"Spindex" (Speech Index) Enhances Menus on Touch Screen Devices with Tapping, Wheeling, and Flicking

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

Users interact with many electronic devices via menus such as auditory or visual menus. Auditory menus can either complement or replace visual menus. We investigated how advanced auditory cues enhance auditory menus on a smartphone, with tapping, wheeling, and flicking input gestures. The study evaluated a spindex (speech index), in which audio cues inform users where they are in a menu; 122 undergraduates navigated through a menu of 150 songs. Study variables included auditory cue type (text-to-speech alone or TTS plus spindex), visual display mode (on or off), and input gesture (tapping, wheeling, or flicking). Target search time and subjective workload were lower with spindex than without for all input gestures regardless of visual display mode. The spindex condition was rated subjectively higher than plain speech. The effects of input method and display mode on navigation behaviors were analyzed with the two-stage navigation strategy model. Results are discussed in relation to attention theories and in terms of practical applications.

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, touch screen, input gestures, tapping, wheeling, flicking

ACM Reference Format:
Jeon, M., Walker, B. N., and Srivastava, A. 2012. "Spindex" (speech index) enhances menus on touch screen devices with tapping, wheeling, and flicking. ACM Trans. Comput.-Hum. Interact. 19, 2, Article 14 (July 2012), 27 pages.
DOI = 10.1145/2240156.2240162 http://doi.acm.org/10.1145/2240156.2240162

1. INTRODUCTION

Research on the use of nonspeech sounds for information display in user interfaces has rapidly grown since the early 1990s [Kramer 1994; Nees and Walker 2009; Walker and Kramer 2006]. The benefits of such auditory displays have been demonstrated in a wide range of different applications, from systems for blind people [Edwards 1989; Jeon and Walker 2011; Kane et al. 2008; Raman 1997] to mobile devices [Brewster and

Portions of this research were supported through the WirelessRERC, funded by NIDRR Grant # H133E060061.
Authors' addresses: M. Jeon, School of Psychology, Georgia Institute of Technology; B. N. Walker, School of Psychology and School of Interactive Computing, Georgia Institute of Technology, email: bruce.walker@psych.gatech.edu; A. Srivastava, School of Interactive Computing, Georgia Institute of Technology.

DOI 10.1145/2240156.2240162 http://doi.acm.org/10.1145/2240156.2240162

Cryer 1999; Brewster et al. 1998; Jeon and Walker 2011; Klante 2004; Leplâatre and Brewster 2000; Li et al. 2008; Palladino and Walker 2007, 2008a, 2008b; Pirhonen et al. 2002; Vargas and Anderson 2003; Walker et al. 2006; Zhao et al. 2007], and ubiquitous/wearable computers [Brewster et al. 2003; Sawhney and Schmandt 2000; Wilson et al. 2007]. Despite much explorative and empirical research, auditory displays have not yet developed a generic, robust theory compared to visual displays, such as visual menu navigation [Norman 1991] or visual search theory [Treisman and Gelade 1980; Treisman and Gormican 1988]. Considerably more research is still needed to set out such a theory for auditory displays.

In line with this thinking, Brewster [2008] pointed to two important areas where nonspeech sounds could be further incorporated. The first area is combining sound with other senses such as visual, tactile, and force-feedback. Multimodal interaction provides a rich experience by utilizing more of the user's senses. In addition, adding sound to interfaces not only improves performance, but also enhances subjective satisfaction and reduces perceived workload (see more detailed reviews in Section 3). The second area for sound incorporation is in mobile and wearable computing devices. The small or nonexistent screens of such devices cause many problems for viewing visual displays such as issues of glare and visibility. Auditory cues can be particularly effective in situations that require eyes-free interaction with these devices in a mobile context (e.g., while walking, cycling, driving, or with the device in a pocket).

Recently, devices such as mobile phones, PDAs, and MP3 players have started to adopt touch screen technology in order to enhance the user experience [Lee and Spence 2008b]. However, Norman and Nielsen [2010] pointed out that in the rush to develop gestural interfaces on touch screen devices, well-established usability standards of interaction design have been ignored. Given that touch screen devices generally lack tactile feedback and have input areas that overlap with the display area, the appropriate use of sounds can provide solutions or compensate for this phenomenon. Moreover, according to Oh et al. [2007], auditory feedback is the most effective modality in physical user interface satisfaction, followed by tactile and motion feedback. For example, one study demonstrated that even task-irrelevant sound can modulate tactile perception delivered via a touch screen [Lee and Spence 2008a]. Another study assessing subjective response to touch screens found that adding only haptic feedback to a visual display did not produce a reliable improvement, whereas adding audio feedback showed improvement over the visual-only display [Pitts et al. 2009]. Furthermore, in the same study, haptic effects were perceived as stronger in the presence of audible feedback. Some commercial products have begun to use audio in this way: the Apple iPod generates a click sound when users move their finger around the touchpad's circumference for separating the items or units. Nevertheless, there is still much room in touch screen interfaces to improve with auditory displays.

Based on this background, the present study investigated whether nonspeech sounds could enhance auditory menus on a touch screen mobile phone, interacting with other modalities. In particular, various gestures were considered as input mechanisms. Usability and overall user experience were measured, including navigation efficiency, perceived workload, and subjective satisfaction.

2. AUDITORY ENHANCEMENTS IN MOBILE DEVICES

Two main pieces of related work are described in this section. Research has been conducted on creating purely auditory interfaces in an attempt to provide a novel auditory-specific system. In comparison, other studies have added various nonspeech sounds to existing interfaces to improve usability. The current study focuses mostly on the latter.

2.1 Purely Auditory Interfaces

A few examples of purely auditory interfaces include SpeechSkimmer [Arons 1997], Nomadic Radio [Sawhney and Schmandt 2000], BlindSight [Li et al. 2008], and earPod [Zhao et al. 2007]. These interfaces each demonstrate how auditory menu navigation can be improved using speech and nonspeech sounds.

SpeechSkimmer is a touchpad system for interactively skimming recorded speech [Arons 1997]. It uses speech-processing techniques to allow users to hear recorded sounds quickly, and at several levels of detail. Through a manual input device, developed for that research purpose, a user controls the speed and detail level of the audio presentation. SpeechSkimmer reduces the listening time in four different ways by incorporating features such as time-compressed speech, pause shortening, automatic emphasis detection, and nonspeech audio feedback.

Nomadic Radio [Sawhney and Schmandt 2000] is a wearable computing platform for managing voice and text-based messages in a nomadic environment. It does not use a touch screen or touch pad. Instead, users wear a microphone and shoulder-mounted loudspeakers that provide a basic spatial audio environment (i.e., left and right). The system uses a context-based notification strategy. Thus, according to the users' focus of attention, it uses seven levels of auditory presentation. At the low level, when users are involved in other tasks, it uses ambient cues based on auditory icons [Gaver 1986], but as the level increases, it uses speech, expanding from a simple message summary up to the full text of a voicemail message.

The navigation functions used by SpeechSkimmer and Nomadic Radio are skimming and retrieving some information in long auditory contents such as novels, news, and email. Thus, they may be different from searching for the designated target item in an MP3 song list or an address book.

BlindSight is an attempt at eyes-free access to mobile phones [Li et al. 2008], and enables users to check, manage, or add data (e.g., a calendar or contact), while avoiding interruption of their phone conversations. Users interact without viewing the screen by using the phone keypad with auditory feedback. The auditory feedback is heard only by a BlindSight user, not by the counterpart. Investigators tried to make a complete, functional system, considering various hardware form factors and types of grip while minimizing mode switching. Consequently, their user evaluation of the interfaces while playing a driving game yielded an overall preference for BlindSight over the visual smartphone interface. However, they failed to show that BlindSight is more usable in terms of objective metrics (i.e., error rate and reaction time). In fact, BlindSight produced significantly more errors and was 200–300 ms slower per key press than the visual interface.

Another recent menu implementation that adopts auditory feedback in touch devices is earPod [Zhao et al. 2007]. It is a type of eyes-free menu navigation technique using touch input and reactive auditory feedback. earPod's auditory feedback involves three main characteristics. First, it uses interruptible audio; that is, each new playback stops the previous one. Second, it uses nonspeech audio like short mechanical click sounds when crossing the boundary from one menu item to the next and a camera shutter sound to confirm an item selection. Finally, it adopts binaural sound cues to reinforce users' cognitive mapping between menu items and spatial locations on the touchpad. In the evaluation study earPod yielded accuracy similar to that of an iPod-like visual menu, when applied to reasonably sized static menus. earPod even outperformed the visual menu in terms of efficiency (reaction time) within 30 minutes of practice. earPod is indeed a good example of eyes-free menu navigation, but it is debatable whether its efficiency mainly comes from the use of sounds. The main benefit of earPod derives from the fact that it does not need visuo-motor cooperation because

it does not have a visual display. Rather, the entire device is effectively an input area. In contrast, with an iPod-like visual menu, users have to combine visual search on the small screen with fine motor control. As described in the paper [Zhao et al. 2007], after moderate learning, users can directly tap the earPod target area based on motor memory, without having to slide their thumb on the circular touchpad and listening to each item. In consideration of this, earPod may be effective for navigation in a restricted hierarchical menu, but may not be optimal for navigation in a long menu like an address book or an MP3 song list, which does not allow the reliance on motor/spatial memory or direct access to menu items.

These attempts at a novel auditory interface demonstrate that auditory displays can stand alone as much as visual displays or sometimes can even outperform visual-only devices. However, the systems just discussed raise at least two major issues: they require a new type of device in order to fully implement those functions and thus, have cost and generalization issues. Consequently, these novel interfaces ask users to learn new interaction methods. An alternative to bypass these issues would be to simply add nonspeech sounds to current devices to which users are already accustomed.

2.2 Adding Nonspeech Sounds to an Existing Auditory Interface

Existing auditory interfaces are typically just speech-based menus. They are fairly straightforward, but have limited usability and efficiency [Brewster 2008]. There have been three main approaches to enhancing the basic text-to-speech (TTS) used in most auditory interfaces. These all tend to include adding sound cues before or concurrent with the spoken menu items. The most well-known types of enhancement cues are categorized as auditory icons [Gaver 1986], earcons [Blattner et al. 1989], and spearcons [Walker et al. 2006]. In addition to these, a new concept, the spindex [Jeon and Walker 2011], has recently been introduced.

2.2.1 Auditory Icons. Auditory icons are sounds that represent an object by capturing the object's essential features, such as functions and events [Gaver 1986]. Thus, they are a type of caricature of naturally occurring sounds such as bumps, scrapes, or even files "hitting" mailboxes. Auditory icons can denote many objects in devices more clearly than some other auditory cues because the relation between the sound and the data is often very direct: for example, a typing sound can represent a typewriter or even a printer. Thus, auditory icons typically require little training and are easily learned. Leveraging these advantages, Gaver [1989] created an auditory icon-enhanced desktop. Other researchers have attempted to convert GUIs to nonvisual interfaces using auditory icons [Mynatt 1997; Mynatt and Weber 1994]. Auditory icons are also suited for presenting dimensional data such as the magnitude of some value. Moreover, they can categorize objects into distinct families. Conversely, it is sometimes difficult to match all functions of a device with proper auditory icons. For example, it may be difficult to create a sound that clearly conveys the idea of "save" or "unit change" [Palladino and Walker 2007, 2008a]. As a result, there have been few systematic uses of auditory icons in auditory interfaces in general, and certainly fewer in auditory menus. However, one could apply auditory icons to auditory menu navigation as well. Consider an address book list on a mobile phone. One could record a friend's voice and register it as feedback for one of the items on the list. This would be similar to the address book on many recent mobile phones, in which users can associate a photo with a person's name listed in the menu. Use of auditory icons within address book lists might enhance users' subjective satisfaction, but might not facilitate navigation efficiency.

2.2.2 Earcons. Earcons are nonspeech audio representations, which are short, rhythmic musical motives with variable intensity, timbre and register, used to provide information to a user about some objects, operations or interactions [Blattner et al. 1989]. Since earcons use an arbitrary mapping between the sound and the object, they can be analogous to a language or a symbolic sign. This arbitrary mapping between earcons and represented items means that earcons can be applied to any type of menu; that is, earcons can represent nearly any concept. However, this flexibility can also be a weakness because the arbitrary mapping of earcons to concepts requires user training. To make more intuitive and logical earcons, detailed guidelines (e.g., timbre, pitch and register, rhythm, duration, and tempo, and intensity) have been proposed based on empricial studies [Brewster 2008]. Earcons can also depict hierarchical menus by logically varying musical attributes. For example, investigators 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]. The results showed improved usability and browsing experience. However, when a new item has to be inserted in a fixed menu structure, it can be difficult to create a new branch sound. Moreover, the structural framework of earcons can be congruent with logical hierarchical menus, whereas it seems difficult to apply them to one-dimensional long menus. If the menu includes hundreds of items, it might be hard for users to memorize those arbitrary mappings. For a recent and more detailed overview of auditory icons and earcons, see Absar and Guastavino [2008].

2.2.3 Spearcons. Spearcons are brief sounds that are produced by speeding up spoken phrases, even to the point where the resulting sound might no longer be comprehensible as a particular spoken word [Walker et al. 2006]. These sounds are analogous to fingerprints because of the unique 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." Another example is if a new name is added to a contact list, the spearcon can be quickly and dynamically created as needed. Also, spearcons are easy to learn because they derive from the original speech [Palladino and Walker 2007]. Spearcons have been shown to enhance performance and preference for auditory menus [Palladino and Walker 2008a, 2008b; Walker and Kogan 2009]. See Walker et al. [in press] for a recent overview of spearcons.

2.2.4 Spindex. An Auditory Index Based on Speech Sounds. A spindex (i.e., speech index) is a brief nonspeech auditory cue based on the pronunciation of the first letter of each menu item [Jeon and Walker 2011]. To illustrate, the spindex cue for "Superstar" would be the sound /es/ or even /s/ based on the spoken sound of "S", the first letter of the item. 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. Also, in the song list or address book of many electronic devices, an alphabetical menu is typically the default setting.

The human factors literature provides a framework about human motions in a general control task (e.g., knob rotation) and this may inform menu navigation studies. In gross-adjustment movement, the operator brings the controlled element to the approximate desired position. This 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]. Similarly, in a search process such as scrolling through an address book on an electronic device (visually or auditorily), there may be

two discrete stages. One is rough navigation and the other is fine navigation [Klante 2004]. In the rough navigation stage, users pass over the nontarget alphabet groups by glancing at their initials. For example, users quickly jump to the "S" section to find "Superstar". Then, once users reach a target zone ('S's), they begin fine navigation, identifying their current location and carefully tuning their search. In auditory menus, people cannot jump around as easily, given the temporal limitations of typical spoken menus. Nonetheless, users still want to pass over the nontarget alphabetical groups as fast as possible.

A previous study on the desktop simulator of a mobile phone showed that the initials of the alphabet of the list—the key structure of a spindex—can provide users with enough information to quickly sort out the nontarget items [Jeon and Walker 2011]. Additionally, the benefits of a spindex can be even more clearly seen in a long menu with a larger number of items (e.g., 150 items) compared to a short menu (50 items) even though the benefits of the spindex cues were reliably demonstrated in both menus. The subsequent study showed that visually impaired users could also benefit from adding spindex cues to plain TTS menus, and they preferred the use of a spindex over plain TTS menus. Furthermore, a pre-made set of spindex sounds does not require a lot of storage or numerous additional audio files and can be added by a simple software update. Finally, because spindex cues are part of the original word and are natural—based on speech sounds—they do not require much training [Jeon and Walker 2011].

3. USER EXPERIENCE METRICS FOR AUDITORY INTERFACES

From the Human Factors and Usability tradition, objective measures for the interface assessment have been well developed. As for auditory interfaces, objective measures have also been emphasized. However, nowadays, the importance of the subjective acceptance and preference level of user interfaces has rapidly been increasing in user experience design circles. For example, Norman [2004] has stressed the importance of visceral design and proposed that an attractive and natural design can improve usability as well as affective satisfaction [Norman 2004, 2007]. Even though many researchers point out that aesthetic and annoyance issues are more important in auditory displays than in visual displays [Brewster 2008; Davison and Walker 2008; Kramer 1994; Nees and Walker 2009], to date, research has mainly focused on performance issues. A fairly early study suggested that the nature of sound aesthetics is independent of performance outcomes [Edworthy 1998]. That is, users might turn off an annoying sound even though the presence of that sound enhances performance with a system or device. Likewise, system sounds can improve the aesthetic experience of an interface without changing performance with the system [Nees and Walker 2009]. Therefore, it is evident that developing universal evaluation metrics in terms of both objective and subjective aspects is crucial to the success of auditory interfaces. From this standpoint, we have attempted to survey a variety of dependent measures for auditory display from the literature.

3.1 Objective Evaluation Metrics

Performance improvement by the addition of auditory cues in menu navigation tasks has been studied by several metrics such as *reaction time*, *number of key presses*, and *error rate*. In earlier work, structured earcons showed a superior *learning rate* (i.e., recognition proportion of mapped visual objects) to nonorganized sound [Brewster et al. 1992]. In research on sonically enhanced buttons and scrollbars, results showed reduced time to recover from errors compared to no-sound conditions [Brewster 1997]. Along the same line, in an experiment with sonified mobile phones, earcons improved

the performance of navigational tasks in terms of the number of errors made and the number of key presses taken to complete the given tasks [Leplâtre and Brewster 2000]. Also, in a hierarchical menu experiment, participants with earcons could identify their location with over 80% accuracy [Brewster et al. 1996]. A study on combining earcons with spoken menu items in a hierarchical menu indicated that the use of earcons improves task performance by reducing the number of keystrokes required, while increasing the time spent for each task [Vargas and Anderson 2003]. Recent research on the addition of auditory scroll bars has demonstrated the potential benefits of applying simple music tones proportionally to each group of list items. The results showed reduced error rates in target search [Yalla and Walker 2008].

Spearcons and spindexes have also shown promising results in objective metrics 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 conditions. 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 another study [Palladino and Walker 2008a], in the 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. In a recent study, undergraduate students showed better performance in navigation time and learning rate with TTS + spindex (mean time of navigation: 10.3 seconds) than with TTS-alone (11.6 seconds) in both visuals-on and visuals-off conditions [Jeon and Walker 2011]. Additional experiments with visually impaired users showed similar results with more efficiency: The spindex + TTS condition (mean time of navigation: 21.3 seconds) enhanced navigation time compared to the TTS-only condition (28.1 seconds).

3.2 Subjective Evaluation Metrics

Subjective evaluation factors can largely be categorized as *subjective preference* and *perceived workload*. Using nonspeech sounds increases *preference for the system*. Experimental comparison of complex and simple sounds in a mobile phone menu demonstrated that a simpler sound was preferred and showed enhanced performance over a complex sound [Marila 2002]. In that research, the researcher posed questions such as "Would you like to have these sounds in your own mobile phone?" and "How distracting and irritating are the sounds?" However, the first question might be affected by sound quality or other confounding variables. Another mobile phone study focused more on subjective reactions of the users and included related questions in their questionnaire [Helle et al. 2001]. Their questions involved first impression, annoyance, aesthetical/musical judgment, opinion of the lengths of sounds, suitability to corresponding functions, effect of usage, and usefulness. More importantly, with respect to the preference evaluations in auditory displays, researchers have to measure annoyance as well as preference. Since users cannot avert their ears from sound, annoying sounds need to be identified. Recently, Andersen and Zhai [2010] showed that adding auditory feedback (i.e., continuous tone, algorithmic rhythmic feedback, and song playback) could make pen-gesture production more stimulating than no-auditory conditions by using three subjective experience rating dimensions: terrible-wonderful, frustrating-satisfying, and dull-stimulating.

Adding nonspeech sounds not only increases preference but also decreases users' subjective workload. In subsequent experiments, sonically enhanced buttons and scrollbars reduced subjective workload as compared to their silent counterparts in a desktop computer [Brewster 1997] and in a pen-based handheld computer [Brewster 2002]. Recent work with spearcons and spindexes began to study more systematically

Table I. Usability Evaluation Metrics Used in This Study

	Objective Metrics			Subjective Metrics		
Dependent Measure	Navigation Efficiency	Learning Rate	Accuracy	Perceived Performance	Subjective Preference	Perceived Workload
Methods	Reaction time (milli second)	RT change according to block	Number of errors	Likert Scale (0∼10)		Electronic NASA-TLX
				Effective and helpful	Likable, fun, and annoying	

the subjective improvements to auditory menus. In a mobile phone study with spearcons and TTS, higher rankings were provided for all audio cues when spearcons were included, both in visual and nonvisual conditions [Walker and Kogan 2009]. Likewise, spindex cues were significantly favored over TTS-alone by undergraduate students (e.g., 8.84 with spindex cues and 5.08 with TTS-alone on 0 to10 functionally helpful scale) and visually impaired users (e.g., 7.77 with spindex cues and 6.08 with TTS-alone) [Jeon and Walker 2011]. In a dual task context such as navigating a menu while playing a driving-like game, all of the sound conditions reduced subjective workload score for overall tasks compared with the no sound condition [Jeon et al. 2009]. Even the spindex + TTS and the spindex + spearcon + TTS condition showed marginally lower *perceived workload* than TTS-only condition. Additionally, perceived performance of the use of the sound can be measured. For example, Jeon and Walker [2011] used functionally helpful and appropriate as perceived performance scale in their previous spindex research. Clearly, there are a large number of metrics that could be used for subjective experience measures. Combining as many factors as possible, objective and subjective metrics used in the current study are shown in Table I.

4. MOTIVATIONS FOR THE CURRENT STUDY AND HYPOTHESES

4.1 Motivations

The present study used the second deployment strategy discussed at the outset, namely the addition of nonspeech sounds to an existing system, via small software tweaks. This strategy can be more universal and cost-effective than making a new auditory device per se. To date, despite attempts to add nonspeech sounds to touch screen interfaces, there has been no research on the use of nonspeech sounds for facilitating the new interaction styles that are becoming more common on touch devices. These new interaction styles include sliding a finger on the full touch screen ("flicking") or circling a finger on the screen ("wheeling"), reminiscent of an iPod scroll wheel. To test these possibilities, one of the suitable spindex variants called attenuated spindex, which contains attenuated cues after the first menu item in a letter category, was selected as an advanced auditory cue type in this study (for the detailed design of the attenuated spindex, see Section 5.3.3 and for other alternative designs of a spindex, see Jeon and Walker [2011]).

Earcons or auditory scroll bars could also have been used, but they have several issues for real applications in the one dimensional menu system. First of all, there is a mapping issue; earcons use arbitrary mappings between sounds and items, so designers have to figure out the best solution for mapping issues such as motive patterns, the number of music notes, and polarity. Another issue arises from that mapping problem; users cannot intuitively determine the meaning of the sound mapping. Thus, they have to learn the meaning of the mapping or be trained. Finally, applying musical sounds for interfaces frequently goes beyond UI designers' job descriptions and skill sets. They might need a specialist sound designer or a musician for good sound implementation.

Auditory scroll bars also have similar issues such as polarity and choice of timbre. Although the meaning of the mapping of auditory scroll bars might be more intuitive than earcons, the meaning of each sound is rather relative than absolute, thereby less intuitive than a spindex.

Spearcons might also be a strong candidate to be adopted for this study. Spearcons have shown positive results in a type of menu list navigation and could be automatically generated. However, in the previous navigation experiment with 150-item lists [Jeon et al. 2009], the spindex-enhanced TTS menu outperformed the spearcon-enhanced condition. Moreover, for input gestures such as wheeling and flicking in the current study, spearcons are still too long to implement in practical applications. Therefore, in the following experiment, we focus on a spindex-enhanced TTS menu vs. a TTS-only menu. Again, the spindex is one of the shortest nonspeech sound enhancement cues and can be made on the fly by adding pre-recorded spindex files to the new menu items.

4.2 Hypotheses

Based on previous spindex research [Jeon and Walker 2011], the spindex is anticipated to be better than TTS alone in both objective and subjective measurements. Target search time, number of errors, and required learning for the TTS + spindex condition should be lower than those of TTS alone for all input gestures particularly in the visuals-off condition. Spindex cues should be favored over plain TTS on perceived performance, subjective preference, and perceived workload evaluation in the visuals-on and visuals-off conditions.

To test these hypotheses empirically, six groups of undergraduate participants navigated an auditorily rendered song list menu (with or without spindex) using different input gestures–tapping, wheeling, and flicking–within visuals-on and visuals-off conditions.

5. METHOD

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.

5.1 Participants

A total of 122 undergraduate students participated 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, gender, handedness, and previous experience with touch screen devices (mean age = 19.7; 56 male, 66 female; 14 left, 108 right handed; mean number of years of touch screen device experience = 1.6).

5.2 Apparatus

Stimuli were presented using a Google Nexus One HTC1, an Android smartphone (version 2.2.2) with a 3.75 inch resistive touch screen panel. The internal sound chip was used for sound rendering. Participants listened to auditory stimuli using Sennheiser HD 202 headphones plugged into the phone's audio jack, and adjusted for fit and comfort.

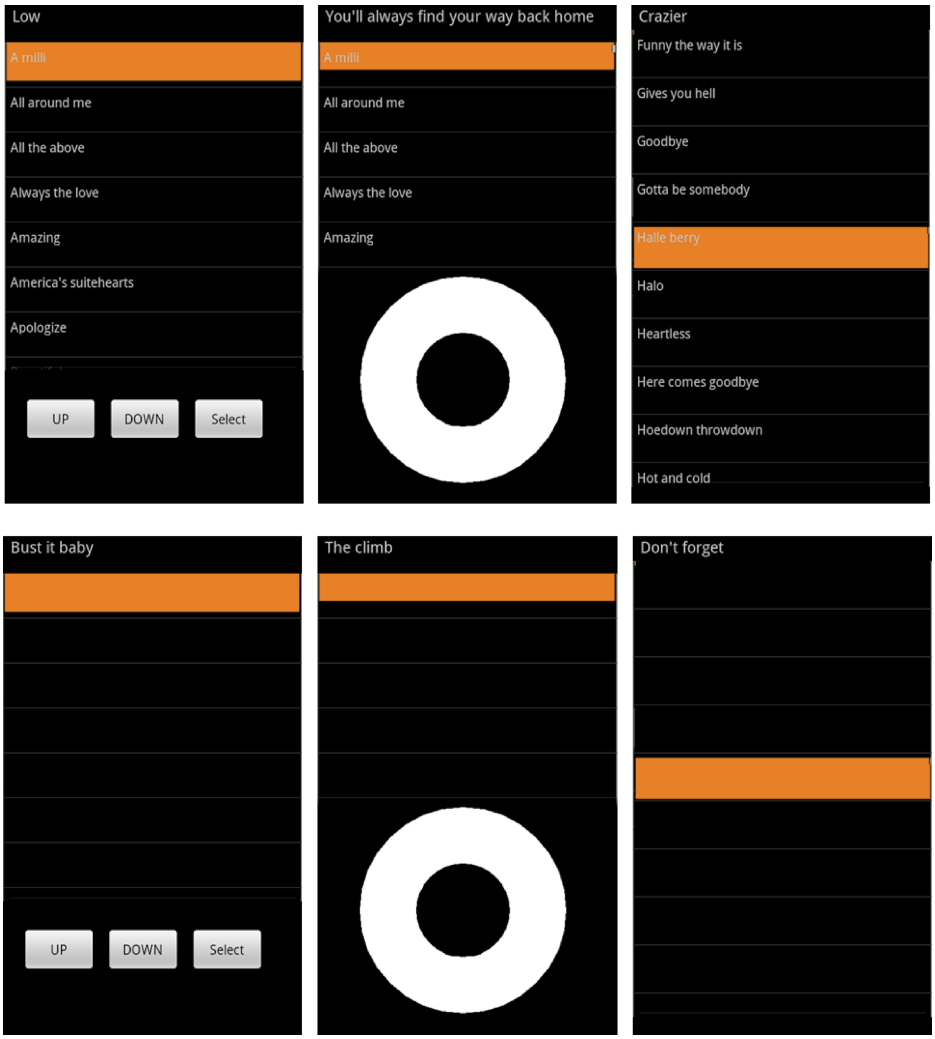

Fig. 1. Screen Visual menus for each input gesture style (Visuals-on condition on the top and Visuals-off condition on the bottom): the Tapping, the Wheeling, and the Flicking condition from left to right. The target song title was visually displayed on the top of the screen in all conditions.

5.3 Stimuli

5.3.1 MP3 Song List Menu. An MP3 song list menu was created with 150 song titles gathered from the Billboard Hot 100 & Pop 100 (2009, 2009)[1] and iTunes Top 100[2]. Each visual menu (see Figure 1) was implemented in Java using the Android SDK programming tool for use as an application ("app") on the smartphone.

The menu items were presented in alphabetical order. For each type of input gesture, participants were able to scroll upward and downward in the menu by (1) tapping on "up" or "down" button areas on the bottom of the screen (Tapping condition),

[1]http://www.billboard.com/bbcom/index.jsp
[2]http://www.apple.com/itunes/top-100/songs/

(2) wheeling on a marked circular area at the bottom of the screen (Wheeling condition), or (3) flicking the list in the desired scrolling direction (Flicking condition) on the touch screen device.

In the Tapping condition, the screen displayed seven song titles in addition to the target item, which was presented on the top line of the screen. The first line of the list was the selection area as indicated by the orange bar; this selection area did not move on the screen in all conditions. When a menu item fell into this area, the device spoke out the item and participants could select the item by tapping a "select" button area. Tapping the "down" button on the screen moved the selection down the list one item by moving menu items up by one menu position.

In the Wheeling condition, there were five lines of song titles underneath the target item on the top. The smaller number of visible lines was necessary to accommodate the wheel area. The location of the selection area was the same as for the Tapping condition. Participants could select an item by touching the center circle of the wheel as it is normally done in such a device. The circular wheeling area was divided into four sections. Thus, sliding the finger clockwise one quarter of the circle moved the list items up by one menu position, so that the item presented in the orange bar came from lower on the list.

In the Flicking condition, there were ten lines of song titles under the target item. The difference in number of visually displayed menu items might add a confounding variable for comparisons among gestures in the Visuals-on condition. On the other hand, it can be more practical because they are quite similar to the real devices. The selection area was located in the fifth line. Menu position was moved by several items, with the exact number depending on the strength of flicking (from one to two items to hundreds of items). However, it is unlikely to get to the last item with one flick (i.e., has no inertia). In all conditions, if participants reached the top or bottom of the menu, the menu list did not wrap around. A more detailed description of gesture interaction for the experiment can be found in 5.5 Procedure.

5.3.2 Text-to-Speech. TTS files (.wav) were generated for all of the song titles using the AT&T Labs TTS Demo program with the male voice Mike-US-English[3]. Menu items in the TTS-only condition simply consisted of an auditory TTS phrase that played for each menu item as participants navigated the song list. All auditory stimuli were interruptible so that when the next item is played, the previous one is stopped. All of the sounds (speech and nonspeech) were prerecorded as a separate file (16000Hz, 16-bit, Mono) for each menu item. The TTS phrases lasted on average 1.07 seconds (range 0.44–2.40 sec).

5.3.3 Spindex Cues. Since the attenuated spindex design has been shown to be the most preferred and simplest to implement with equal performance to other designs [Jeon and Walker 2011], it was used in this experiment. The attenuated version of the spindex contains cues that are attenuated by 20 dB after the first menu item in a letter category (e.g., AAAA. . . BBBB. . . CCCC. . .). Spindex cues were created by generating TTS files for each letter (e.g., "A"). Each spindex cue pronounced one letter of the alphabet. In the cases of letters which generate a longer pronunciation such as A, F, H, I, J, K, S, W, X, and Y, the longer sound was used for the first cues, then the shorter sound was used for the subsequent cues in that letter group (e.g., /es/ then /s/ for "S," see Table II). The subsequent shorter cues were made of part of the first cues with attenuation except A (from "Agora"), H (from "Harbor"), I (from "Israel"), W (from "What"), and Y (from "Yoyo"). The lengths of the subsequent cues and the words for them derived from pilot tests.

[3]http://www.research.att.com/~ttsweb/tts/demo.php

Table II.

A spindex cue set used in this experiment. Cases in which the pronunciation for the 1st and 2nd cues is different are emphasized in bold.

1st	**ei**	bi	si	di	i	**ef**	di	**eit**	**ai**	**dei**	**Kei**	el	em
2nd	**a**	bi	si	di	i	**f**	di	**h**	**i**	**d**	**k**	el	em
1st	En	o	pi	kju	a(r)	**es**	ti	ju	vi	**dblju**	**eks**	**wai**	zi
2nd	en	o	pi	kju	a(r)	**s**	ti	ju	vi	**wa**	**s**	**yo**	zi

Note that this distinction between the longer cues and the shorter cues is different from the previous study [Jeon and Walker 2011]. Spindex cues used in the list were presented before the TTS cues, such that, the "All around me" target item would produce the sound "a"-pause-"All around me." The interval between the spindex and the TTS was 250ms as used in the previous study. If participants tapped, wheeled, or flicked the appropriate area quickly, the spindex cues were generated preemptively, without a lag between items. The first spindex cues lasted on average 0.46 seconds (range 0.20–0.59 sec) and the subsequent attenuated spindex cues lasted on average 0.28 seconds (range 0.10–0.45 sec).

5.4 Experimental Design

A split-plot design was used in this experiment. The two between-subjects variables included input gesture (tapping, wheeling, and flicking) and visual display mode (on and off). The two within-subjects variables included auditory cue type (TTS-only and TTS + spindex) and block (1–3). Our experiment was designed in this way to focus more on the intra-participant's effects of auditory cue type and its learning effects, considering task completion time and plausible fatigue.

5.5 Procedure

After the informed consent procedure, participants were randomly assigned to one of the six groups (3 input gesture x 2 visual mode). According to the assigned condition, the experimenter explained the detailed procedure and demonstrated how to interact with the menu system on the phone. Participants wore the headphones and could adjust them for fit and comfort, as well as the volume level on the phone. Next, they had a short practice session (10–30 seconds) for one or two trials with TTS cues in order to be familiar with how to control the device. Then, the experimental session began. The overall goal of the participants was to reach the target song title in the song list menu as fast as possible and select it by touching the selection area.

In the experiment, each block included 15 trials of different songs as targets. To evenly spread out the target menu positions across conditions, one target in each block was randomly selected from menu items 1–10 (Bin 1), one from 11–20 (Bin 2), and so on to 141–150 (Bin 15). Moreover, the order of these 15 targets was also randomized in the block. Each condition was composed of three successive blocks. Every participant completed two conditions (TTS-only and TTS + spindex), which were counterbalanced across participants.

In each trial, the target name was visually presented at the top of the phone screen (Figure 1). In the Visuals-off condition, the song list was not shown, but the target item was still presented visually. The timer started when participants first touched the available area. Participants could navigate through the menu system to find the assigned target song with their preferred hand and fingers. In the Tapping condition, participants tapped the up and down button area of the touch screen to navigate the list menu and touched the selection button area. In the Wheeling condition, they used their finger to wheel around the circular area and to press the center circle selection

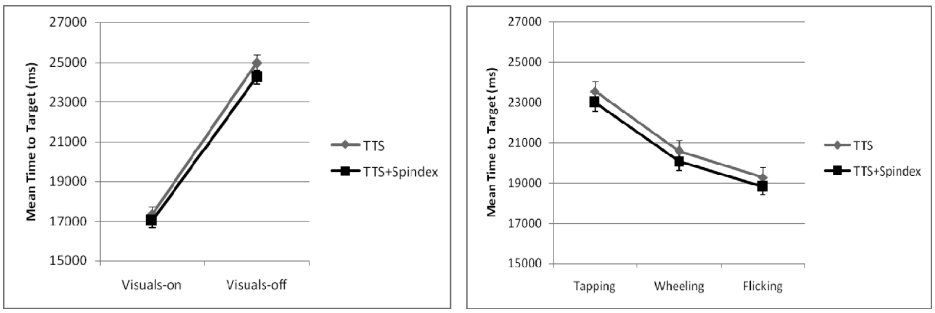

Fig. 2. Time to target for visual mode and auditory cue type (left) and time to target for input gesture type and auditory cue type (right). The enhancement effect of the spindex showed consistently in both Visuals-on and Visuals-off conditions across all input gestures.

area. In the Flicking condition, they flicked the list area using their finger to navigate the list and touched the selected item itself. Pressing the selection area (Tapping and Wheeling conditions) or the focused item itself on the orange bar (Flicking condition) indicated the selection of the requested target and recorded the end time. Besides, pressing the menu item instead of the selection area (Tapping and Wheeling conditions) or items outside the orange bar (Flicking condition) did not work as a selection. This procedure was repeated for all 15 targets in a block. Then, participants were shown a screen that indicated that the next block of 15 trials was ready to start. When the participants were ready, they pressed the OK button on the screen and started the next block. After three blocks of the first condition, participants completed the electronic version of the NASA TLX [Hart 2006] on a desktop computer to report their perceived workload for the navigation task. Then, they repeated the same procedure for the second condition (15 trials x 3 blocks and NASA TLX) with either TTS-only or TTS + spindex. After finishing both auditory cue conditions, participants filled out a short subjective questionnaire. An eleven-point Likert-type scale was used for the self-rated levels of perceived performance (how effective and functionally helpful) and subjective preference (how likable, fun, and annoying) with regards to auditory cues. Finally, participants provided comments on the study.

6. RESULTS

To look at representative objective and subjective evaluation results in one dimension, a 3 (Input gesture) x 2 (Visual mode) x 2 (Auditory cue type) multivariate analysis of variance (MANOVA) was conducted, considering both time to target and subjective workload score (NASA TLX) as dependent variables. The MANOVA found a significant positive effect of adding spindex, $F(2, 115) = 3.818$, $p < .05$, Pillai's Trace = .062, $\eta_p^2 = .06$. There was no interaction or trade-off between the two dependent variables, so subsequent univariate tests (adding block as a variable) for each dependent measure are described in the following sections.

6.1 Objective Evaluation

6.1.1 Accuracy. Errors in both the TTS condition ($M = 2.52$, $SD = 2.81$) and the TTS + spindex condition ($M = 2.51$, $SD = 2.81$) were minimal and not significantly different. Therefore, we will focus more on the mean time to target and learning rate in the objective evaluation analyses.

6.1.2 Navigation Efficiency and Learning Rate. The time to target results are depicted in Figures 2–4. Results were analyzed with a 3 (Input gesture) x 2 (Visual mode) x 2

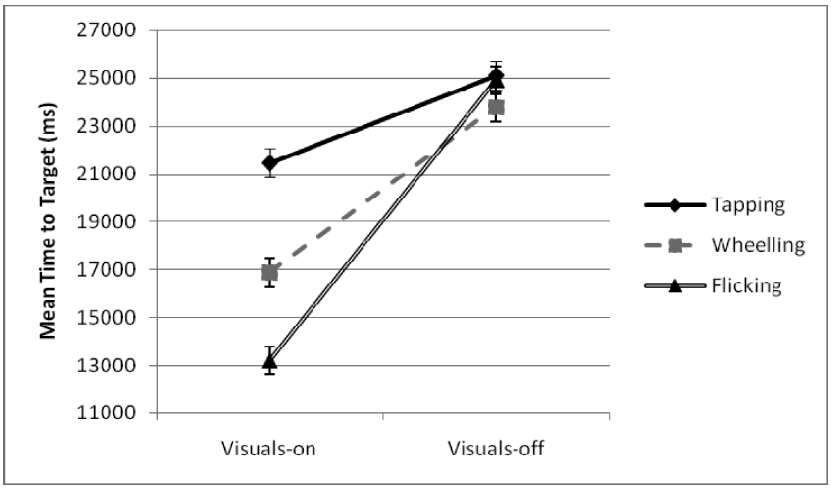

Fig. 3. The interaction between visual mode and input gesture. The Flicking condition showed a sharp increase in time to target in the Visuals-off condition 1.

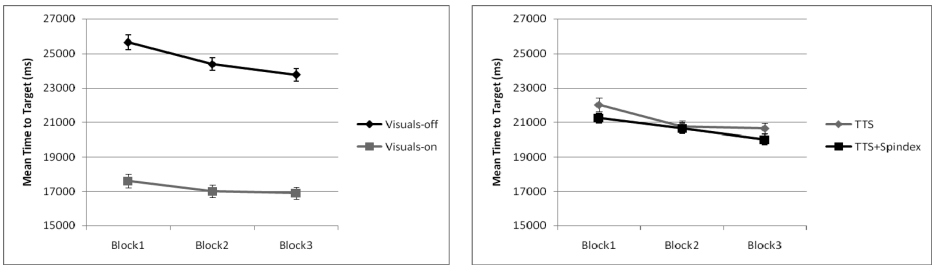

Fig. 4. The interaction between block and visual mode (left). A greater learning effect occurred in the Visuals-off condition than in the Visuals-on condition. Time to target for block and auditory cue type (right). There was no interaction between block and auditory cue type. However, more learning took place between Block 2 and Block 3 in the TTS + spindex condition compared to the TTS condition.

(Auditory cue type) x 3 (Block) repeated measures analysis of variance (ANOVA). The analysis revealed that participants reached the target item significantly faster in the TTS + spindex condition (M = 20637, SD = 3225) than in the TTS-only condition (M = 21145, SD =2806), $F(1, 116) = 4.04$, $p < .05$, $\eta_p^2 = .03$. This spindex enhancement effect appeared consistently, regardless of visual mode and input gesture (see Figure 2). Participants in the Visuals-on condition (M = 17177, SD = 2679) had faster search times than those in the Visuals-off condition (M = 24604, SD = 2679), $F(1, 116) = 234.44$, $p < .001$, $\eta_p^2 = .67$. Also, the main effect for block (i.e., practice) was statistically significant $F(1.86, 215.2) = 22.79$, $p < .001$, $\eta_p^2 = .16$. In addition, the input gesture showed a significant main effect, $F(2, 116) = 26.53$, $p < .001$, $\eta_p^2 = .31$. Pairwise comparisons revealed that the Flicking condition (M = 19058, SD = 2677) was significantly faster than the Wheeling condition (M = 20339, SD = 2694), ($p < .05$), which was significantly faster than the Tapping condition (M = 23275, SD = 2694), ($p < .001$). However, this main effect was moderated by the interaction between input gesture and visual mode, $F(2, 116) = 23.82$, $p < .001$, $\eta_p^2 = .29$. This occurred because the Flicking condition showed a sharper increase in navigation time in the Visuals-off condition than other input gestures (Figure 3). In the Visuals-off condition, time to target for

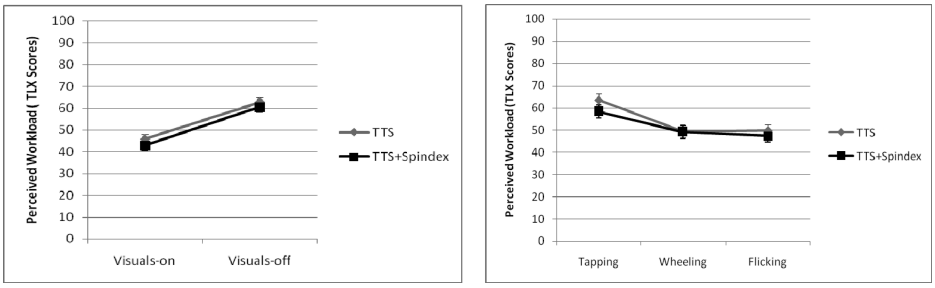

Fig. 5. Perceived workload for visual mode and auditory cue type (left) and perceived workload for input gesture type and auditory cue type (right). The spindex consistently reduced perceived workload both in Visuals-on and Visuals-off conditions across all input gestures.

the flicking gesture (M = 24929) increased to the same level as the tapping gesture (M = 25094), $t(39) = .17$, $p = .87$. The interaction between block and visual mode was also significant, $F(1.86, 232) = 4.41$, $p < .05$, $\eta_p^2 = .04$ (Figure 4). This interaction reflects the fact that as block number increased, more learning occurred in the Visuals-off condition than in the Visuals-on condition. The interaction between auditory cue type and block was not significant. However, we can find interesting results. Between Block 1 and 2, there was a learning effect in both auditory cue types. While in the TTS condition, there was no more learning between Block 2 (M = 20721) and Block 3 (M = 20630), $t(121) = .38$, $p = .70$, in the TTS + spindex condition, there was more learning effect between Block 2 (M = 20624) and Block 3 (M = 19977), $t(121) = 2.52$, $p < .05$ (Figure 4). In sum, the TTS condition showed higher reaction time from the first block and the learning effect due to familiarity took place only at the first block. In contrast, the TTS + spindex condition showed lower reaction time from the first block and the benefit of adding spindex increased more as experience increased.

6.2 Subjective Evaluation

6.2.1 Perceived Workload. Perceived workload scores (NASA TLX) were also analyzed with a 3 (Input gesture) x 2 (Visual mode) x 2 (Auditory cue type) repeated measures ANOVA. Perceived workload results are depicted in Figures 5–7. The analysis revealed that adding spindex cues to TTS (M = 51.63, SD = 18.56) reduced perceived workload significantly, compared to the plain TTS condition (M = 54.23, SD = 17.67), $F(1, 116) = 4.09$, $p < .05$, $\eta_p^2 = .03$. The spindex enhancement effect on workload appeared consistently, regardless of visual mode and input gesture (see Figure 5). Participants in the Visuals-on condition (M = 44.29, SD = 16.64) rated perceived workload significantly lower than those in the Visuals-off condition (M = 61.57, SD = 16.64), $F(1, 116) = 32.83$, $p < .001$, $\eta_p^2 = .22$. In addition, the input gesture showed a significant main effect on perceived workload, $F(2, 116) = 7.04$, $p = .001$, $\eta_p^2 = .11$. Pairwise comparisons revealed that the Tapping condition (M = 60.94, SD = 16.63) showed significantly higher workload than the Wheeling condition (M = 49.26, SD = 16.63), ($p = .002$) and the Flicking condition (M = 48.59, SD = 16.66), ($p = .001$). However, the Wheeling and the Flicking conditions were not significantly different from each other ($p > .05$). This main effect was moderated by the interaction between input gesture and visual mode, $F(2, 116) = 3.96$, $p < .05$, $\eta_p^2 = .06$. This occurred because the Flicking condition showed a sharp increase in workload in the Visuals-off condition (Figure 6), similar to the results for navigation time. Workload scores for the flicking type (M = 62.90) in the Visuals-off

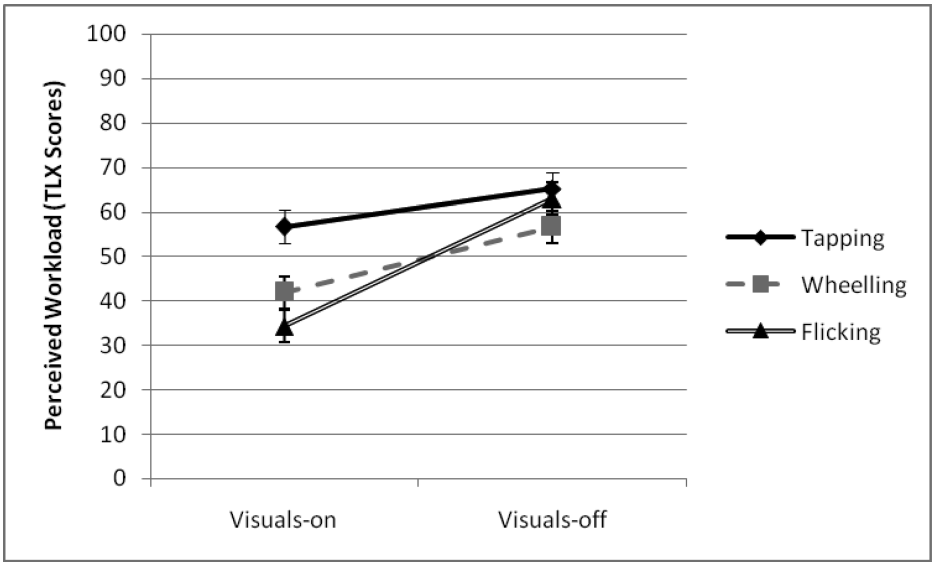

Fig. 6. The interaction between visual mode and input gesture. The flicking condition showed a sharp increase in perceived workload in the Visuals-off condition.

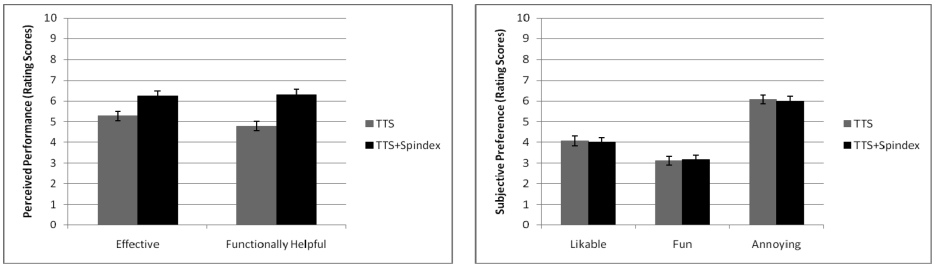

Fig. 7. Perceived performance for auditory cue type (left). Participants rated the TTS + spindex condition significantly higher than the TTS condition on both perceived performance scales. Subjective preference for auditory cue type (right). There was no difference on subjective preference scores between auditory cue types.

condition increased to the same level as in the Tapping condition (M = 65.17), $t(39)$ = .46, p = .65.

6.2.2 Perceived Performance. Perceived performance was measured by rating scores on the Effective and the Functionally Helpful scale. Paired-samples t-tests showed that participants rated the TTS + spindex condition (M = 6.23, SD = 2.56) significantly higher than the TTS condition (M = 5.28, SD = 2.48), t(121) = −3.77, p < .001 on the Effective scale. Similarly, on the Functionally Helpful scale, the TTS + spindex condition (M = 6.30, SD = 2.68) was significantly higher than the TTS condition (M = 4.79, SD = 2.65), t(121) = −5.58, p < .001.

6.2.3 Subjective Performance. In this study, subjective preference was also measured by Likert type scales including Likable, Fun, and Annoying. However, for the subjective preference data, there was no statistically significant difference between auditory cue types on the Likable scale, t(121) = 0.21, p = .83, on the Fun scale t(121) = −0.29, p = .77, or on the Annoying scale t(121) = 0.30, p = .76.

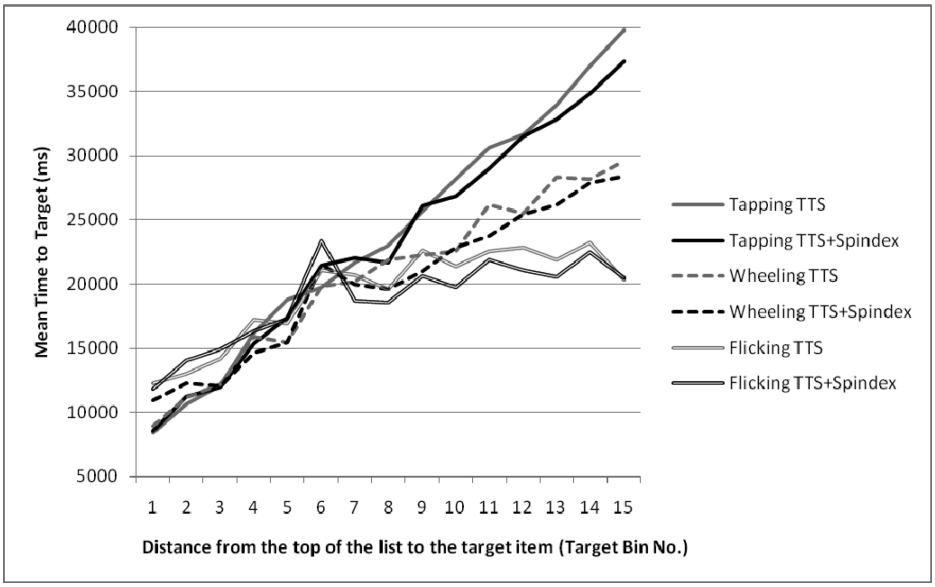

Fig. 8. Time to target as a function of target distance. See the text for more detailed analysis.

6.3 Navigation Behavior Pattern Analysis

In addition to the objective and subjective data analyses, we analyzed participants' navigation behavior patterns using navigation time data, according to the interaction between the input gesture and the output mode (i.e., auditory cue type). These analyses revealed more clearly where and how the spindex cues facilitated navigation efficiency.

6.3.1 Where to Facilitate Navigation. Figure 8 plots the mean time to target as a function of distance from the top of the menu to the target item (i.e., bin number of the target). Overall, as expected, as the target distance increased, the navigation time increased. Also, the disparity between the navigation times for the three input gestures also increased as the bin number increased. Regression lines for each input gesture were created using the mean time to target by the target distance. The Tapping condition was best fit to a linear model (TTS condition: R^2 = .995, y = 2161x + 6532; TTS + spindex condition: R^2 = .990, y = 2005x + 7166). The Wheeling condition also showed a linear increase (TTS condition: R^2 = .968, y = 1447x + 8954; TTS + spindex condition: R^2 = .954, y = 1279x + 9885), but the slope of the Tapping condition is steeper than that of the Wheeling condition. This is because while one tap moves only one item, one "wheel" moves four items. In contrast to other input gestures, the Flicking condition showed a power function increase in speed with the distance to the target (TTS condition:R^2 = .886, y = $11758x^{0.26}$, TTS + spindex condition: R^2 = .839, y = $12133x^{0.23}$). That is, in the Flicking condition, participants could get to a distant point faster by increasing flicking strength as opposed to other gesture conditions where the number of input actions had to be increased for the far target. For the same reason, in Bin 15 of the Flicking condition, both auditory conditions showed a decrease in navigation time. Participants might get to the last area with strong flicking, without thorough scanning. On the other hand, the data from Bin 1 showed that flicking had some starting cost, more than the other input gestures. In all input gestures, the slope of the TTS + spindex condition was less steep than that of the TTS condition. Note that in

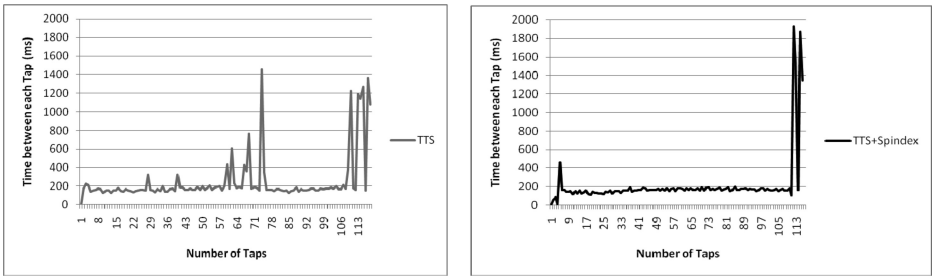

Fig. 9. Time between each tap as a function of the number of taps: Participant A in the Visuals-off TTS condition Block 1 Trial 7 Target No. 118 (left) and in the Visuals-off TTS + spindex condition Block 3 Trial 14 Target No. 113 (right).

near bins (e.g., Bin 1, 2, 3), a spindex effect did not show well, but as the target distance increased (e.g., after Bin 7), the spindex effect appeared more clearly. The increase in navigation time of the spindex condition in Bin 6 and 7 was due to the fact that in those bins, 17 items started with "i" (which is much more than the average number of songs started with each alphabet character, 6.5), so participants had to listen to the TTS part more in those zones than in other bins. See more detailed analysis for the trade-offs of the spindex according to the number of menu items in Section 7.4.

6.3.2 How to Facilitate Navigation. As seen above, spindex cues were more helpful for farther targets than closer targets. How, then, did spindex cues make navigation time faster? To answer this question on a more detailed level, the next analyses looked into how navigation behaviors were changed in one specific trial as a function of the number of input behaviors for each input gesture. As described in Section 5, one tapping gesture means one item movement and one wheeling gesture means four item unit movement. One flicking gesture can include several item unit movement depending on the strength. For each tap and flick, release action is needed, whereas wheeling can be continuously done without any release. These analyses selected a trial with a relatively distant target which showed clearer spindex effects. Because a target was randomly chosen for both Bin and item in every trial, it is impossible to compare the exact same target number for the TTS and TTS + spindex conditions. Therefore, we selected representative trials in which targets are in a similar distance for both auditory cue conditions in order to illustrate the spindex effects.

Because tapping and wheeling involved similar linear relations between target distance and navigation time, they also showed similar behavior patterns in one trial analysis. In the TTS condition, a participant appeared to frequently pause to check his or her location in an early stage (Figure 9 and 10). In contrast, in the TTS + spindex condition, the participant paused fewer times before reaching the target zone. This behavioral difference is in accordance with the two-stage menu navigation strategy. This is discussed more in Section 7.2. In both cumulative figures of the tapping and the wheeling trial (Figure 11), spindex benefits increased as the number of input actions increased.

In the Flicking condition, participants' common strategy was to flick strongly in an early stage (nontarget zone), and then flick more gently several times near the target zone (Figure 12). For flick number $1 \sim 4$ in both conditions, time between each flick was relatively longer because the user might have flicked more strongly and skipping menu items took longer. In the TTS condition, there were more soft flicks than in the TTS + spindex condition. On the other hand, in the TTS condition of the Tapping and Wheeling conditions participants needed more breaks for the status check between

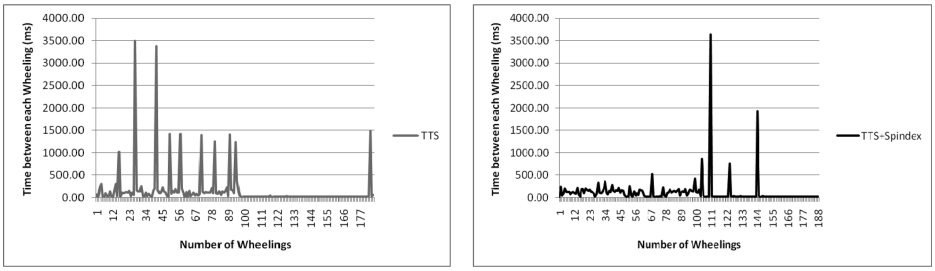

Fig. 10. Time between each wheeling as a function of the number of wheelings: Participant B in the Visuals-off TTS Condition Block 1 Trial 4 Target No. 112 (left) and in the Visuals-off TTS + spindex condition Block 1 Trial6 Target 114 (right).

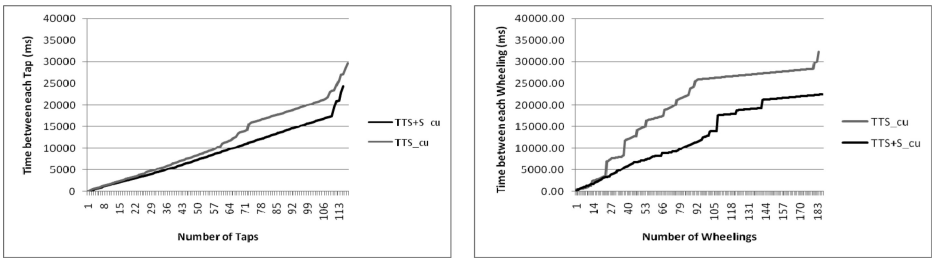

Fig. 11. Cumulative time between each tap as a function of the number of taps (Participant A in the Visuals-off condition) (left) and cumulative time between each wheeling as a function of the number of wheelings (Participant B in the Visuals-off condition) (right).

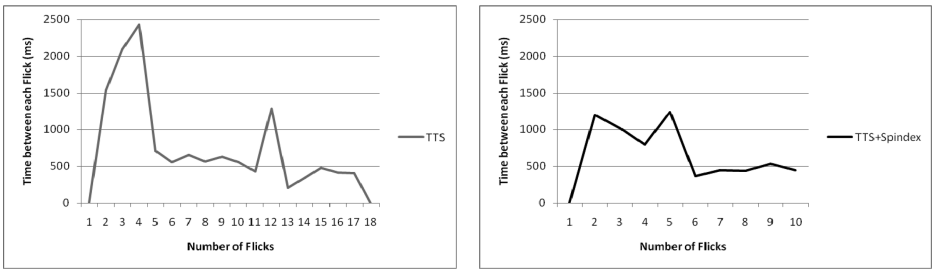

Fig. 12. Time between each flick as a function of the number of flicks: Participant C in the Visuals-off TTS condition Block Block 1 Trial 5 Target 109 (left) and in the Visuals-off TTS + spindex condition Block 1 Trial 10 Target 110 (right).

inputs; this resulted in more flicking times in the Flicking condition. This is supported by analysis with a 2 (Visual mode) x 2 (Auditory cue type) x 3 (Block) repeated measures ANOVA for the number of flicks. The results revealed a statistically significant difference in auditory cue type and visual mode for the mean number of flicks. The TTS + spindex condition (M = 68.49, SD = 36.49) led to significantly fewer flicks than the TTS condition (M = 75.76, SD = 28.46), $F(1, 38) = 4.29$, $p < .05$, $\eta_p^2 = .10$ (Figure 13). Also, the Visuals-on condition (M = 46.24, SD = 30.73) led to significantly fewer flicks than the Visuals-off condition (M = 98.01, SD = 30.75), $F(1, 38) = 28.29$, $p < .001$, $\eta_p^2 = .43$. In addition, block showed a significant practice effect for the number of flicks, $F(2, 76) = 5.31$, $p = .007$, $\eta_p^2 = .12$.

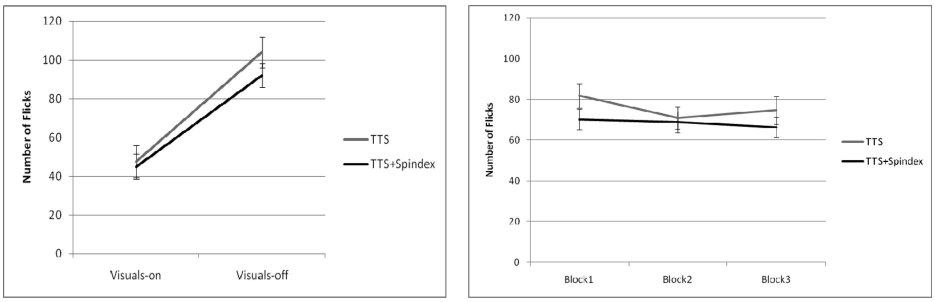

Fig. 13. Number of flicks for visual mode (left) and for block (right). The TTS + spindex condition required fewer number of flicks than the TTS condition in both visual types. Also, the TTS + spindex condition reduced the number of flicks more consistently as block increased than the TTS condition.

7. DISCUSSION

The Fairly recently, the spindex, a new type of nonspeech auditory cue was introduced and showed promise for performance and preference in one-dimensional auditory menu navigation in several studies [Jeon and Walker 2011; Jeon et al. 2009]. Correspondingly, results in the present study strongly supported the benefits of adding spindex cues to speech menus on a touch screen mobile device using various input gestures.

In the present experiment, the TTS + spindex condition showed better actual performance (navigation time), lower perceived workload (NASA TLX), and higher perceived performance (effective and functionally helpful ratings). These spindex enhancement effects were shown both in the Visuals-on and Visuals-off conditions across all three input gestures. In terms of universal design, the enhancement of the spindex even in the Visuals-on condition showed that this improvement has the potential to help not only visually impaired people but also sighted users. Also, because there was no difference in error rates between the TTS-only and the TTS + spindex conditions, there was no trade-off between speed and accuracy. Only subjective preference ratings (likable, fun, and annoying) showed no difference between the TTS condition and the TTS + spindex condition. That is, adding the spindex was not irritating for the user, even though the overall auditory output was lengthened slightly.

7.1 Navigation Interaction among Input Gestures

In addition to spindex benefits, the unique characteristics of each input gesture were also identified. For instance, tapping showed higher workload ratings than wheeling or flicking because tapping required more physical movements than the other input gestures. On the other hand, in the Visuals-off condition, flicking showed a sharp increase in both navigation time and workload scores. It might be because the Flicking condition has more visible items in the Visuals-on condition than other conditions. However, as Figure 3 shows, in the Visuals-on condition, the Wheeling condition outperformed the Tapping condition even though it has smaller visible items at a time than the Tapping condition. Therefore, the number of visible items on the menu may not be linearly correlated with the navigation performance. Rather, it can be analyzed that flicking is a more visually demanding task than the others. This notion can be supported by experimenter's anecdotal observations. Because the number of items passed was not constant in the Flicking condition, participants often had to clutch the menu or go backward when passing by the target item mistakenly, which occurred more in the Visuals-off condition.

7.2 Navigation Behavior Pattern and Two Stage Navigation Model

The navigation time data revealed exactly where and how the spindex benefits occurred for the three input gestures. As can be seen in Figure 8, the spindex changed the slope of the function relating item location and time to target. The spindex effect in navigation efficiency increased as the target distance increased. Again, this is due to the fact that even small per-item enhancements lead to important and noticeable navigation time gains in the menu search. In addition, more microlevel analysis showed how the spindex worked in one trial. In the TTS condition of the tapping and wheeling gestures, participants frequently paused in between taps and wheels to figure out their status or location, but in the TTS + spindex condition, they did not need to do that. In the Flicking condition, participants in both auditory cue conditions showed a similar behavior pattern in which they flicked strongly in an early stage, and then flicked gently in the target zone. Adding the spindex allowed them to make fewer flicks overall.

We can infer that these three behaviors correspond to the two-stage navigation model: rough navigation and fine navigation. As mentioned in Section 2, in the rough navigation stage, users exclude nontargets until they approach the alphabetical area that includes 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 information about the nontargets. It is sufficient for them to obtain only enough information to decide whether they are in the target zone or not. After users perceive that they have reached the target zone, they then need the detailed information about each menu item to compare it with the target. The spindex-enhanced auditory menu can contribute significant per-item speedups during the rough navigation, and then, the TTS phrase still supports detailed item information in the fine navigation stage. Figures 9 to 12 show that time consumption in the rough navigation of the TTS condition was much longer than that in the fine navigation of the TTS + spindex condition.

7.3 Perceived Workload, Preattentive Processes, and Task Shift

The spindex seems to leverage what users are already familiar with from tangible examples of long list 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. Previous research [Beck and Elkerton 1989] suggested that visual indexes could decrease visual search time with list menus. We would explain that the spindex is a successful translation of the index from the visual display into the auditory display realm (hence the name spindex).

Besides a functional approach to navigation efficiency, the spindex effect in terms of perceived workload can be explained by attention theories on a psychological level. For example, in visual search theory, finding a red "O" among various colors of different alphabet characters (e.g., "As, Bs, Cs, Gs, Qs") is not easy. However, finding a red "O" among many white characters is easier because the oddball color is automatically processed *pre-attentively*, without the full use of attention [Treisman and Gelade 1980; Treisman and Gormican 1988]. In the latter condition, the target "O" will *pop out* in the pre-attentive processing stage due to its color. Similarly, in auditory search, if we make distracters (i.e., nontargets) unified (e.g., replace the variable sounds of "C," namely /si/, /k/, /cha/, /t□i/,... with /si/), people can easily *filter out* nontargets with no attentional limits, although the target cannot pop out because auditory processing is serial. This filtering may occur at a surface and acoustic processing level rather than a deep and linguistic (or semantic) level. It should require merely pre-attentive and automatic processing. In hearing as well as in vision, similar concepts have been suggested as *filtering* and *pigeonholing* [Broadbent 1977]. The preattentive filtering

processes segregate detailed information into bundles or segments whether they are to be attended to or rejected as a whole. The filtering is the selection of the stimulus for attention because it possesses a certain feature that is absent from distracters (e.g., capital letters among others in lower case; the spindex of the target zone among others in different spindex cues in our experiment). However, in pigeonholing, the target and distracters do not differ by any single feature (e.g., every different sound of all TTS cues). Due to the larger amount of information that needs to be compared, the pigeonholing requires more processing than filtering does.

Reflecting this exact notion, one participant commented that "the softer voice [spindex cue] was better for finding song titles later on in the alphabet because I didn't have to put all my attention on looking at the screen to see if I was at that letter yet. Softer voices [spindex cues] are better on the ears." This reflects the fact that participants reported lower workload in the spindex condition than in the TTS-only condition.

The benefit of the spindex in the Visuals-on condition as well as in the Visuals-off condition can also be explained by a similar but slightly different perspective. Auditory display has been well known for its advantages of displaying time-varying data, such as in a monitoring task [Kramer 1994] or detecting the correct auditory signal in streaming [Walker and Kramer 2004]. Therefore, if adding spindex cues to the TTS can change an auditory search task (which needs active attentional processing) into a monitoring task (which does not require that effort), the benefit of the spindex is not surprising. Also, we can infer that participants, even in the Visuals-on condition, might depend more on the auditory signal in such a monitoring task than the visual signal, which passed by rapidly and may appear blurred.

7.4 Trade-Offs between Gain and Loss of the Spindex and Improvements

Even with these positive results, the spindex can still be improved considering its trade-offs between performance gain and loss. Although users can benefit from adding spindex cues in the rough navigation stage, it takes more time for them to hear both the spindex and TTS parts for fine tuning in the fine navigation stage. For example, one participant reported, "Perhaps somehow implement another condition in which the first letter is only read when scrolling quickly. That would be more useful in my opinion because reading the first letter makes it easier to reach each letter [section of the list], but it adds time to pinpointing the exact song within the letter [section]." Indeed, in our experiment, navigation time increased in Bins 6 and 7 where there are relatively many items that start with "i." Then, if there are more items in a menu, would the spindex benefits decrease? How can we analyze the trade-offs more systematically and maximize spindex effects? To answer this question, we roughly formularized the elements that can contribute to the performance gain and loss of the spindex in Figure 14.

Performance (time) loss of the spindex in the fine navigation can approximately be figured out by

= [length of spindex + interval (250 ms) + length of some portion of TTS] x (jT − 1) in AT +- errors

Once getting to the target zone, users need to spend time to listen to the spindex, interval, and some part of the TTS phrases of each item to identify if the item is the target until they reach the target item. Here, the spindex and interval are fixed and necessary parts, but the portion of the TTS may vary depending on participants and situations. In some cases, participants can notice if it is the target by hearing the first word of the TTS or less than that, but in other cases, they cannot. There might

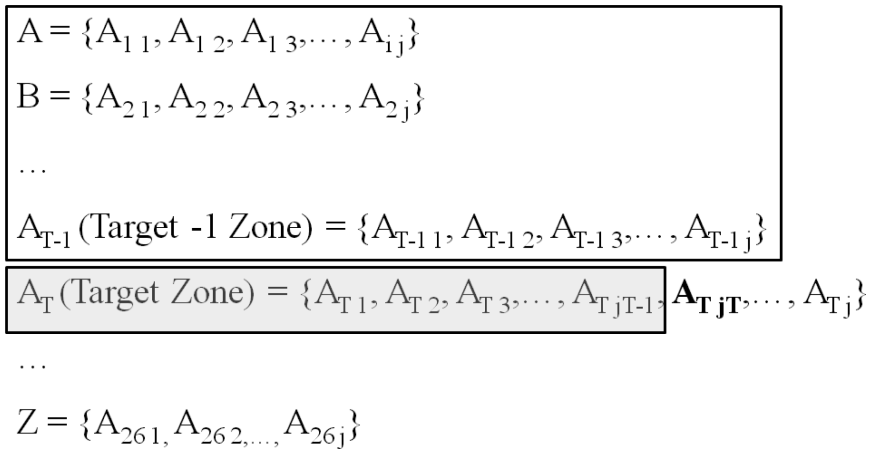

Fig. 14. A conceptual formula of the spindex benefit trade-offs. A white box means performance gain of the spindex and a grey box means performance loss of the spindex. A_{ij} = an item in i^{th} alphabet zone (row) and j^{th} location (column), (1 <= i <= 26 and 0 <= j < unlimited). The target item is A_{TjT}.

be additional errors that this formularization does not identify (e.g., going back when users passed by the target item mistakenly).

On the other hand, performance (time) gain of the spindex in the rough navigation can approximately be figured out by

$= \Sigma_{1\,1}^{T-1j}$ (length of some protion of TTS − length of spindex) + Sum of length of each status check in TTS only condition +-*errors*.

Performance gain of the spindex has to be calculated compared to the use of TTS-only. In general, performance gain of the spindex is obtained from the first item of the menu to the last item (A_{T-1j}) in A_{T-1} zone. The amount of spindex benefit of each item equals the differences between the length of some portion of the TTS to perform the status checks in the TTS-only condition and the length of the spindex in the TTS + spindex condition. Frequent status check in the TTS-Only condition until getting to the target zone (A_T) adds time loss in the TTS-only condition. It may also include some unknown errors. In brief, adding menu items can influence both performance gain and loss of the spindex. The more items each alphabet zone has, the more loss in AT (Target Zone) occurs, but the more gain in a range from A to AT-1 (Target -1 Zone) occurs. Status check is more likely to happen as the total number of menu item increases in the TTS condition. Several variables can be changed depending on situations and even interaction gestures. Therefore, a simple profit and loss statement may not be obtained with this conceptual formula. However, empirical results [Jeon and Walker 2011] still support that spindex benefits may increase as the total number of menu items increases. In the first experiment in that study, spindex benefits were shown more clearly in the longer menu (with 150 menu items) than in the shorter menu (with 50 menu items) even though in both cases, the spindex-enhanced menu was significantly more effective than the TTS-only menu. Further, in the present study where the menu has a considerable number of items in only one alphabet zone, the spindex loss occurred only when the target item was in the latter part of that zone. The total results still showed spindex benefits because that zone also helped increase spindex benefits when the target item was located after that zone. Note that if the

target item is in a zone 'A' or the first zone of the menu, spindex benefits may fall into zero and there would be only performance loss per item. However, even from the second alphabet zone, spindex benefits would take place. In conclusion, because the spindex benefits result in saving small amount of time per item just before getting to the target zone, if the menu has more items, the target zone is the latter alphabet, and the target item is in the earlier location of the target zone, the spindex benefits will increase.

Considering these analyses, there are two plausible solutions to compensate for performance loss of the spindex. First, the interval between the spindex and the item could be decreased. From our experience in several experiments, the presence of the spindex-TTS interval seems necessary to distinguish the cue and the item during fast search, but it could be shortened from the 250 ms used in this experiment. Second, the spindex can be applied to a menu system in a more adaptive way. If the user input (whether it is tapping, wheeling, or flicking) is slow or gentle, the speech menu system could speak out only the TTS item. On the other hand, if the user input is faster or stronger, the device would generate spindex cues only. With these adaptations, there would be less or no time sacrifice due to the spindex, even in the fine navigation stage.

7.5 Advantages in Practical Applications of the Spindex

From the perspective of practical applications for touch screen devices, the spindex has several advantages. First, the spindex does not require any major change to the programming architecture, nor does it take up large storage space on a device. A little tweaking of the software, such as parsing the newer items and adding prerecorded files, could fulfill the requirements for a fast implementation of the spindex. Second, the fact that participants gave higher scores to the spindex menu on the subjective ratings indicated that they did feel that the spindex provided a better user experience in their navigation task. Especially with the new input gestures on touch screen mobile devices (e.g., wheeling and flicking), user experience should be fun, engaging, and creative [Blythe et al. 2003; Russell and Bryan 2009] because to a certain user population, such as the "thumb generation," the mobile phone serves as an entertainment object [Jeon et al. 2008]. Finally, it is also encouraging that the benefit of spindex cues comes after little or no practice, which indicates a low threshold for helping new users. These advantages can significantly increase the possibility of applying spindex cues to real devices.

For constructing a comprehensive auditory menu navigation theory, this research can be extended in several ways. First, in addition to this 'search' behavior in a 'time' domain, researchers can also examine 'browsing' or 'exploring' behaviors in a 'spatial' domain without a specific target. Second, we may be able to figure out when and where nonspeech sound cues (e.g., earcons, spearcons, auditory scrollbars, and spindexes) can be optimally used together or alone. Finally, future research should be carried out in real mobile contexts such as walking, jogging, or driving to gain more practical relevance [Fisk and Kirlik 1996]. We hope that researchers and designers can use the usability metrics and practical results of the present study for implementation of auditory user interfaces and further design evaluation.

ACKNOWLEDGMENTS

We would like to thank Dr. Gregory Corso and Dr. Frank Durso for their insightful comments and feedback on this study. We also acknowledge the valuable contributions of Yarden Moskovitch during data collection for this project.

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).*

ANDERSEN, T. H. AND ZHAI, S. 2010. "Writing with music": Exploring the use of auditory feedback in gesture interfaces, *ACM Trans. App. Percept. 7,* 3, 17:1–24.

ARONS, B. 1997. SpeechSkimmer: A system for interactively skimming recorded speech. *ACM Trans. Comput.-Hum. Interact. 4*, 1, 3–38.

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.

BLYTHE, M., MONK, A., OYERBEEKE, C., AND WRIGHT, P. Eds. 2003. *Funology: From Usability to User Enjoyment*. Kluwer Academic Publishers.

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

BREWSTER, S. A. 2002. Overcoming the lack of screen space on mobile computers. *Pers. Ubiq. Comput. 6*, 3, 188–205.

BREWSTER, S. A. 2008. Chapter13: 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 BCS Conference on Human-Computer Interaction (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.

BROADBENT, D. E. 1977. The hidden preattentive processes. *Amer. Psych. 32*, 2, 109–118.

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)*.

EDWARDS, A. D. N. 1989. Soundtrack: An auditory interface for blind users. *Hum.-Comput. Interact. 4*, 45–66.

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.

FISK, A. D. AND KIRLIK, A. 1996. Practical relevance and age-related research: Can theory advance without practice? In *Aging and Skilled Performance: Advances in Theory and Application*, W. A. Rogers, A. D. Fisk, and N. Walker Eds., Erlbaum, NJ, 1–15.

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.

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.

HART, S. G. 2006. NASA-Task Load Index (NASA-TLX); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting (HFES'06)*.

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).*

JEON, M. AND WALKER, B. N. 2011. Spindex (speech index) improves auditory menu acceptance and navigation performance. *ACM Trans. Access. Comput. 3*, 3, 10:1–26.

JEON, M., NA, D., AHN, J., AND HONG, J. 2008. User segmentation & UI optimization through mobile phone log analysis. In *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI'08)*. 495–496.

JEON, M., DAVISON, B. K., NEES, M. A., WILSON, J., AND WALKER, B. N. 2009. Enhanced auditory menu cues improve dual task performance and are preferred with in-vehicle technologies. In *Proceedings of the the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI'09)*. 91–98.

KANE, S. K., BIGHAM, J, P., AND WOBBROCK, J. O. 2008. Slide rule: Making mobile touch screens accessible to blind people using multi-touch interaction techniques. In *Proceedings of the Annual ACM Conference on Assistive Technologies (ASSETS'08)*. 73–80.

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.

LEE, J. AND SPENCE, C. 2008a. Feeling what you hear: Task-irrelevant sounds modulate tactile perception delivered via a touch screen. *J. Multimodal User Interfaces 2*, 3–4, 1783–7677.

LEE, J. AND SPENCE, C. 2008b. Spatiotemporal visuotactile interaction. In *Proceedings of the Haptics: Perception, Devices and Scenarios, 6th International Conference (EuroHaptics'08)*. Lecture Notes in Computer Science, vol. 5024, 826–831.

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)*.

LI, K. A., BAUDISCH, P., AND HINCKLEY, K. 2008. BlindSight: Eyes-free access to mobile phones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'08)*. 1389–1398.

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)*.

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 (ASSETS98)*.

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

MYNATT, E. AND WEBER, G. 1994. Nonvisual presentation of graphical user interfaces: Contrasting two approaches. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'94)*. 166–172.

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

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

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

NORMAN, D. A. AND NIELSEN, J. 2010. Gestural interfaces: A step backward in usability. *Interactions 17*, 5, 46–49.

NORMAN, K. 1991. *The Psychology of Menu Selection: Designing Cognitive Control at the Human/Computer Interface*. Ablex Publishing Corp., Norwood, New Jersey.

OH, J. W., PARK, J. H., JO, J. H., LEE, C., AND YUN, M. H. 2007. Development of a kansei analysis system on the physical user interface. In *Proceedings of the Korean Conference on Human Computer Interaction.*

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 (HFES'08)*. 1262–1266.

PIRHONEN, A., BREWSTER, S. A., AND HOLGUIN, C. 2002. Gestural and audio metaphors as a means of control for mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'02)*. 291–298.

PITTS, M. J., WILLIAMS, M. A., WELLINGS, T., AND ATTRIDGE, A. 2009. Assessing subjective response to haptic feedback in automotive touchscreens. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI'09)*. 11–18.

RAMAN, T. V. 1997. *Auditory User Interfaces: Toward the Speaking Computer*. Kluwer Academic Publishers.

RUSSELL, D. C. AND BRYAN, R. 2009. To touch or not to touch: A brief guide for designing or selecting touch screen computers and touch software for consumer use. In *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting (HFES'09)*. 980–984.

SANDERS, M. S. AND MCCORMICK, E. J. 1993. Chapter 11: 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.

SAWHNEY, N. AND SCHMANDT, C. 2000. Nomadic radio: Speech and audio interaction for contextual messaging in nomadic environments. *ACM Trans. Comput.-Hum. Interact. 7,* 3, 353–383.

TREISMAN, A. M. AND GELADE, G. 1980. A feature-integration theory of attention. *Cogn. Psych. 12*, 97–136.

TREISMAN, A. M. AND GORMICAN, S. 1988. Feature analysis in early vision: Evidence from search asymmetries, *Psych. Rev. 95*, 1, 15–48.

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. 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, 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. AND KRAMER, G. 2006. Auditory displays, alarms, and auditory interfaces. In *International Encyclopedia of Ergonomics and Human Factors* 2nd Ed., W. Karwowski Ed., CRC Press, New York, 1021–1025.

WALKER, B. N., AND NEES, M. A. 2012. Theory of sonification. In *Handbook of Sonification*. T. Hermann, A. Hunt, and J. Neuhoff Eds., Academic Press: New York, 9–39.

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.

WALKER, B. N., LINDSAY, J., NANCE, A., NAKANO, Y., PALLADINO, D. K., DINGLER, T., AND JEON, M. in press. Spearcons (Speech-based earcons) improve navigation performance in advanced auditory menus. *Human Factors.*

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.

ZHAO, S., DRAGICEVIC, P., CHIGNELL, M., BALAKRISHNAN, R., AND BAUDISCH, P. 2007. earPod: Eyes-free menu selection using touch input and reactive audio feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'07)*. 1395–1404.

Received May 2011; revised November 2011; accepted March 2012