952385 http://doi.acm.org/10.1145/1952383.1952385

Certainly, speech is excellent for accuracy and requires little learning. On the other hand, speech (even in fast talking screen readers) leads to the same problems as text in text-based systems, as this is also a serial medium [Brewster 2008]. Because of the temporal characteristics of speech, auditory scanning of speech-based menus is more difficult than that of visual menus, especially for very long menus. Moreover, modern visual menus often include much more than the text of the menu items. Divider lines, icons, shortcuts, scrollbars, and tabs are all common nontext menu elements that are difficult to convey using just speech.

Fairly recently, nonspeech auditory cues have been considered as a way to convey some of those nontext menu elements, compensating for the limitations of speech-only interfaces. Non-speech sounds may be used in addition to, or even instead of speech in an interface [Jeon et al. 2010]. Speech is still generally a required element when a user begins to interact with the device; however, with practice the speech sounds may no longer be necessary and the nonspeech sounds can play a more prominent role, thereby addressing many of the problems associated with speech-only interfaces.

Research on applications of nonspeech sounds in auditory menus has mainly included mobile phones [Brewster et al. 1998; Leplâtre and Brewster 2000; Palladino and Walker 2007, 2008a, 2008b; Vargas and Anderson 2003; Walker et al. 2006], PDAs [Brewster and Cryer 1999; Klante 2004], and wearable computers [Brewster et al. 2003; Wilson et al. 2007]. We can divide these nonspeech menu sounds into two basic categories: menu *item-level* cues; and menu *structure-level* cues. For example, auditory icons [Gaver 1986] and spearcons [Walker et al. 2006] belong to the item-level set; they focus on "what" an item is and provide one-to-one mapping between sound and meaning. In contrast, for menu structure-level cues, the focus is "where" the user is in the menu; the nonspeech sounds provide contextual information such as the menus' structure and size, and the user's location or status. For example, Blattner et al. [1989] showed that earcons can allow for good performance in hierarchical menu navigation with structural information. Auditory scrollbars also enhanced users' estimation of menu size and their relative location in a one-dimensional menu [Yalla and Walker 2008].

In a similar vein, the current project attempts to improve menu navigation with a new type of contextual sound cue, called a "spindex" [Jeon and Walker 2009]. Spindex cues help a user know where he or she is in an alphabetical list, and may be especially useful in navigating long lists such as the names of songs on an MP3 player.

2. BACKGROUND

2.1 Types of Nonspeech Auditory Cues

Before the spindex, there have been four main approaches to adding nonspeech sounds to the basic TTS used in most auditory menus. These all tend to include adding sound cues before or concurrent with the spoken menu items. The main types of enhancement cues are auditory icons, earcons, spearcons, and auditory scrollbars, as previously mentioned.

2.1.1 Auditory Icons. Auditory icons [Gaver 1986] are nonmusical sounds that have some resemblance (actual or analogous) to the objects, functions, and events they are representing. That is, an auditory icon representing a printer might sound like a dot-matrix printer or typewriter. Because of this analogic relationship between sounds and objects, auditory icons generally require little learning. Clearly, the level of direct resemblance between the auditory icon and the represented item can vary, just as with a visual icon. At some point, the direct iconic representation gives way to a metaphorical representation [Walker and Kramer 2004]. Whereas researchers have attempted to

convert GUIs on the desktop computer to sets of auditory icons [Gaver 1989; Mynatt 1997], there seem to be few examples of the addition of auditory icons leading to significantly better performance with auditory menu-based interfaces, in terms of navigation or menu item identification. Moreover, sometimes it is difficult to create a proper auditory icon for menu item such as "search" or "save as HTML" [Palladino and Walker 2007, 2008a].

2.1.2 Earcons. Earcons [Blattner et al. 1989] are nonspeech audio cues, typically composed of short, rhythmic sequences of musical notes with variable intensity, timbre, and register. Since earcons use an arbitrary mapping between the (musical) sound and the object they represent, they can be analogous to a language or a symbol. This arbitrary mapping between earcon and represented item means that earcons can be applied to any type of menu; that is, earcons can represent pretty much any concept. On the other hand, this flexibility provides a weakness because the arbitrary mapping of earcons requires training. Earcons can also be used to represent hierarchical menus by logically varying musical attributes. Investigators have designed auditory systems for visually impaired users to enable efficient navigation on the web or hypermedia using auditory icons and earcons [Goose and Moller 1999; Morley et al. 1998]. Those results improved usability and browsing experience. However, when a new item has to be inserted in a fixed hierarchy (e.g., adding a new name to a contact list menu), it can be difficult to create a new branch sound. This makes menus that use earcons "*brittle*" [Walker et al. 2006]. Finally, the musical nature of earcons makes them generally unsuitable for very long menus; you simply run out of notes when dealing with 30,000 songs on an MP3 player. Absar and Guastavino [2008] provide a recent overview of auditory icons and earcons.

2.1.3 Spearcons. Spearcons are brief sounds that are produced by speeding up spoken phrases, even to the point where the resulting sound is no longer comprehensible as speech [Walker et al. 2006]. These unique sounds are analogous to fingerprints because of the acoustic relation between the spearcons and the original speech phrases. Spearcons are easily created by converting the text of a menu item to speech via text-to-speech. This allows the system to cope with dynamically changing items in menus. For example, the spearcon for "Save" can be readily extended into the spearcon for "Save As." Or, if a new name is added to a contact list, the spearcon can be created as needed, even on the fly. Also, spearcons are easy to learn, whether they are comprehensible as a particular word or not, because they derive from the original speech [Palladino and Walker 2007]. Finally spearcons can enhance menus that are arbitrarily long.

2.1.4 Auditory Scrollbars. In visual menus sometimes a scrollbar is displayed to help convey contextual information. Scrollbars in menus are obviously used in many applications, and on many platforms. The location of the "thumb" of the scrollbar (the filled rectangle that moves up or down) conveys the user's location within the menu, while the size of the thumb conveys the size of the menu. If the thumb size is small, there are many items in the menu. If the thumb size is large, there are few items in the menu. An auditory scrollbar can be designed analogous to the visual scrollbar. Previously, this has been suggested for enhancing a visual scrollbar as found on a computer application [Brewster 1998]. That study focused on a continuous scrollbar (the thumb could be located anywhere along the bar). More recently, Yalla and Walker [2008] examined the possibility of the use of discrete auditory scrollbars (the thumb can only be located at discrete, designated points) in mobile device menus. In that context, the scrollbar is purely a "display," in that it is not used to actually move the cursor up or down in the

list (contrast this with the scrollbar on a desktop computer application, which is both display and control). A mobile phone contacts list (or "address book") was a good example of where a menu became long (easily 50 items or more), and the scrollbar could play an important role in the display. Yalla and Walker demonstrated the potential benefits of the proportionally mapped auditory scrollbars for visually impaired participants.

2.2 Usability Improvements by Nonspeech Auditory Cues

2.2.1 Performance. Performance improvements from the addition of auditory cues in menu navigation tasks has been quantified by several metrics such as reaction time, the number of key presses, accuracy, and error rate. In earlier work, earcons have shown superior performance compared to some other less systematic sounds [Brewster et al. 1992], and compared to no sound in a desktop computer [Brewster 1997], in a PDA [Brewster and Cryer 1999], and in a mobile phone [Leplâtre and Brewster 2000]. Further, musical timbres are more effective than simple tones [Brewster et al. 1992]. Also, in a hierarchical menu experiment, participants with earcons could identify their location with over 80% accuracy. These findings showed that the logic of earcons is promising to apply to hierarchy information [Brewster et al. 1996].

Spearcons have also recently shown promising performance results in menu navigation tasks. Walker et al. [2006] demonstrated that adding spearcons to a TTS menu leads to faster and more accurate navigation than TTS-only, auditory icons + TTS, and earcons + TTS menus. Spearcons also improved navigational efficiency more than menus using only TTS or no sound when combined with visual cues [Palladino and Walker 2008a, 2008b]. According to [Palladino and Walker 2008a], in their visuals-off condition, the mean time-to-target with spearcons + TTS was shorter than that with TTS-only, despite the fact that adding spearcons made the total system feedback longer.

2.2.2 Learnability. Learnability is an important component of system usability [Preece et al. 1996]. Since auditory display is time dependent and sometimes less familiar than visual display, the learnability of the information presentation type is crucial. Hutchins et al. [1986]once suggested the term "articulatory directness" to describe the relationship between the form and meaning of a statement. For example, a link between a paintbrush and a swishing sound involves more articulatory directness than a link between an incorrect entry and a beep. From this point of view, Gaver [1986] asserted that an increase in articulatory directness should be accompanied by an increased ease of learning. Therefore, in auditory icons, a very direct or "nomic" mapping is regarded as easier to learn than a metaphoric or symbolic mapping. Auditory icons provide benefits in natural mapping and learnability because the mapping between object and sound has often been learned throughout the user's life. In terms of naturalness, speech can claim a similar advantage because it is also learned for our entire lifetime. In this regard, spearcons and spindex cues (being speech-based) seem to adopt more natural mappings than earcons and even auditory icons; as such, they should be easily learned.

Brewster et al. [1996] reported an earcon learning study in which participants were tested to see if they could identify where previously unheard earcons would fit in the hierarchy. In that experiment training was divided into two parts. In the first part, the experimenter showed the hierarchy of the test menu, with earcons playing. In the second part, participants were given only five minutes to learn the earcons by themselves. The results of the study showed 90% accuracy of this task despite the short training sessions. Another more recent study [Dingler et al. 2008] compared the learnability of several auditory cues including auditory icons, earcons, spearcons, and

speech. In that study, participants were first trained to match a single target word with the sound associated with that environmental feature. After the training phase, they repeated the testing sessions until they had successfully identified all 20 features correctly. The results showed that spearcons are as learnable as speech, followed by auditory icons, and earcons are much more difficult to learn. The framework of the Dingler et al. study is different from the previous earcon study [Brewster et al. 1996] in that the sounds represent nonorganized common environmental features. In other words, researchers measured "what" the sounds represent, instead of "where" in the hierarchy. However, their results were consistent with Palladino and Walker's [2007] study on learning rates for earcons and spearcons using hierarchal menu items such as a cell phone menu list and a noun list. In that study, participants were asked to match the auditory cue to the proper location on the map after training. The result also supported that spearcons significantly outperform earcons in terms of learning rate.

2.2.3 Preference and Annoyance. Long or loud sounds may easily be annoying and can disturb the work of others. Therefore, sounds used in applications and devices should be designed carefully, especially with respect to duration, amplitude, and aesthetic quality. Although many researchers point out that aesthetic and annoyance issues are more important in auditory display than in visual display [Brewster 2008; Davison and Walker 2008; Kramer 1994; Leplâtre and McGregor 2004; Nees and Walker 2009], to date, research mainly has focused on performance issues. A similar point could be made about assistive technologies in general: aesthetics unfortunately seems to take a back seat to performance in nearly every case. We suggest a more even weighting is appropriate.

As a case in point, one of the widely used definitions of assistive technology is from Public Law (PL) 100-407, the Technical Assistance to the States Act in the United States: "Any items, piece of equipment or product system whether acquired commercially off the shelf, modified, or customized that is used to increase, maintain or improve functional capabilities of individuals with disabilities" [Cook and Hussey 2002]. According to this definition, assistive technology seems "sufficient" if better performance is obtained, regardless of user preference. Indeed, there has often been an emphasis on performance on the part of designers, which has been exemplified in the oft-cited paper by Andre and Wickens [1995], which discusses the challenges a designer can face in optimizing performance, even "when users want what's not best for them".

However, if there are many options (such as various iterations or variations of a design), it is also possible that performance is about the same across alternatives. In that case, researchers or designers might be asked to determine which option users prefer. Actually, the importance of the acceptance and the preference level of the interface has recently been increasing in mainstream user interaction and usability circles. As just one example, Norman [2004] has stressed the importance of visceral design. Further, he proposed that an attractive and natural design can sometimes improve usability as well as affective satisfaction [Norman 2004, 2007].

Moreover, Edworthy [1998] suggested that the nature of sound aesthetics is independent of performance outcomes. Users might turn off an annoying sound, even if the presence of that sound enhances performance with the system or device. Likewise, system sounds can improve the aesthetic experience of an interface without changing performance with the system [Nees and Walker 2009]. Traditionally, researchers of auditory warning signals have focused first on performance, based on "better *safe* than sorry" principles [Patterson 1985, 1990]. However, even in that area, designers have begun to change their strategy, and have more recently attempted to reduce annoyance and startle responses. Similarly, in the auditory interface design of everyday products, "better *efficient* than sorry" principles do not always work. For example, there are

some studies on polarity preference of sound [Walker 2002; Yalla and Walker 2008] and some studies have investigated preference or annoyance of earcons [Helle et al. 2001; Leplâtre and Brewster 2000; Marila 2002]. Earcons are sometimes preferred aesthetically because they are based on musical motives. Nevertheless, frequently played sounds in devices can make users annoyed. [Helle et al. 2001] investigated the subjective reactions of users who used sonified mobile phones. Users did not prefer the sonified phone; in fact they found it disturbing and annoying. It is possible that users would have rated the interface higher if the sounds had been less complex, especially since [Marila 2002] demonstrated that simpler sounds were preferred (and enhanced performance) over complex sounds. It is also possible for sounds to be disliked when they are of low quality [Helle et al. 2001]. Nowadays, sound quality limitations have generally been overcome as technology develops. However, there are many questions remaining to be answered on aesthetics, preference, and annoyance. Recent work has begun to study the subjective improvements to auditory menus from spearcons and other similar enhancements [Walker and Kogan 2009]. Following this line, the present study will show how the user's acceptance level can be changed as a function of preference and annoyance level, despite maintaining a similar level of objective performance.

2.3 Overview of the Current Research

2.3.1 Motivation and Inspiration. The Georgia Tech Sonification Lab and the Center for the Visually Impaired (CVI) in Atlanta have regular shared research meetings. Attendees demonstrate and discuss currently available technologies as well as a variety of novel interface ideas with end users, assistive technology specialists, teachers, and experts from industries. The idea of the spindex stemmed naturally from an expressed desire for faster auditory navigation than what is possible, even with previous techniques such as sped-up speech, spearcons, and auditory scrollbars. Blind users can comprehend spoken material at a speech rate that is 1.6∼2.5 times faster than sighted users [Asakawa et al. 2003; Moos and Trouvain 2007]. We have observed that very fast speech can enhance auditory menu navigation for many experienced visually impaired users, but for those who are not familiar with it (including most sighted users), fast speech is not sufficient (or even useful) by itself.

In consideration of the need for a better way to speed up menu navigation, the spindex concept was developed and empirically compared to TTS-alone. Experiment 1 presents the results of that investigation. Note that partial results of Experiment have been presented previously [Jeon and Walker 2009], and have now been extended and expanded here. Then, an iterative design process including several prototyping and evaluation cycles was completed, as described in Experiments 2 and 3.

2.3.2 Initial Design: Spindex (An Auditory Index Based on Speech Sounds). The spindex is created by associating an auditory cue with each menu item, in which the cue is based on the pronunciation of the first letter of each menu item. For instance, the spindex cue for "Apple" would be a sound based on the spoken sound "A." Note that the cue could simply be the sound for "A," but could also be a derivative, such as a sped-up version of that sound, reminiscent of the way spearcons are based on, but not identical to, sped-up speech sounds. The set of spindex cues in an alphabetical auditory menu is analogous to the visual index tabs that are often used to facilitate flipping to the right section of a thick reference book such as a dictionary or a telephone book.

When people control devices, there are two types of human motions in such tasks. In a *gross*-adjustment movement, the operator brings the controlled element to the approximate desired position. This gross movement is followed by a *fine*-adjustment,

in which the operator makes adjustments to bring the controlled element precisely to the desired location [Sanders and Mccormick 1993]. Likewise, when it comes to a search task such as navigation through an address book (visually or auditorily), the process can be divided into two stages: rough navigation followed by fine navigation [Klante 2004]. In rough navigation, users pass over the nontarget alphabet groups by glancing at the initials. For example, users quickly jump to the "T" section by passing quickly over "A" through "S." Then, once users reach a target zone (e.g., the "T"s) and begin fine navigation, they check where they are and cautiously tune their search. In auditory menus, people cannot jump as easily, given the temporal characteristics of spoken menu items. However, they still want to pass over the nontarget alphabetical groups as fast as possible. If a sound cue is sufficiently informative, users do not need to listen to the whole TTS phrase [Palladino and Walker 2007]. Although users can certainly pass over the TTS phrase without listening to the whole phrase, truncated speech sounds are complex and are not consistent, and thus do not allow users to make sure where they are in the list.

In contrast, the alphabet sounds ("A," "B," "C," etc.) can give enough information to users when sorting out the nontarget items because they provide simple and consistent sounds. The benefit of a cue structure like the spindex is even more obvious in long menus with many items in many categories or sections. Given that they are likely most useful in long menus, it is fortunate that spindex cues can be generated quickly (on the fly) by TTS engines, and do not require the creation and storage of many additional audio files for the interface. This is an important issue for mobile devices which, despite increasing storage for content files, are not designed to support thousands of extra files for their menu interface (as would be required with other cue types). Finally, because spindex cues are part of the original words and they are natural (based on speech sounds), they are expected not to require much, if any, training, and should be well liked.

2.3.3 Hypotheses. In this project, the goal was to make more intuitive and acceptable non-speech menu enhancements for navigation of (long) auditory menus. For this purpose, members of various user populations contributed to an iterative evaluation and redesign process. In Experiment 1, undergraduates navigated auditory menus with TTS + spindex and TTS-only to examine whether adding a spindex would improve navigation efficiency and preference. It was predicted that target search time and required learning for TTS + spindex would be shorter than that of TTS alone. Spindexes should also score higher than plain TTS on subjective rating scales. Experiment 2 included several spindex design alternatives and examined how these alternatives would affect users' actual and perceived performance as well as subjective evaluation. Experiment 3 extended Experiment 1 to include visually impaired participants. Further, Experiment 3 included design alternatives which had been devised in Experiment 2 and investigated how performance and preference results from visually impaired participants could be different from those of sighted participants.

3. EXPERIMENT 1

Experiment 1 compared spindex + TTS to plain TTS, with sighted undergraduate participants. In order for more systematic examinations, the study investigated both performance (objective and perceived) and subjective impressions of the spindex design.

3.1 Method

3.1.1 Participants. Twenty-seven undergraduate students participated in this study for partial credit in psychology courses. They reported normal or corrected-to-normal

Fig. 1. Screen grab of mobile phone simulation with name list used to collect data for Experiments 1, 2, and 3.

hearing and vision, signed informed consent forms, and provided demographic details about age and gender. In data analysis, we excluded two of them: one (male) answered all the trials the same, regardless of the correct answer, in Block 3; and another one (female) showed extreme outlier results (above 52 *Std.Error*) in Block 1, indicating she had not completed the task as instructed. Therefore, 25 participants' data were analyzed (14 female; mean age = 20.4 years).

3.1.2 Apparatus and Equipment. Stimuli were presented using a Dell Optiplex GX620 computer, running Windows XP on a Pentium 4, 3.2 GHz processor and 1 GB of RAM. An external Creative Labs Soundblaster Extigy sound card was used for sound rendering. Participants listened to auditory stimuli using Sennheiser HD 202 headphones, adjusted for fit and comfort. A 17" monitor was placed on a table 40 cm in front of the seated participant.

3.1.3 Stimuli. Two address book lists were composed for this study. The short list included 50 names and the long list contained 150 names. These names (e.g., "Allegra Seidner") were created using a random name generator.[1] Visual stimuli consisted of a mobile phone simulation with this list of names displayed on the simulated phone's screen in alphabetical order by first name. Ten names appeared on screen at a time, and the list scrolled downward and upward by pressing arrow keys on the keyboard (see Figure 1). The enter key was used to select the appropriate menu item. For both the auditory and visual components, if the participant reached the top or bottom of the list, the list did not wrap around. The experiment was built using Macromedia Director MX and Lingo.

[1]http://www.xtra-rant.com/gennames/ and http://www.seventhsanctum.com

3.1.3.1 Text-to-Speech. TTS files (.wav) were generated for all of the names using the AT&T Labs Text-To-Speech Demo program with the male voice "Mike-US English."[2]

3.1.3.2 Spindex. Spindex cues were also created by the AT&T Labs TTS Demo program. Each spindex cue consisted of only one syllable (except "W," which has three), pronouncing each of the 26 letters that represented the initial letter of the names. Spindex cues used in the address book were presented before each TTS cue, with 250 ms interval between them (similar to spearcons, [Palladino and Walker 2008a, 2008b). This led to stimuli like, "A. . . Allegra Seidner" or "D. . . Denice Trovato," where the leading "A" was pronounced as the letter name. If a participant pressed and held the up or down arrow key, only the spindex cues were presented, in rapid succession. That is, the initials of the names were generated without interval, in a preemptive manner (e.g., "A. . . A. . . B. . . B. . . C. . . D. . . Denice Trovato").

3.1.4 Design and Procedure. A split plot design was used in this experiment with one between-subjects variable and three within-subjects variables. The between-subjects variable was visual display type (On and Off). The within-subjects variables included auditory cue type (TTS-only and TTS + spindex), block (1∼4) and list length (Short and Long). The overall goal of the participant was to reach the target name in the address book menu as fast as possible, and press the enter key. There were no practice trials before the experiment blocks. The experiment was composed of four blocks in each condition. One block included 15 trials of different names as targets. All participants experienced the same procedure for each block, regardless of the assigned menu display conditions. The order of appearance of the condition was counterbalanced within and between participants.

After the informed consent procedure, participants were randomly assigned to one of the display type conditions. A simulated mobile phone address book menu was presented that contained items constructed with auditory and visual representations. On each trial, the target name was presented visually on the top of the computer screen. In the visuals-off group, the address book list was not shown in the phone simulator's display, but the target name was still presented visually (see Figure 2). When the participants first pressed the down arrow key, the timer started. Participants navigated through the menu system to find the assigned target name, and pressed the enter key to indicate selection of the requested target. This procedure was repeated for all 15 names in the block. Participants were then shown a screen that indicated that the next block of 15 trials was ready to start. After four conditions, participants filled out a short questionnaire. An eleven-point Likert-type scale was used for the self-rated levels of appropriateness (appropriate; functionally helpful) and likability (likable; fun; annoying) with regard to auditory cues.

3.2 Results of Experiment 1

3.2.1 Objective Performance. Errors in the tasks were very rare, so we focus here on the mean time-to-target analyses. The results are depicted in Figures 3–5. In particular, Figure 3 shows the interaction between list length and auditory cue type. These results were analyzed with a 2 (Visual types) x 2 (Auditory cue types) x 4 (Blocks) x 2 (List types) repeated measures ANOVA. The analysis revealed that participants searched significantly faster in the visuals-on condition than in the visuals-off condition (see

[2]http://www.research.att.com/∼ttsweb/tts/demo.php

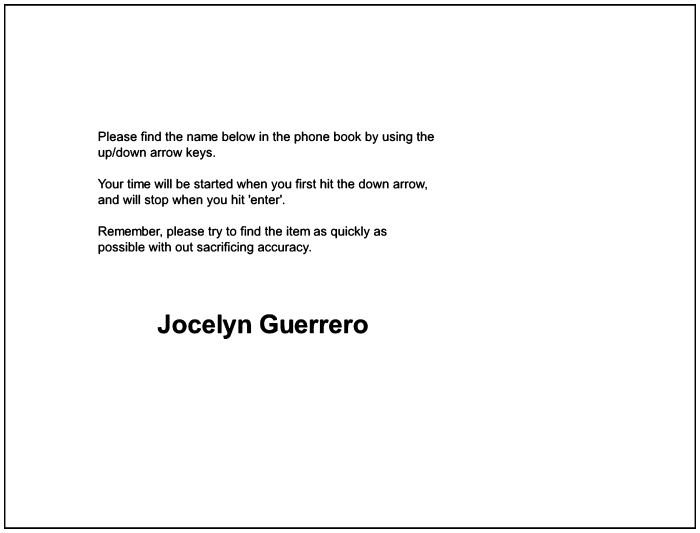

Fig. 2. Data collection screen of visual off condition used in Experiments 1, 2, and 3, showing the target name only on one trial. Participants listened to an auditory menu, and searched for the target name in that menu.

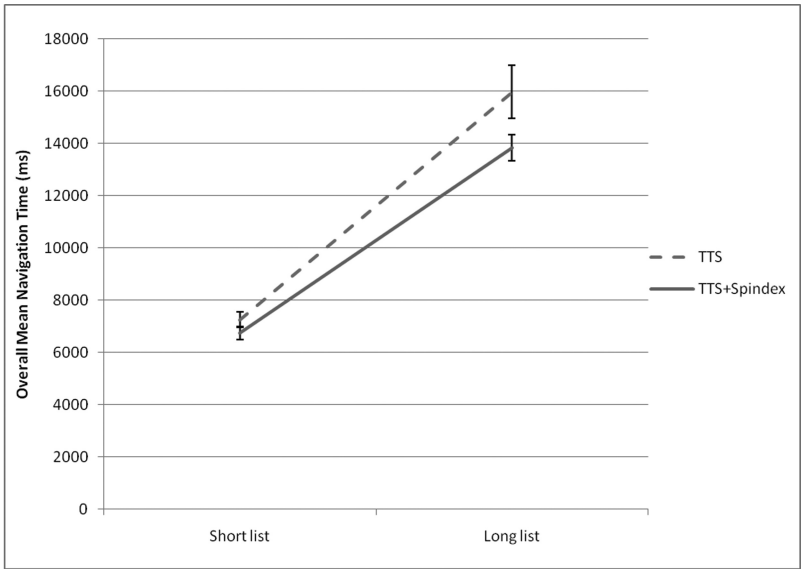

Fig. 3. Interaction of list length and auditory cue type for Experiment 1. Error bars show standard error of the mean.

Table I for statistics in Experiment 1). Participants in the TTS + spindex condition searched faster than those in the TTS-only condition. Also, the main effect for block was statistically significant, showing a practice effect. In addition, a short list led to significantly shorter search times than a long list. The interaction between list length and auditory cue type was also significant. This interaction term reflects the fact that the spindex is more beneficial in the long list than in the short list.

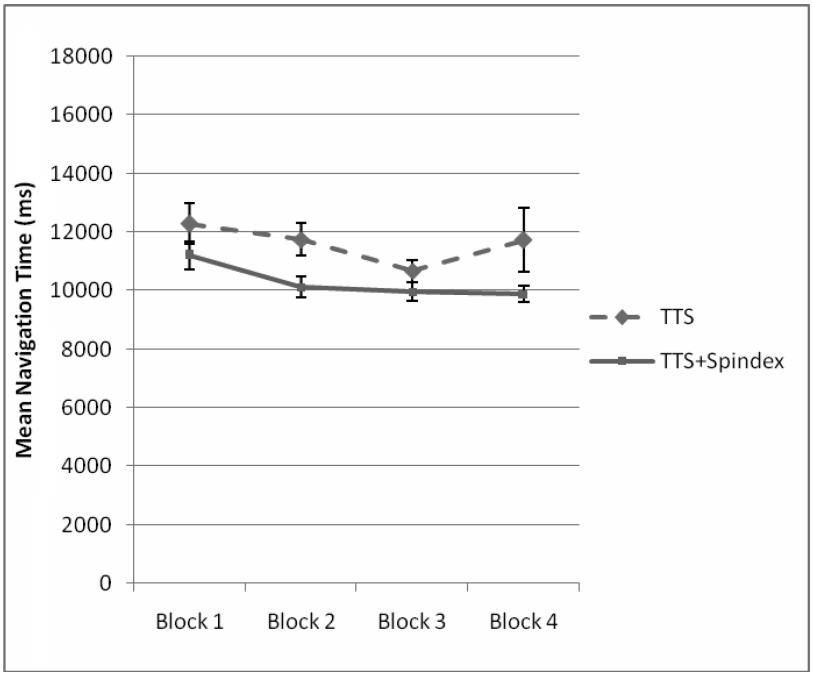

Fig. 4. Learning effect of auditory cue types across blocks in Experiment 1.

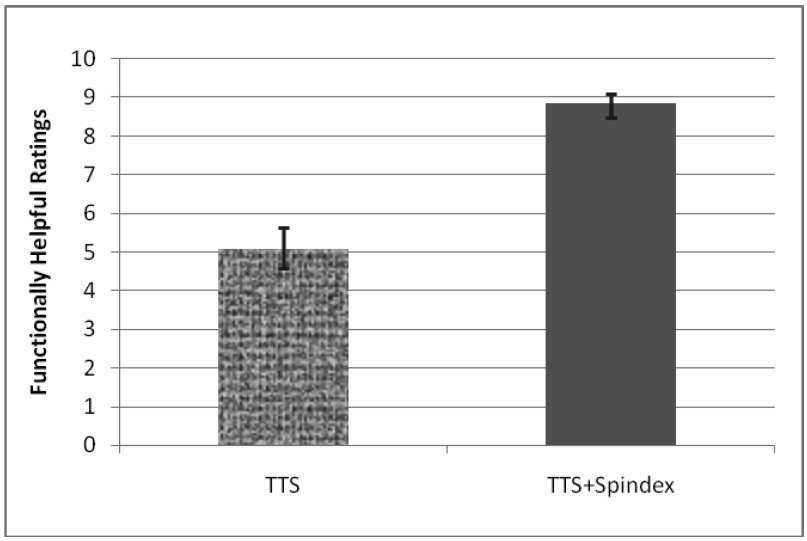

Fig. 5. Functionally helpful-rating results for Experiment 1. Likert scales from 0 (least helpful) to 10 (most helpful) were used. Higher ratings indicate preferred auditory menu enhancement types.

Figure 4 illustrates the learning effect of both auditory cue types. Time to target for the TTS + spindex condition reached the lowest (fastest) point in Block 2. In contrast, the TTS-only condition reached the lowest point in Block 3. This implies that the spindex required less learning than the TTS alone. Paired-samples t-tests for each condition supported this result. In the TTS + spindex condition, the mean time to

Table I. Statistics for Experiment 1

Measures	Conditions		Statistics
Objective Performance (Navigation Time (ms))	Visuals-on M = 8474, SD = 640	Visuals-off M = 13423, SD = 667	$F(1, 23) = 28.67$, $p < .001$, $\eta_p^2 = .56$
	TTS-only M = 11606, SD = 638	TTS + Spindex M = 10292, SD = 326	$F(1, 23) = 10.04$, $p < .05$, $\eta_p^2 = .30$
	Main Effect for Block		$F(3, 69) = 9.25$, $p < .001$, $\eta_p^2 = .29$
	Short List M = 7003, SD = 238	Long List M = 14896, SD = 731	$F(1, 23) = 190.15$, $p < .001$, $\eta_p^2 = .89$
	Interaction between List Length and Auditory Cue Type		$F(1, 23) = 8.95$, $p < .05$, $\eta_p^2 = .28$.
	Block 1 (TTS + Spindex) M = 11101, SD = 3514	Block 2 (TTS + Spindex) M = 10038, SD = 2752	Block 2 (TTS + Spindex) $t(24) = 4.241$, $p < .001$
	Block 1 (TTS-only) M = 12154, SD = 4824)	Block 3 (TTS-only) M = 10573, SD = 3044	$t(24) = 2.704$, $p < .05$
Perceived Performance Functionally Helpful Scale	TTS-only M = 5.08, SD = 2.72	TTS + Spindex M = 8.84, SD = 1.25	TTS + Spindex $t(24) = -5.914$, $p < .001$
Appropriate Scale	M = 7.08, SD = 2.18	M = 7.92, SD = 1.82	$t(24) = -1.731$, $p = .096$
Subjective Preference Likable Scale	TTS-only M = 5.88, SD = 2.51	TTS + Spindex M = 4.92, SD = 2.40	TTS-only $t(24) = 1.439$, $p = .163$
Fun Scale	M = 4.32, SD = 2.30	M = 4.68, SD = 2.21	$t(24) = -.725$, $p = .475$
Annoying Scale	M = 4.92, SD = 6.24	M = 6.24, SD = 2.39	$t(24) = -1.894$, $p = .070$

target of Block 1 and Block 2 were significantly different. The mean time of Block 2 and the remaining blocks showed no difference. However, in the TTS-only condition, the mean time to target of Block 1 and Block 3 were significantly different. Moreover, in Block 4, the mean time numerically increased again though it was not significantly different from Block 3. These data show that learning in the TTS-only condition was relatively slower than in the TTS + spindex condition.

3.2.2 Perceived Performance. Perceived performance was measured by ratings on Likert scales of "appropriate" and "functionally helpful." Figure 5 depicts the "functionally helpful" rating results, showing that undergraduate students rated the TTS + spindex types more highly than the TTS-only. "Appropriate" scale scores did not produce statistically reliable differences.

3.2.3 Subjective Preference. We also measured subjective preference including "likable," "fun," and "annoying" using Likert scales. However, for the subjective preference results, there was no statistically significant difference between auditory cue types, on the "likable" scale, on the "fun" scale, or on the "annoying" scale.

3.3 Discussion of Experiment 1

In this experiment, undergraduate participants showed better performance in the TTS + spindex condition than in the TTS-only condition. The spindex enhancement effect was larger for longer auditory menus (150 items) than for relatively short menus (50 items). This is due to the fact that even small per-item enhancements lead to important and noticeable improvements in navigation times in long lists. This bodes very well for using spindex cues in extemely long menus, such as those found in MP3 players.

The spindex seems to leverage what users are familiar with from tangible examples of long menus. For example, dictionaries and reference books often have physical

and visual tabs that serve the same function in visual search as the spindex does in auditory search. Beck and Elkerton [1989] already showed that visual indexes could decrease search time with lists. Thus, we might explain that the spindex is also the successful translation from the visual display into the auditory display.

The benefit of an auditory index (spindex) can be explained by the users' different strategies in the search processes: In the rough navigation stage, users exclude nontargets until they approach the alphabetical area containing the target. This is possible because they already know the framework of alphabetic ordering and letters. Thus, during this process, they do not need the full text of nontarget menu items. It is enough for them to obtain only a little information in order to decide whether they are in the target zone or not. After users perceive that they have reached the target zone (e.g., the "T"s), they then need the detailed information to compare items with the target. Between these processes, the spindex-enhanced auditory menu can contribute significant per-item speedups in the rough search stage, then the TTS phrase still supports detailed item information in the fine search stage.

The fact that participants gave equal or higher scores to the spindex menu on the subjective rating (though only the "functionally helpful" scale showed statistically significant difference) indicated that they did feel that the spindex was helpful in the navigation task. Of course, it is encouraging when objective performance (navigation time) and subjective assessments match. Even the few participants whose search times were not statistically better in the spindex condition said that their strategy for navigation was to hear the initial alphabet sound of the names. This validates the spindex approach, even if in some cases it did not lead to a measurable improvement. At least, it did no harm. It may simply be the case that a reliable improvement from a spindex menu would come at even longer list lengths for these participants; that remains to be determined. It is encouraging that using spindex cues requires little or no learning, which means a low threshold for new users. These advantages can increase the possibility of application of the spindex to real devices for all users.

Nevertheless, the details of adding a spindex need to be refined, in order to minimize annoyance. Thus, we implemented alternative designs such as attenuating the intensity of the spindex cues after the first one, and the use of a spindex cue only when crossing sub-list boundaries (e.g., for the first item starting with B, then the first C, and so on). We report on these, next.

4. EXPERIMENT 2

With alternatives developed after Experiment 1, we conducted Experiment 2 to examine how newer spindex design alternatives could improve users' subjective satisfaction and reduce annoyance, while showing at least comparable performance.

4.1 Method

4.1.1 Participants. Twenty-six undergraduate students (15 female; mean age = 19.5 years) participated in Experiment 2 for partial credit in psychology courses. All reported normal or corrected-to-normal vision and hearing, signed informed consent forms, and provided demographic details about age and gender. None had participated in Experiment 1.

4.1.2 Apparatus and Equipment. The apparatus was the same as in Experiment 1.

4.1.3 Stimuli. The same phone contact list as in Experiment 1 was used. In Experiment 2, we used only the long list condition (150 names) and the visuals-off condition. However, participants could still see the target name on the screen (see Figure 2).

Fig. 6. Types of spindex in Experiment 2 (From left, the basic, attenuated, decreased, and minimal type). The basic spindex played a sound before each menu item. The attenuated and the decreased spindex varied the loudness of the audio enhancements. The minimal spindex only played in front of the first item in each letter category.

4.1.3.1 Text- to- Speech. TTS files (.wav) were the same as in Experiment 1.

4.1.3.2 Spindex Cues. The *basic* spindex cues (used in Experiment 1) were created by generating TTS files for each letter (e.g., "A"), and editing them in Cool Edit Pro 2.0. Each spindex cue (other than "W") consisted of only one syllable, pronouncing one of the 26 English letters.

For Experiment 2, we created three alternative types in addition to the *basic* spindex: *attenuated*, *decreased*, and *minimal* cues (see Figure 6). The attenuated version contained lower amplitude cues that were attenuated by 20 dB from the first menu item in a letter category. For the decreased spindex, the intensity of the spindex cues gradually decreased after the first one. The amplitude of the last item in a category (e.g., the last name starting with "A") was set at −20 dB from the first item. Moving upwards from the last cue of a category, the loudness level of the cues increased +2dB for each menu item. For example, if an alphabet category included seven names, the level of the spindex cues were composed of the first one with original intensity, then the following levels: −10dB, −12dB, −14dB, −16dB, −18dB, and −20dB. For the minimal spindex, we used the spindex cues only when it crossed the category boundaries (e.g., for the first menu item starting with A, then the first item starting with B, and so on).

4.1.4 Design and Procedure. There were five within-subjects conditions, based on cue type: TTS-only and TTS + spindex, with four types of spindex. The overall goal of the participants was to reach the target name in the contact list menu as fast as possible, and press the enter key.

There were no practice trials before the experiment blocks. Each condition contained two blocks. One block included 15 trials of different names as targets. All participants experienced the same procedure for each block, regardless of the assigned auditory cue conditions. The order of appearance of the spindex conditions was fully counterbalanced; the TTS-only condition was always presented as the last condition (see Discussion).

On each trial, participants could see only the target name on the computer screen (see Figure 2). Participants navigated through the menu to find the assigned target name and pressed the enter key. This procedure was repeated for all 15 names in a block. After five conditions (two blocks each), participants filled out a short questionnaire. The same eleven-point Likert-type scale as in Experiment 1 was also used for the self-rated levels of perceived performance and subjective preference with regard to auditory cues. Finally, users were asked to provide comments on the study.

4.2 Results of Experiment 2

4.2.1 Objective Performance. The results are depicted in Figures 7–10. In particular, Figure 7 shows overall mean time to target (i.e., "search time", in ms) for each of the auditory cue types. These results were analyzed with a 5 (Auditory cue type) x 2 (Block) repeated measures ANOVA, which revealed a statistically significant difference between auditory cue types in mean search time (see Table II for statistics in

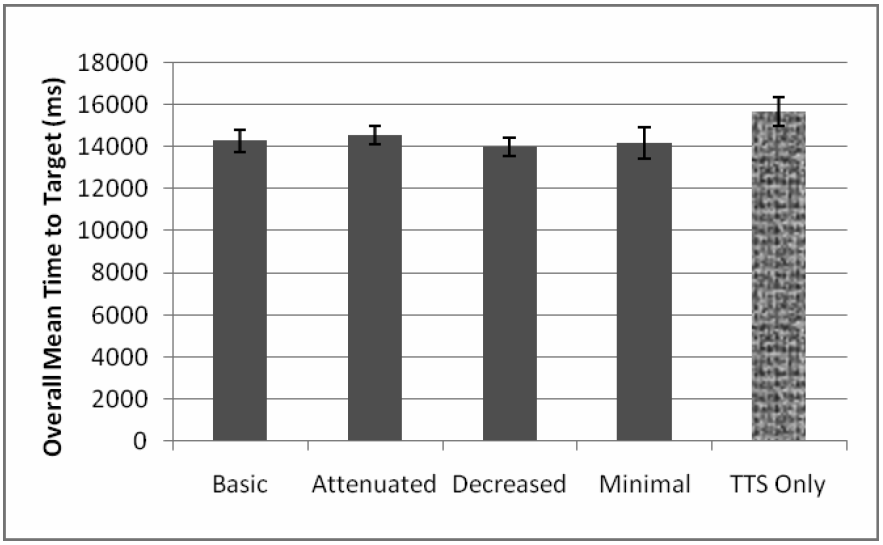

Fig. 7. Overall mean time to target (ms) for Experiment 2. Lower times indicate better performance.

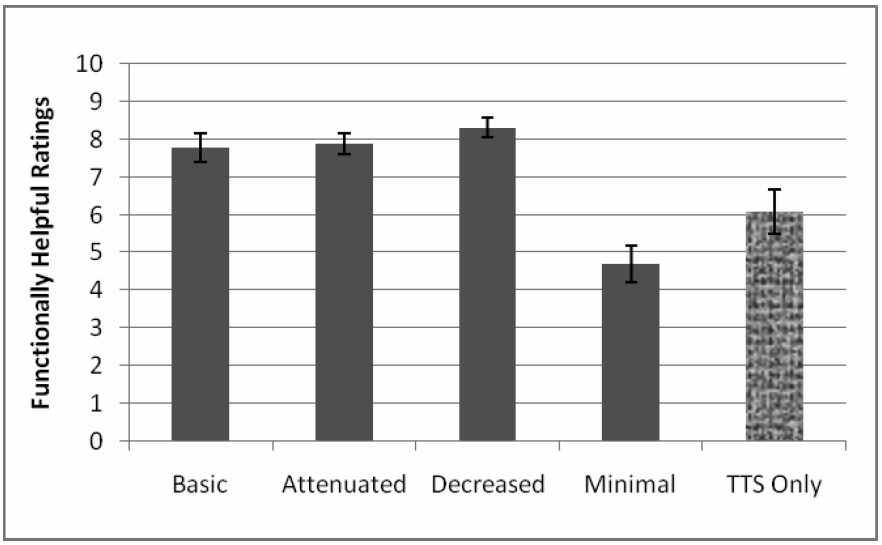

Fig. 8. Functionally helpful-rating results for Experiment 2.

Experiment 2). Also, Block 2 led to significantly shorter search times than Block 1, reflecting a practice effect. For the multiple comparisons between the auditory cue types, we conducted paired-samples *t*-tests. Participants searched significantly faster in all of the spindex conditions than in the TTS-only condition.

4.2.2 Perceived Performance. Perceived performance was measured by ratings on Likert scales of "appropriate" and "functionally helpful." Here, we focus only on the analysis of "functionally helpful" because both showed similar results. Figure 8 contains the "functionally helpful" rating results, showing that users rated the basic, the attenuated, and the decreased spindex types more highly than the minimal

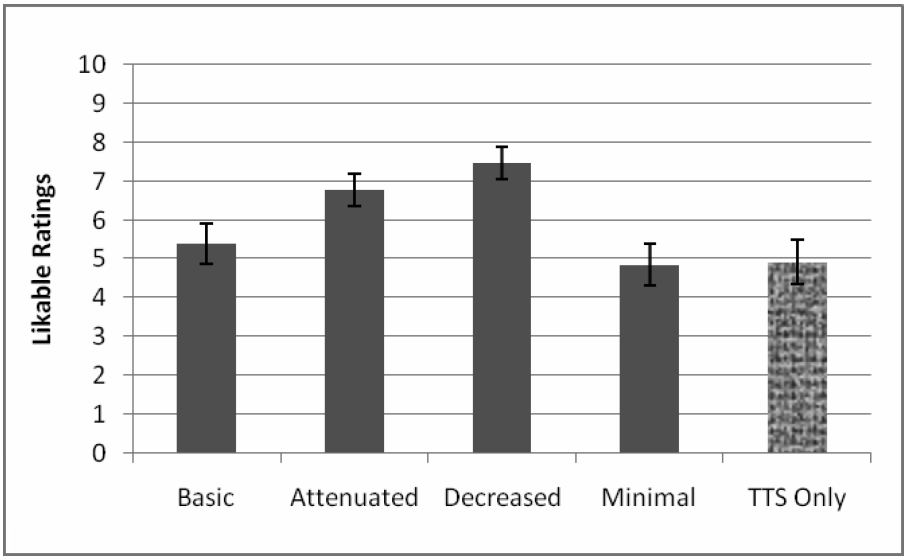

Fig. 9. Likable-rating results for four spindex enhancement types, versus TTS-only for Experiment 2. Higher ratings indicate preferred types.

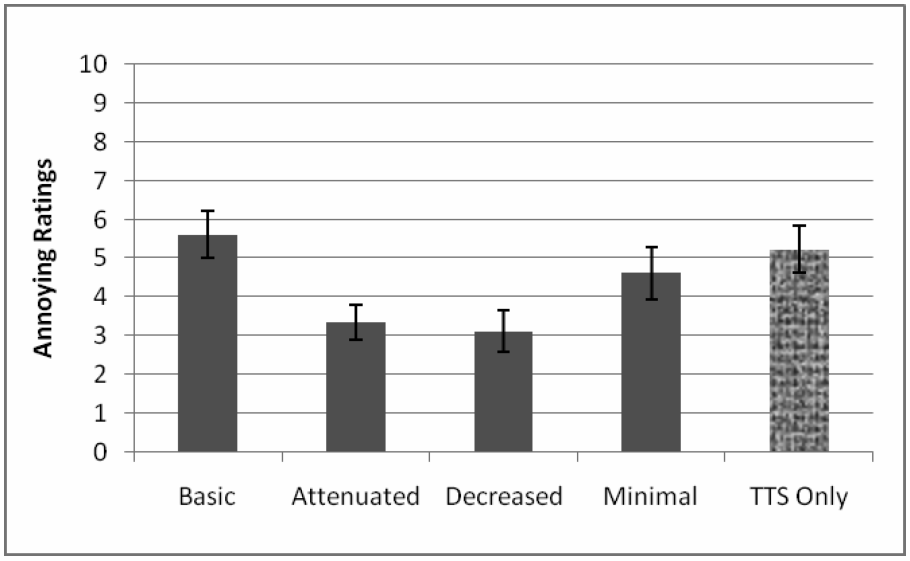

Fig. 10. Annoying-rating results for Experiment 2. Lower values (lower ratings of "annoyance") indicate preferred types.

spindex type and the TTS-only. Repeated measures ANOVA showed that there was a statistically significant difference between auditory cue types in "functionally helpful" rating values. Paired-samples *t*-tests showed that participants rated the basic, the attenuated, and the decreased spindex types significantly higher than the TTS-only on the "functionally helpful" scale. Similarly, the rating of the basic, the attenuated, and the decreased spindex type were higher than that of the minimal type. The "appropriate" scale showed similar patterns except two differences from the "functionally

Table II. Statistics for Experiment 2

Measures	Conditions		Statistics
Objective Performance (Navigation Time (ms))	Main Effect for Auditory Cue Type		$F(4, 100) = 2.814$, $p < .05$, $\eta_p^2 = .10$
	Block 1 M = 15011, SD = 533	Block 2 M = 14077, SD = 421	$F(1, 25) = 11.91$, $p < .05$, $\eta_p^2 = .32$
	TTS + Spindex (Basic) M = 14287, SD = 2772	TTS-only M = 15685, SD = 3368	$t(25) = 2.194$, $p < .05$
	TTS + Spindex (Attenuated) M = 14551, SD = 2216		$t(25) = 2.281$, $p < .05$
	TTS + Spindex (Decreased) M = 13999, SD = 2239		$t(25) = 2.647$, $p < .05$
	TTS + Spindex (Minimal) M = 14197, SD = 3889		$t(25) = 2.905$, $p < .001$
Perceived Performance	Main Effect for Auditory Cue Type		$F(4, 100) = 13.928$, $p < .001$, $\eta_p^2 = .36$
Functionally Helpful Scale	TTS + Spindex (Basic) M = 7.77, SD = 1.95	TTS-only M = 6.08, SD = 3.03	$t(25) = -3.085$, $p < .05$
	TTS + Spindex (Attenuated) M = 7.88, SD = 1.40		$t(25) = -3.093$, $p < .05$
	TTS + Spindex (Decreased) M = 8.31, SD = 1.32		$t(25) = -3.280$, $p < .05$
	TTS + Spindex (Basic) M = 7.77, SD = 1.95	TTS + Spindex (Minimal) M = 4.69, SD = 2.48	$t(25) = 4.293$, $p < .001$
	TTS + Spindex (Attenuated) M = 7.88, SD = 1.40		$t(25) = 5.366$, $p < .001$
	TTS + Spindex (Decreased) M = 8.31, SD = 1.32		$t(25) = 6.720$, $p < .001$
Appropriate Scale	TTS + Spindex (Decreased) M = 8.23, SD = 1.11	TTS + Spindex (Basic) M = 7.15, SD = 2.29	$t(25) = -2.409$, $p < .05$
	TTS-only M = 6.19, SD = 3.00		$t(25) = -1.642$, $p = .113$
Subjective Preference	Main Effect for Auditory Cue Type		$F(4, 100) = 7.076$, $p < .001$, $\eta_p^2 = .22$
Likable Scale	TTS + Spindex (Basic) M = 5.38, SD = 2.64	TTS + Spindex (Attenuated) M = 6.77, SD = 2.07	$t(25) = -2.825$, $p < .001$
	TTS + Spindex (Minimal) M = 4.84, SD = 2.71		$t(25) = 2.931$, $p < .001$
	TTS-only M = 4.92, SD = 2.83		$t(25) = -3.081$, $p < .001$
	TTS + Spindex (Basic) M = 5.38, SD = 2.64	TTS + Spindex (Decreased) M = 7.77, SD = 1.95	$t(25) = -3.823$, $p < .001$
	TTS + Spindex (Minimal) M = 4.84, SD = 2.71		$t(25) = 3.456$, $p < .001$
	TTS-only M = 4.92, SD = 2.83		$t(25) = -3.418$, $p < .001$
Fun Scale	TTS + Spindex (Attenuated) M = 5.35, SD = 2.64	TTS + Spindex (Decreased) M = 6.27, SD = 2.86	$t(25) = 3.157$, $p < .001$
Annoying Scale	Main Effect for Auditory Cue Type		$F(4, 100) = 5.027$, $p < .05$, $\eta_p^2 = .17$
	TTS + Spindex (Basic) M = 5.62, SD = 3.01	TTS + Spindex (Attenuated) M = 3.35, SD = 2.31	$t(25) = 3.789$, $p < .001$
	TTS-only M = 5.23, SD = 3.05		$t(25) = 2.942$, $p < .001$
	TTS + Spindex (Basic) M = 5.62, SD = 3.01	TTS + Spindex (Decreased) M = 3.12, SD = 2.70	$t(25) = 4.011$, $p < .001$
	TTS-only M = 5.23, SD = 3.05		$t(25) = 2.779$, $p < .05$
	TTS + Spindex (Attenuated) M = 3.35, SD = 2.31	TTS + Spindex (Minimal) M = 4.84, SD = 2.71	$t(25) = -1.649$, $p > .05$
	TTS + Spindex (Decreased) M = 3.12, SD = 2.70		$t(25) = -1.971$, $p > .05$

helpful" scale. In this case, even the decreased type was rated as more appropriate than the basic type. However, there was no difference between the basic type and the TTS-only.

4.2.3 Subjective Preference. We also measured subjective preference such as "likable," "fun," and "annoying" using Likert scales. Figures 9 and 10 show the results of "likable" and "annoying." These two figures suggest that participants favor the attenuated and the decreased types only. Repeated measures ANOVA showed a statistically significant difference between auditory cue types for the "likable" rating values and "annoying" rating values. Paired-samples *t*-tests showed that on the "likable" scale (Figure 9) participants rated the attenuated spindex significantly higher than the basic spindex, the minimal spindex, and the TTS-only. Also, participants rated the decreased spindex significantly higher on the "likable" scale than the basic spindex, the minimal spindex, and the TTS-only. Interestingly, the "fun" scale (not shown in a figure) showed only one difference from the "likable" scale. In that case, even the decreased spindex type was preferred over the attenuated type.

The result of paired-samples *t*-tests of the "annoying" scale (Figure 10) was similar to the "likable" scale except for the minimal spindex type: the attenuated spindex led to significantly lower "annoying" scores than the basic spindex and the TTS-only. Finally, the decreased spindex also reduced the "annoyance" significantly more than the basic spindex and the TTS-only. However, there was no significant difference between the attenuated and the minimal spindex nor between the decreased and the minimal spindex.

In summary, as shown in Figures 7–10, all types of the spindex enhanced the navigation performance relative to TTS-only. However, in perceived performance, participants felt that the minimal spindex type was not helpful. Despite its actual benefit, in the end, in subjective satisfaction the basic spindex type, which used in Experiment 1, was least preferred.

4.3 Discussion of Experiment 2

Experiment 1 showed that adding spindex cues to an auditory menu led to better performance than TTS alone. However, despite the performance benefits, there were some concerns that the particular spindex which we implemented was rated as somewhat annoying. Thus, Experiment 2 investigated whether alternative spindex designs could lead to higher ratings of acceptability, along with equal (or better) performance. This second study compared various speech indexes in terms of objective performance, perceived performance, and subjective preference. The results again showed that all types of spindex led to faster navigation than the TTS-only, even though the TTS-only condition was always at the end of the experiment order so that it might benefit from any learning effects. That is, the last condition for each participant should have a slight advantage due to learning; the fact that the TTS-only condition still fared the worst (although it was always in the final blocks) suggests the spindex enhancements are even more effective. This bodes well for this type of auditory menu enhancement.

In terms of perceived performance of the participants, the results showed some interesting complexities. Not all spindexes are created equally, it seems; at least not in terms of participants' reactions to them. Participants did not feel that the minimal spindex type was functionally helpful for their navigation task (although it actually was; see Figure 7). Moreover, in terms of subjective likability and annoyance, participants gave both the basic and the minimal spindex types lower ratings than the others. Overall, participants were positive toward only the attenuated and the

decreased types of spindex (again, despite the fact that all spindexes improved performance). Therefore, from these results either the attenuated or the decreased spindex type should be recommended for deployment in a real application. It should also be pointed out that the attenuated spindex design is somewhat simpler to implement, in that only two loudness levels of the cues are required, as compared to several levels required for the decreased spindex.

It may be instructive to discuss these results in terms of the display design guideline issue relating to, "How much is too much?" and "How little is too little?" [Sheridan 2005]. That is, display designers and engineers have to decide not only what to show, but also how to display it [Smallman 2005]. This study partially answered that question for this type of auditory interface enhancement. Abstract information was sometimes stronger than naïve realism [Sheridan 2005; Smallman and St. John 2005], but too much abstract information could also lead to failure: The basic spindex contains too much information, and the minimal type includes too little.

The minimal type was too little for users to recognize its presence, even though it helped them. The reason why participants did not like the minimal design is likely because they might want more contextual information. Ellis [2005] suggested that displays can be more effective when they provide redundancy. In a similar fashion, Burns [2005] pointed out that there needs to be a margin of safety in display design. Even if users overtly obtained enough aid with the minimal type, they seemed to want more continuous, but not intrusive information. Note that on the "annoying" scale, the minimal spindex was higher than the TTS-only.

We can also discuss the success of the attenuated and the decreased spindex designs in a theoretical acoustics framework. Historically, auditory warning designers focused on performance, given the safety implications. It has typically meant that warning sounds are tuned to possess a rapid onset time, high intensity and long duration so that they avoid any masking by environmental noise. This emphasis on performance sometimes led to unacceptable sounds, from the listener's perspective. However, even in auditory warning design, the notion of "better safe than sorry" has given way, because users' first reaction has been to figure out how to turn the "noise" off, rather than addressing the reason the warning was on in the first place. Thus, researchers have attempted to decrease the annoyance of the warning signal [Edworthy et al. 1991; Patterson 1990]. Design of the attenuated and the decreased spindex is in the same line as this approach. Consequently, it is evident that addressing both preference and performance is crucial to auditory display success.

5. EXPERIMENT 3

After several spindex studies with sighted participants, Experiment 3 included visually impaired participants. All end user populations are needed to participate in the universal design process, to confirm whether they respond similarly or differently. While visually impaired people can benefit most from this type of auditory menu enhancement, their later participation in the research was due to the typical difficulty in recruiting large enough samples for statistically reliable results. That is, while visually impaired advisors were actively consulting and contributing throughout the project, it was best to wait to involve a wider group of visually impaired participants until there were some pre-screened conditions for them to evaluate.

5.1 Method

5.1.1 Participants. A total of 17 blind and visually impaired adults participated and received $20 compensation for their participation. In the analysis, data from one

female were excluded because the experimenter had incorrectly operated the experimental device. Therefore, 16 participants' data were analyzed. Seven participants were clients of the Center for the Visually Impaired (CVI) in Atlanta (4 female; mean age 50.86 years, range 32–65 years). Nine participants were workers for the Georgia Industries for the Blind (Griffin Plant) (5 female; mean age 47.56 years, range 28–57 years). This led to a total of seven male and nine female participants (overall mean age 49.20 years, range 28–65 years). Three out of 16 participants were totally blind and 10 participants were legally blind. All were familiar with screen reader software. All participants reported normal hearing and provided demographic details about age, gender, visual impairments, and usage of a desktop computer and a moblie phone. Every participant in this experiment provided signed informed consent, with a sighted impartial reader/ witness.

5.1.2 Apparatus and Equipment. Stimuli were presented using a laptop computer, running MS Windows. The up, down, and arrow keys were marked with tape to guide participants' fingers. Participants used Sennheiser HD 202 headphones to listen to the auditory stimuli.

5.1.3 Stimuli. The address book menu from Experiments 1 and 2 was used, with only the long menu (150) and visuals-off condition. For the experiment session, TTS-only and TTS + (basic) spindex cues were used. Then, the attenuated, the decreased, and the minimal spindex cues were demonstrated for further subjective evaluation.

5.1.4 Design and Procedure. There were two within-subjects variables: cue type (TTS-only and TTS + (basic) spindex) and block (1∼2). As before, the overall goal of the participant was to reach the target name in the contact list menu as fast as possible, and press the enter key. The experiment was composed of two 15-trial blocks in each condition with no practice. All participants experienced the same procedure for each block, regardless of the assigned menu display conditions. The order of appearance of the condition was counterbalanced across participants.

On each trial, the target name was aurally presented by the experimenter. After hearing the target item, participants repeated the name once and they could ask the experimenter to repeat the target name at any time again. Participants were allowed to use either one hand or two hands for navigation and response. Since it was the visuals-off condition, the address book menu was not shown (see Figure 2). After the two conditions, participants filled out a short questionnaire. The same eleven-point Likert-type scale was used for the self-rated levels of perceived performance (appropriate; functionally helpful) and subjective preference (likable; fun; annoying) with regard to the auditory cues. The experimenter read the questions and recorded answers for all participants. Then, the experimenter demonstrated three alternative designs: the attenuated, the decreased, and the minimal spindex cues. The order of the presentation of alternatives was also counterbalanced across participants. In this demonstration session, participants were not asked to find the target, but asked to just browse the auditory menu until they felt sufficient familiarity with the sounds. Then, they had to choose the best auditory cue type among all five conditions (the TTS-only, the basic, the attenuated, the decreased, and the minimal spindex). Finally, participants provided comments on the study.

5.2 Results of Experiment 3

5.2.1 Objective Performance. The results are depicted in Figures 11 and 12. In particular, Figure 11 shows mean time to target (i.e., "search time," in ms) per block for each of the auditory cue types. These results were analyzed with a 2 (Auditory cue type) x 2

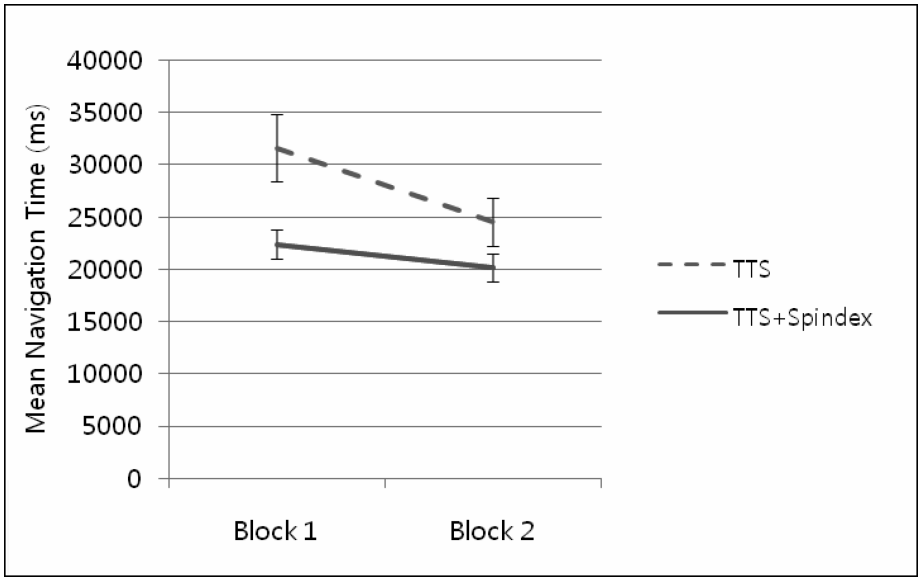

Fig. 11. Learning effect of auditory cue types across blocks in Experiment 3.

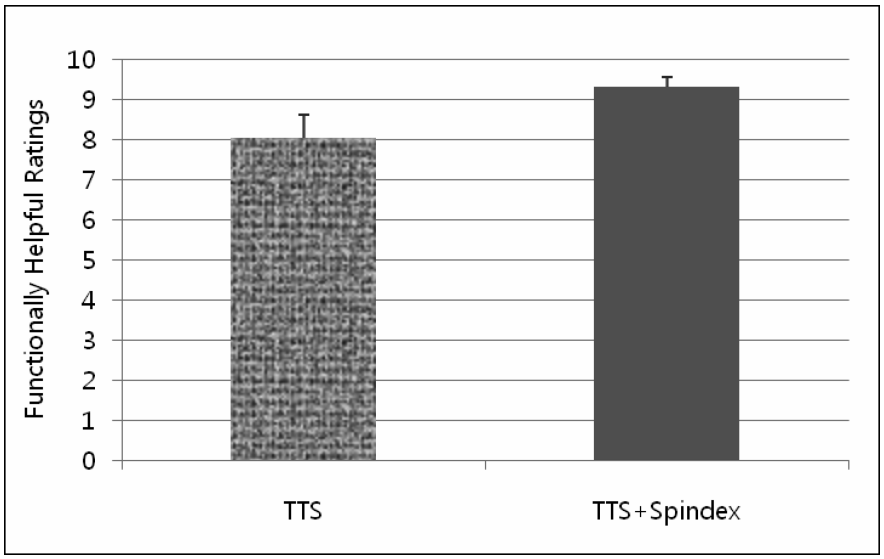

Fig. 12. Functionally helpful-rating results for Experiment 3.

(Block) repeated measures ANOVA, which revealed statistically significant differences in auditory cue types and in blocks for mean search time (see Table III for statistics in Experiment 3). The TTS + spindex condition led to significantly shorter search times than the TTS-only condition. Also, Block 2 led to significantly shorter search times than Block 1, reflecting a practice effect.

5.2.2 Perceived Performance. Perceived performance was measured by ratings on Likert scales of "appropriate" and "functionally helpful." Figure 12 contains the

Table III. Statistics for Experiment 3

Measures	Conditions		Statistics
Objective Performance (Navigation Time (ms))	TTS-only M = 28117, SD = 11637	TTS + Spindex M = 21308, SD = 5499	$F(1, 15) = 10.832$, $p < .01$, $\eta_p^2 = .42$
	Block 1 M = 27038, SD = 10940	Block 2 M = 22387, SD = 7664	$F(1, 25) = 11.91$, $p < .01$, $\eta_p^2 = .53$
Perceived Performance Functionally Helpful Scale	TTS-only M = 6.08, SD = 3.03	TTS + Spindex M = 7.77, SD = 1.95	TTS + Spindex $t(15) = -2.236$, $p < .05$
Appropriate Scale	M = 7.94, SD = 2.02	M = 8.63, SD = 1.71	$t(15) = -1.405$, $p = .180$
Subjective Preference Likable Scale	TTS-only M = 8.06, SD = 2.17	TTS + Spindex M = 8.81, SD = 1.68	TTS + Spindex $t(15) = -1.667$, $p = .118$
Fun Scale	M = 7.19, SD = 3.19	M = 7.81, SD = 2.48	$t(15) = -1.667$, $p = .116$
Annoying Scale	M = 1.31, SD = 1.82	M = 1.44, SD = 2.00	$t(15) = -.243$, $p = .812$.
Best Choice	Actual frequencies were significantly different from the case in which all frequencies are equal		$\chi^2(3, 16) = 9.50$, $p < .05$

"functionally helpful" rating results, showing that users rated the TTS + spindex type more highly than the TTS-only. Paired-samples *t*-tests showed that participants rated the TTS + spindex type significantly higher than the TTS-only on the "functionally helpful" scale. "Appropriate" scores did not produce statistically reliable differences.

5.2.3 Subjective Preference. We also measured subjective preference including "likable," "fun," and "annoying" using Likert scales. However, for the subjective preference data, there was no statistically significant difference between auditory cue types, on the "likable" scale, on the "fun" scale, or on the "annoying" scale.

5.2.4 Best Choice. For the best choice among the five alternatives, there were clear preferences. Eight people found the basic spindex type was the best and six people specified the attenuated spindex type as their most preferred type. Only one person chose the decreased and the minimal condition as the best. No one preferred the TTS alone. This distinction was statistically confirmed by analyzing the frequency of the choice: actual frequencies were significantly different from the case in which all frequencies are equal.

5.2.5 Comments. Participants' comments reflect their preference for spindex cues. Some participants favored spindex, directly noting, "I liked it. It is useful when looking at any files in alphabetical order,... I wish this speech index [were available] in my phone" and "hurry up with creating it on a phone book." One participant accurately pointed out the benefit of spindex, "spindex is better than TTS because [with spindex I] could monitor it better, with TTS I just had to guess and count how far it would go."

We found that many of the visually impaired participants were interested in obtaining more information from the more sophisticated sound designs: "in my phone, it generates TTS, I think it should be louder volume in the upper case and be lower in the lower case," "Liked attenuated sound with volume changes between letters, it helped separate lists," "Liked different cues, depending on the application of the use," and "Liked attenuated the best because it gave more feedback as to where I was on the list." Despite these promising comments, one was concerned about the interval between spindex cue and TTS: "Spindex was nice to hear letters pronounced, but difficult and slow to process letters and names."

Further, we could gather some insights about their mobile phones. For example, most participants preferred to have spoken functions on their mobile phone: "I like that my phone reads out number or name of people who are calling instead of just ringing," "I like feedback when hitting the key to make sure I hit the right number," "I like having phone stats (number of bars and how much battery left) read out at the press of one button." Nevertheless, they still need more sound feedback: "On my phone, I have physical dots on the screen [pixels] for numbers, without any auditory cues. Have trouble getting out of menus, get stuck in menus. Need more accessibility to menus."

5.3 Discussion of Experiment 3

Overall results from Experiment 3 were quite similar to results from Experiment 1 and Experiment 2. In Experiment 3, visually impaired participants demonstrated better performance in the TTS + spindex condition than in the TTS-only condition. For the perceived performance, visually impaired participants also rated the spindex-enhanced menu higher than the TTS-only menu.

However, we found some slightly different patterns on the subjective rating scores. In contrast to Experiment 1, subjective rating scores for both auditory menu conditions were relatively higher in Experiment 3. For example, in Experiment 1 the average score of "appropriate," "functionally helpful," "likable," and "fun" was 5.59 out of 10 for the TTS-only and 6.59 for the TTS + spindex. On the other hand, in Experiment 3, the average score of "appropriate," "functionally helpful," "likable," and "fun" was 7.85 for the TTS-only and 8.60 for the TTS + spindex. On the annoyance scale, visually impaired participants gave lower (less annoying) scores than sighted participants, which meant that visually impaired people favored auditory menus more. In Experiment 1, the "annoyance" score was 4.92 for the TTS-only and 6.24 for the TTS + spindex. In contrast, in Experiment 3, annoyance scores were 1.35 for the TTS-only and 1.88 for the TTS + spindex. These lower annoyance scores reflect that for visually impaired people, auditory displays may be a necessity, more than just an option. Thus, they tend to rate both auditory menus as better or more useful than do sighted people. Also, visually impaired participants are more likely familiar with TTS, and its often mediocre speech quality.

This tendency was revealed again in the comparison between the subjective preference rating scores in Experiment 2 and the best choice results in Experiment 3. In Experiment 2, we did not ask participants to select the best condition, but the rating scores of the "likable" and the "fun" scales showed that they preferred the attenuated and the decreased types. Even on the "fun" scale, the decreased type was significantly preferred over the attenuated type. However, in Experiment 3, visually impaired people preferred the basic and the attenuated types to others. We can infer that for visually impaired people, the clarity of the basic version is more important than the fun of the decreased version. Moreover, because some of visually impaired participants recognize the benefit of the attenuated version, structural benefit and less annoyance, this type can fall into an overlapping preference region between sighted people and people with visual impairments.

6. CONCLUSION AND FUTURE WORK

The use of mobile devices with smaller or nonexistent screens, and longer and longer menus, is increasing. These devices are being used in more dynamic and mobile contexts, even with the device in a pocket. These devices are much more difficult to use for people with vision loss. Auditory menus, especially when well designed and sophisticated, can make these devices accessible. The present set of experiments assesses

a new type of auditory menu enhancement, the spindex. Further, these studies attempted to optimize design alternatives in order to provide enhanced design guidelines for auditory menus.

The spindex requires relatively minimal programming to be added to an auditory menu; and, it is certainly possible with these small additions to significantly improve both the performance and preference of auditory menus. As auditory menus and all auditory interfaces become more effective and more acceptable, they will lead to more universally accessible systems and devices, which will lead to increased accessibility for users with vision loss. To ensure this momentum continues, parallel considerations of both sighted users and visually impaired users must be encouraged. We hope that this attempt would show that design for disability in applications could also lead to beneficial innovations for users who are not disabled [Glinert and York 2008].

While the current results stand on their own and inform us about making auditory menus more accessible, one can look forward to the extended results with broader contexts. The spindex could be applied to other systems such as touch screen devices, which have limited tactile feedback, and thus may require more auditory feedback than others. Also, the combination of nonspeech auditory cues (e.g., spindexes and auditory scrollbars) can be further explored. In summary, the contextual enhancement of auditory menus using the spindex can improve subjective impressions, usability, and universal accessibility of the devices and can ultimately provide essential information to a more inclusive range of user populations.

REFERENCES

ABSAR, R. AND GUASTAVINO, C. 2008. Usability of non-speech sounds in user interfaces. In *Proceedings of the International Conference on Auditory Display (ICAD'08).*

ANDRE, A. D. AND WICKENS, C. D. 1995. When users want what's not best for them. *Ergonom. Design 3*, 4, 10–14.

ASAKAWA, C. AND ITOH, T. 1998. User interface of a home page reader. In *Proceedings of the Annual ACM Conference on Assistive Technologies (ASSETS'98)*. 149–156.

ASAKAWA, C., TAKAGI, H., INO, S., AND IFUKUBE, T. 2003. Maximum listening speeds for the blind. In *Proceedings of the International Conference on Auditory Display (ICAD'03)*.

BECK, D. AND ELKERTON, J. 1989. Development and evaluation of direct manipulation list. *SIGCHI Bull. 20*, 3, 72–78.

BLATTNER, M. M., SUMIKAWA, D. A., AND GREENBERG, R. M. 1989. Earcons and icons: Their structure and common design principles. *Hum.-Comput. Interact.,* 4, 11–44.

BREWSTER, S. A. 1997. Using non-speech sound to overcome information overload. *Displays, 17*, 179–189.

BREWSTER, S. A. 1998. The design of sonically-enhanced widgets. *Interact. Comput. 11*, 2, 211–235.

BREWSTER, S. A. 2008. *Nonspeech auditory output*. In *The Human Computer Interaction Handbook,* A. Sears and J. Jacko, Eds., Lawrence Erlbaum Associates, New York, 247–264.

BREWSTER, S. A. AND CRYER, P. G. 1999. Maximising screen-space on mobile computing devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'99)*. 224–225.

BREWSTER, S. A., WRIGHT, P. C., AND EDWARDS, A. D. N. 1992. A detailed investigation into the effectiveness of earcons. In *Proceedings of the 1st International Conference on Auditory Display (ICAD'94)*. 471–478.

BREWSTER, S. A., RATY, V. P., AND KORTEKANGAS, A. 1996. Earcons as a method of providing navigational cues in a menu hierarchy. In *Proceedings of the HCI'96*. Springer, 167–183.

BREWSTER, S. A., LEPLÂTRE, G., AND CREASE, M. G. 1998. Using non-speech sounds in mobile computing devices. In *Proceedings of the 1st Workshop on Human Computer Interaction with Mobile Devices*.

BREWSTER, S. A., LUMSDEN, J., BELL, M., HALL, M., AND TASKER, S. 2003. Multimodal 'eyes-free' interaction techniques for wearable devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'03)*. 473–480.

BURNS, C. M. 2005. Choosing the best from the good: Display engineering principles. In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting (HFES'05)*. 1556–1560.

COOK, A. M. AND HUSSEY, S. M. 2002. *Assistive Technologies: Principles and Practice.* Mosby, Inc.

DAVISON, B. D. AND WALKER, B. N. 2008. AudioPlusWidgets: Bringing sound to software widgets and interface components. In *Proceedings of the International Conference on Auditory Display (ICAD'08)*.

DINGLER, T., LINDSAY, J., AND WALKER, B. N. 2008. Learnability of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech. In *Proceedings of the International Conference on Auditory Display (ICAD'08)*.

EDWORTHY, J. 1998. Does sound help us to work better with machines? A commentary on Rauterberg's paper 'About the importance of auditory alarms during the operation of a plant simulator'. *Interact. Comput. 10*, 401–409.

EDWORTHY, J., LOXLEY, S., AND DENNIS, I. 1991. Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Hum. Factors 32*, 2, 205–231.

ELLIS, S. R. 2005. On redundancy in the design of spatial instruments. In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting (HFES'05)*. 1561–1563.

GAVER, W. W. 1986. Auditory icons: Using sound in computer interfaces. *Hum.-Comput. Interact. 2*, 167–177.

GAVER, W. W. 1989. The SonicFinder, a prototype interface that uses auditory icons. *Hum.-Comput. Interact. 4*, 67–94.

GLINERT, E. P. AND YORK, B. W. 2008. Computers and people with disabilities. *ACM Trans. Access. Comput. 1*, 2, 1–7.

GOOSE, S. AND MOLLER, C. 1999. A 3D audio only interactive web browser: Using spatialization to convey hypermedia document structure. In *Proceedings of the 7th ACM International Conference on Multimedia* (Part 1) *(MULTIMEDIA'99).* 363–371.

HELLE, S., LEPLÂTRE, G., MARILA, J., AND LAINE, P. 2001. Menu sonification in a mobile phone: A prototype study. In *Proceedings of the International Conference on Auditory Display (ICAD'01).*

HUTCHINS, E. L., HOLLAN, J. D., AND NORMAN, D. A. 1986. *Direct Manipulation Interfaces.* Lawrence Erlbaum Associates Inc.

JAMES, F. 1998. Lessons from developing audio HTML interfaces. In *Proceedings of the Annual ACM Conference on Assistive Technologies (ASSETS'98)*. 27–34.

JEON, M. AND WALKER, B. N. 2009. "Spindex": Accelerated initial speech sounds improve navigation performance in auditory menus. In *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society (HFES09)*. 1081–1085.

JEON, M., DAVISON, B., WILSON, J., RAMAN, P., AND WALKER, B. N. 2010. Advanced auditory menus for universal access to electronic devices. In *Proceedings of the 25th Annual International Technology and Persons with Disabilities Conference (CSUN'10)*.

KEHOE, A. AND PITT, I. 2006. Designing help topics for use with Text-To-Speech. In *Proceedings of the 24th Annual ACM International Conference on Design of Communication (SIGDOC'06)*. 157–163.

KLANTE, P. 2004. Auditory interaction objects for mobile applications. In *Proceedings of the 7th International Conference on Work with Computing Systems (WWCS'04)*.

KRAMER, G. 1994. An introduction to auditory display. In *Auditory Display: Sonification, Audification, and Auditory Interfaces*, G. Kramer Ed., Addison-Wesley, 1–77.

LEPLÂTRE, G. AND BREWSTER, S. A. 2000. Designing non-speech sounds to support navigation in mobile phone menus. In *Proceedings of the International Conference on Auditory Display (ICAD'00)*. 190–199.

LEPLÂTRE, G. AND MCGREGOR, I. 2004. How to tackle auditory interface aesthetics? Discussion and case study. In *Proceedings of the International Conference on Auditory Display (ICAD'04)*.

MARILA, J. 2002. Experimental comparison of complex and simple sounds in menu and hierarchy sonification. In *Proceedings of the International Conference on Auditory Display (ICAD'00)*.

MOOS, A. AND TROUVAIN, J. 2007. Comprehension of ultra-fast speech - blind vs. 'normally hearing' persons. In *Procceedings of the 16th International Congress of Phonetic Sciences*. 677–680.

MORLEY, S., PETRIE, H., AND 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 (ASSETS'98)*.

MYNATT, E. 1997. Transforming graphical interfaces into auditory interfaces for blind users. *Hum.-Comput. Interact. 12*, 7–45.

NEES, M. A. AND WALKER, B. N. 2009. Auditory interfaces and sonification. In *The Universal Access Handbook*, C. Stephanidis Ed., CRC Press Taylor & Francis, New York, 507–521.

NEWELL, A. F. 2008. Accessible computing: Past trends and future suggestions. *ACM Trans. Access. Comput. 1*, 2, 9:1–7.

NORMAN, D. A. 2004. Emotional Design. Basic Books, New York.

NORMAN, D. A. 2007. The Design of Future Things. Basic Books, New York.

PALLADINO, D. K. AND WALKER, B. N. 2007. Learning rates for auditory menus enhanced with spearcons versus earcons. In *Proceedings of the International Conference on Auditory Display (ICAD'07)*. 274–279.

PALLADINO, D. K. AND WALKER, B. N. 2008a. Efficiency of spearcon-enhanced navigation of one dimensional electronic menus. In *Proceedings of the International Conference on Auditory Display (ICAD'08)*.

PALLADINO, D. K. AND 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 (HFES08)*. 1262–1266.

PATTERSON, R. D. 1985. Auditory warning system for high-workload environments. *Ergon. Interact. 85*, 163–165.

PATTERSON, R. D. 1990. Auditory warning sounds in the work environments. *Phil. Trans. Roy. Soc. Lond. B*, 327, 484–492.

PITT, I. J. AND 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 (ASSETS'96)*. 124–130.

PREECE, J., ROGERS, Y., SHARP, H., BENYON, D., HOLLAND, S., AND CAREY, T. 1996. *Human-Computer Interaction*. Addison-Wesley, London.

SANDERS, M., S. AND MCCORMICK, E. J. 1993. Chapter11: Controls and data entry devices. In *Human Factors in Engineering and Design*, M. S. Sanders and E. J. McCormick Eds., McGraw-Hill, Inc, New York, 334–382.

SHERIDAN, T. B. 2005. Allocating bits to displays for dynamic control: When more is more and when more is less. In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting (HFES'05)*. 1569–1572.

SMALLMAN, H. S. 2005. Overarching principles of display design. In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting (HFES'05)*. 1554–1557.

SMALLMAN, H. S. AND ST. JOHN, M. 2005. Naive realism: Limits of realisms as a display principle. In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting (HFES'05)*. 1564–1568.

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 (ASSETS'94)*. 39–46.

VARGAS, M. L. M. AND ANDERSON, S. 2003. Combining speech and earcons to assist menu navigation. In *Proceedings of the International Conference on Auditory Display (ICAD'03)*.

WALKER, B. N. 2002. Magnitude estimation of conceptual data dimensions for use in sonification. *J. Experi. Psych.: Appl. 8*, 4, 211–221.

WALKER, B. N. AND KOGAN, A. 2009. Spearcons enhance performance and preference for auditory menus on a mobile phone. In *Universal Access in HCI*, Part II: Lecture Notes in Computer Science vol. 5615, C. Stephanidis Ed., Springer-Verlag, Berlin, 445–454.

WALKER, B. N. AND KRAMER, G. 2004. Ecological psychoacoustics and auditory displays: Hearing, grouping, and meaning making. In *Ecological Psychoacoustics*, J.G. Neuhoff, Ed., Academic Press, New York, 150–175.

WALKER, B. N., NANCE, A., AND LINDSAY, J. 2006. Spearcons: Speech-based earcons improve navigation performance in auditory menus. In *Proceedings of the International Conference on Auditory Display (ICAD'06)*. 95–98.

WILSON, J., WALKER, B. N., LINDSAY, J., CAMBIAS, C., AND DELLAERT, F. 2007. SWAN: System for wearable audio navigation. In *Proceedings of the 11th International Symposium on Wearable Computers (ISWC'07)*.

YALLA, P. AND WALKER, B. N. 2008. Advanced auditory menus: Design and evaluation of auditory scrollbars. In *Proceedings of the Annual ACM Conference on Assistive Technologies (ASSETS'08)*. 105–112.

Received June 2010; revised December 2010; accepted December 2010