"SPINDEX" (SPEECH INDEX) ENHANCES MENU NAVIGATION USER EXPERIENCE OF TOUCH SCREEN DEVICES IN VARIOUS INPUT GESTURES: TAPPING, WHEELING, AND FLICKING

A Thesis
Presented to
The Academic Faculty

by

Myounghoon Jeon

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Psychology

Georgia Institute of Technology
December, 2010

"SPINDEX" (SPEECH INDEX) ENHANCES MENU NAVIGATION USER EXPERIENCE ON TOUCH SCREEN DEVICES IN VARIOUS INPUT GESTURES: TAPPING, WHEELING, AND FLICKING

Approved by:

Dr. Bruce N. Walker, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Gregory M. Corso
School of Psychology
Georgia Institute of Technology

Dr. Frank Durso
School of Psychology
Georgia Institute of Technology

Date Approved: September 23rd, 2010

I dedicate this thesis to my family, for their never-ending support.

ACKNOWLEDGEMENTS

I wish to thank Dr. Bruce Walker for his guidance. I would also like to thank the Sonlabbers for their support and discussion.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

In a large number of electronic devices, users interact with the system by navigating through various menus. Auditory menus can complement or even replace visual menus, so research on auditory menus has recently increased with mobile devices as well as desktop computers. Despite the potential importance of auditory displays on touch screen devices, little research has been attempted to enhance the effectiveness of auditory menus for those devices. In the present study, I investigated how advanced auditory cues enhance auditory menu navigation on a touch screen smartphone, especially for new input gestures such as tapping, wheeling, and flicking methods for navigating a one-dimensional menu. Moreover, I examined if advanced auditory cues improve user experience, not only for visuals-off situations, but also for visuals-on contexts. To this end, I used a novel auditory menu enhancement called a "*spindex*" (i.e., speech index), in which brief audio cues inform the users of where they are in a long menu. In this study, each item in a menu was preceded by a sound based on the item's initial letter. One hundred and twenty two undergraduates navigated through an alphabetized list of 150 song titles. The study was a split-plot design with manipulated auditory cue type (text-to-speech (TTS) alone vs. TTS plus spindex), visual mode (on vs. off), and input gesture style (tapping, wheeling, and flicking). Target search time and subjective workload for the TTS + spindex were lower than those of the TTS alone in all input gesture types regardless of visual type. Also, on subjective ratings scales, participants rated the TTS + spindex condition higher than the plain TTS on being 'effective' and 'functionally helpful'. The interaction between input methods and output

modes (i.e., auditory cue types) and its effects on navigation behaviors was also analyzed based on the two-stage navigation strategy model used in auditory menus. Results were discussed in analogy with visual search theory and in terms of practical applications of spindex cues.

CHAPTER 1

INTRODUCTION

Research on the use of non-speech sounds for information display in user interfaces has rapidly grown since the early 1990s (Kramer, 1994). Their benefits have been demonstrated in a wide range of different applications, from systems for blind people (Edwards, 1989; Jeon & Walker, 2010; Raman, 1997) to mobile devices (Brewster & Cryer, 1999; Brewster, Leplatre, & Crease, 1998; Jeon & Walker, 2009; Klante, 2004; Leplatre & Brewster, 2000; Palladino & Walker, 2007, 2008a, 2008b; Vargas & Anderson, 2003; Walker, Nance, & Lindsay, 2006), and ubiquitous/wearable computers (Brewster, Lumsden, Bell, Hall, & Tasker, 2003; Sawhney & Schmandt, 2000; Wilson, Walker, Lindsay, Cambias, & Dellaert, 2007).

Brewster (2008) pointed out two important areas into which non-speech 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, utilizing more of the user's senses. In a similar vein, adding sound to interfaces does not only improve performance, but also enhances subjective satisfaction and reduces perceived workload (see more detailed reviews in Chapter 3).

The second arena for sound incorporation is in mobile and wearable computing devices. The small or non-existent 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, IT devices such as mobile phones, PDAs, and MP3 players have started to adopt touch screen technology in order to enhance the user experience (Lee & Spence, 2008b). According to a study (Oh, Park, Jo, Lee, & Yun, 2007), auditory feedback is the most effective modality in physical user interface satisfaction, followed by tactile and motion feedback. Moreover, use of auditory displays is increasingly important, especially considering the fact that touch devices generally lack tactile feedback and have input areas that overlap with display area. Nevertheless, the prevalence of auditory displays has not reflected the great prevalence of touch screen interfaces (Fritz, 2000). One study demonstrated that task-irrelevant sound modulates tactile perception delivered via a touch screen (Lee & Spence, 2008a). In another study on assessing subjective response to automotive touch screens, adding only haptic feedback to visual display did not produce a reliable difference from visual-only display (Pitts, Williams, Wellings, & Attridge, 2009). In contrast, adding audio feedback showed significant differences from visual-only display. Moreover, in the same study, haptic effects were perceived as stronger in the presence of audible feedback.

Consequently, auditory feedback has been presented in some touch devices as a default display. For example, the HP Touch Smart® relays a noise from the rear of the computer whenever users touch the screen (Simms, 2008). Also, the Windows 7 Beta® shows a ripple on-screen display every time users tap the screen. iPod also generates a click sound when users move their finger around the touchpad's circumference for separating the items or units.

In summary, use of sounds with another modality, such as visual and tactile/haptic, seems to enrich the user experience in terms of multi-modal interaction. In addition,

sounds can play a more prominent role when another modality is not possible with mobile devices. Based on these premises, this study investigated the use of non-speech sounds for auditory menu on a touch screen device for improving navigation efficiency and subjective satisfaction.

In this thesis, first, auditory enhancements in mobile devices are reviewed. These include purely auditory interfaces and the addition of non-speech sound cues to an existing system (Chapter 2). Then, user experience metrics for auditory interfaces are discussed in terms of objective and subjective evaluation (Chapter 3). This is followed by motivations for the current study and hypotheses for the experiment (Chapter 4). Finally, the experiment is presented, implementing a novel auditory menu enhancement called a *spindex* (Jeon & Walker, 2009) for various input gestures such as tapping, wheeling, and flicking on a touch screen device (Chapter 5).

CHAPTER 2

AUDITORY ENHANCEMENTS IN MOBILE DEVICES

Two main pieces of related work are described in this chapter. On the one hand, research on creating purely auditory interfaces has been conducted, which attempts to provide a novel auditory-specific system. On the other, various non-speech sounds have been added to existing interfaces to improve usability. The current study focuses mostly on the latter.

Purely Auditory Interfaces

Purely auditory interfaces include SpeechSkimmer (Arons, 1997), Nomadic Radio (Sawhney & Schmandt, 2000), and earPod (Zhao, Dragicevic, Chignell, Balakrishnan, & Baudisch, 2007). These interfaces each demonstrate how auditory menu navigation can be improved using speech and non-speech sounds.

SpeechSkimmer

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, user interaction renders continuous real-time control of the speed and detail level of the audio presentation. SpeechSkimmer reduces the time needed to listen in four different ways by incorporating features such as time-compressed speech, pause shortening, automatic emphasis detection, and non-speech audio feedback.

Nomadic Radio

Nomadic Radio (Sawhney & 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.

In Nomadic Radio, a preview for text messages extracts the first 100 characters of the message. Sawhney and Schmandt (2000) suggest this heuristic generally provides sufficient context for the listener to anticipate the overall theme and urgency of the message. Further, a preview for an audio source such as a voice message or news broadcast presents one-fifth of the message at a gradually increasing playback rate of up to 1.3 times faster than normal.

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

earPod

One of the most recent menu implementations 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. In the evaluation study, comparing earPod with an iPod-like visual menu technique on reasonably-sized static menus indicates that they are similar in accuracy. earPod even outperforms the visual menu in terms of efficiency (reaction time) within 30 minutes of practice. 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 non-speech audio like short mechanical click sounds when crossing the boundary in touchpad and a camera-shutter sound to confirm item selection. Finally, it adopts binaural spatial sound cues to reinforce users' cognitive mapping between menu items and spatial locations on the touchpad.

earPod is indeed a good example of eyes-free menu navigation, but it is arguable 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. Instead, all of the sections of the device are effective input areas. In contrast, in the case of an iPod-like visual menu, users have to combine visual search on the small screen with motor control. As described in the paper, after moderate learning, users can directly tap the target area depending on their motor memory without sliding their thumb on the circular touchpad and listening to each item's name. Therefore, earPod may be ideal for navigation in the restricted hierarchical menu, but not be proper for navigation in a long list menu like an address book or an MP3 song list, which does not allow motor/spatial memory or direct access.

These attempts to invent 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, these studies are different from the present study in two ways: they require a new type of device in order to fully implement those functions, and they have cost and generalization issues. Consequently, these novel interfaces ask users to learn new interactions for use. An alternative to bypass these issues would enable users to benefit from simply adding non-speech sounds to current devices to which they are already accustomed.

Adding Non-speech Sounds to the Existing System

There have been three main approaches to enhancing the basic TTS used in most auditory interfaces. These all tend to include adding sound cues before or concurrent with the spoken menu items. The main types of enhancement cues are categorized as auditory icons (Gaver, 1986), earcons (Blattner, Sumikawa, & Greenberg, 1989), and spearcons (Walker, et al., 2006). In addition to these, a new concept—the spindex— is introduced.

Auditory Icons

Auditory icons are audio representations of objects, functions, and events (Gaver, 1986). They are caricatures of naturally occurring sounds such as bumps, scrapes, or even files hitting mailboxes. As caricatures, auditory icons capture an object's essential features, by presenting a representative sound of the object. Auditory icons can represent various objects in devices more clearly than other auditory cues because the relation between a source of sound and a source of data is more natural than others. For example, a typing sound can represent a typewriter. Thus, auditory icons typically require little training and are easily learned. Adopting these advantages, Gaver (1989) created an auditory icon-enhanced desktop. Other researchers have attempted to convert GUIs to non-visual interfaces using auditory icons (Mynatt, 1997; Mynatt & 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 devices 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 & 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 in particular. However, one could apply auditory icons for the auditory menu navigation study as well. Consider an address book list on the mobile phone. One can record a friend's voice and register it as feedback of the item on the list. In fact, in an address book of recent touch screen phones, users can save the picture taken with the contact as a visual icon or short cut for the item. Use of auditory icons as address book contacts might enhance users' subjective satisfaction, but might not facilitate navigation efficiency.

Earcons

Earcons are non-speech audio representations, composed of musical motives, which are short, rhythmic sequences of pitches with variable intensity, timbre and register, to provide information to the user about some objects, operations or interactions (Blattner, et al., 1989). Since earcons use an arbitrary mapping between sound and object, they can be analogous to a language or a symbolic sign. This arbitrary mapping between earcon and represented item means that earcons can be applied to any type of menu; that is, earcons can represent any concept in general. However, this flexibility can also be a weakness because the arbitrary mapping of earcons to concepts requires user training. Earcons can also represent 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 & Moller, 1999; Morley, Petrie, & McNally, 1998). The results showed improved usability and browsing experience. However, when a new item has to be inserted in a fixed menu structure, it might be difficult to create a new branch sound. The structural framework of earcons can be congruent with logical hierarchical menus, but it is difficult to apply to one-dimensional long list menus. If the menu includes hundreds of items, it is hard to memorize those arbitrary mappings. For the most recent overview of auditory icons and earcons, see Absar and Guastavino (2008).

Spearcons

Spearcons are brief sounds that are produced by speeding up spoken phrases, even to the point where the resulting sound is no longer comprehensible as 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." Or, if a new name is added to a contact list, the spearcon can be quickly and spontaneously created as needed. Also, spearcons are easy to learn whether they are comprehensible as a particular word or not, because they derive from the original speech (Palladino & Walker, 2007).

Spindex: An Auditory Index Based on Speech Sounds

A spindex (i.e., Speech Index) is created by associating an auditory cue with each menu item, in which the cue is based on the pronunciation of the first letter of each menu item (Jeon & Walker, 2009). For instance, the spindex cue for "Apple" would sound /ei/ or even /a/ based on the spoken sound of "A", 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.

When people control devices, there are two types of human motions in such tasks. In a *gross*-adjustment movement, the operator brings the controlled element to the approximate desired position. This gross movement is followed by a *fine*-adjustment, in which the operator makes adjustments to bring the controlled element precisely to the desired location (Sanders & McCormick, 1993). When it comes to a search task such as the navigation through an address book (visually or auditorily), similarly, the process can be divided into two stages. One is *rough navigation* and the other is *fine navigation* (Klante, 2004). In the *rough navigation* stage, users pass or jump the non-target alphabet groups by glancing at their initials. For example, users quickly jump to the "T" section. Then, once users reach a target zone, they begin the *fine navigation*, check where they are and cautiously tune their search. In auditory menus, people cannot jump around as easily, given the temporal characteristics of spoken menu items. However, they still want to pass over the non-target alphabetical groups as fast as possible. If a sound cue is sufficiently informative, users do not need to listen to the whole TTS phrase (Palladino & Walker, 2007). The initials of the alphabet of the list can yield enough information to users when

they sort out the non-target items. A previous study showed that the benefits of a cue structure of a spindex could be realized more clearly in a long menu with a number of items (*Number of item* = 150) than a short menu (50) even though the benefits of the spindex cues were reliably demonstrated in both menus (Jeon & Walker, 2009). Given that spindexes are likely more useful in long menus, it is fortuitous that spindex cues can be generated quickly by TTS engines, and do not require the storage of numerous additional audio files for the interface. This is an important issue for mobile devices, which, despite increasing storage for content file, are not designed to support considerable extra files for their menu interface. Finally, because spindex cues are the part of the original word and are natural—based on speech sounds—they do not require much training.

CHAPTER 3

USER EXPERIENCE METRICS FOR AUDITORY INTERFACES

The importance of subjective acceptance and preference level of the user interface has been increasing in user experience design circles. For example, Don Norman (2004) has stressed the importance of visceral design. Furthermore, he proposed that an attractive and natural design can sometimes 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 display than in visual display (Brewster, 2008; Davison & Walker, 2008; Kramer, 1994; Nees & Walker, 2009), to date, research has mainly focused on performance issues. A study once suggested that the nature of sound aesthetics is independent of performance outcomes (Edworthy, 1998). 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 & Walker, 2009). Therefore, it is evident that developing universal evaluation metrics including objective and subjective aspects is crucial to the success of auditory interfaces. From this standpoint, I attempted to collect various dependent measures for auditory display from literature.

Objective Evaluation Metrics

Performance improvement by the addition of auditory cues in menu navigation tasks has been studied by several metrics such as the measurement of reaction time, the number of key presses, accuracy, and error rate.

In earlier work, structured earcons have shown a superior learning rate (i.e., recognition proportion of mapped visual objects) compared to non-organized sound (Brewster, Wright, & Edwards, 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 the experiment of sonified mobile phones, earcons improved the performance of navigational tasks in terms of the number of errors made and the number of keypresses taken to complete the given tasks (Leplatre & Brewster, 2000). Also, in a hierarchical menu experiment, participants with earcons could identify their location with over 80% accuracy (Brewster, Raty, & Kortekangas, 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 & College, 2003). Recent research on the addition of auditory scroll bars has demonstrated the potential benefits of applying earcons proportionally to each group of list items. The results showed reduced error rates in target search (Yalla & Walker, 2008).

Spearcons and spindex 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 & Walker, 2008a, 2008b). According to another study (Palladino & 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 than with TTS-alone in visuals-on and visuals-off conditions (Jeon & Walker, 2010). Additional experiments with visually impaired users showed similar results: The spindex + TTS condition enhanced navigation time compared to the TTS-only condition.

Subjective Evaluation Metrics

From the literature review, subjective evaluation factors can be categorized as perceived performance, subjective preference, and perceived workload.

Using non-speech 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). The researcher used three queries including, “would like to have these sounds in own mobile phone?”, “how distracting and irritating sounds are?” However, the first question might be contaminated 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, Leplatre, Marila, & Laine, 2001). Questions involved first impression, annoyance, aesthetical/musical judgement, opinion of the lengths of sounds, suitability to corresponding functions, effect of usage, and usefulness. What is more important in the preference scale relevant to auditory display is

that researchers have to measure annoyance as well as preference. Since users cannot avert their ears from sound, annoying sound should not be used even if some users prefer it.

Adding non-speech sounds not only improves 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 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 non-visual conditions (Walker & Kogan, 2009). Likewise, spindex cues were favored over TTS-alone with undergraduate students and visually impaired users (Jeon & Walker, 2010). In dual task contexts such as menu navigation while playing a driving-like game, all of the sound conditions reduced subjective workload score for overall tasks compared with the no sound condition. Even the spindex + TTS and the spindex + spearcon + TTS condition showed marginally lower perceived workload than TTS-only condition (Jeon, Davison, Nees, Wilson, & Walker, 2009). In conclusion, combining as many factors as possible, I include objective and subjective metrics in this thesis as follows:

Table 1. Usability evaluation metrics used in this thesis.

	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) Effective and helpful	Likert Scale (0~10) Likable, fun, and annoying	Electronic NASA-TLX

CHAPTER 4

MOTIVATIONS FOR THE CURRENT STUDY AND HYPOTHESES

In this study, I choose the second deployment strategy discussed at the outset, namely the addition of non-speech sounds to the existing system with small software tweaks. This strategy can be more universal and cost-effective than making a new auditory device per se. To date, despite various attempts at adding non-speech sounds to touch screen interfaces, there has been no research on the use of non-speech sounds for facilitating the new interaction styles becoming more readily available on touch devices, such as sliding a finger on the full touch screen ("flicking") or circling a finger on the iPod-like wheel ("wheeling").

To test these possibilities, one of the suitable spindex variants called *attenuated spindex* (Jeon & Walker, 2010) was selected as an advanced auditory cue type in this study. Musical sounds such as earcons or auditory scroll bars could be alternatives, but they might result in several issues for real applications in the one dimensional menu system. First of all, there is a mapping issue. Because earcons use arbitrary mappings between sounds and items, engineers 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 engineers' job descriptions and skillsets. They might need a sound designer or a musician for good sound quality to be implemented.

Spearcons might be a strong candidate to be adopted for this thesis. 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 lists (Jeon, et al., 2009), the spindex-enhanced TTS menu outperformed the spearcon-enhanced condition. Moreover, for input gestures such as flicking and wheeling in the current study, spearcons are still too long to implement in practical applications. Therefore, in the following research, I focus on spindex-enhanced TTS menu vs. TTS-only menu. Again, spindex is one of the shortest non-speech sounds and can be made on the fly, adding pre-recorded spindex files to the new menu items.

Research Questions

In this thesis, I attempt to attain a deeper understanding of auditory menu navigation on touch screen devices using spindex cues. More specifically, I am interested in the following research questions:

1) Can the benefit of use of spindex cues in auditory menus be extended or generalized to touch screen devices, especially for new input gestures such as tapping the screen (tapping), scrolling the wheel (wheeling), and flicking the list (flicking)?
2) Can the advanced auditory cues enhance user experience not only for visuals-off situations, but also for visuals-on contexts?
3) How does the interaction between input method and output mode (i.e., auditory cue type) affect the nature of navigation behaviors?

In addition to these research questions, I pose the question of how to build usability metrics of auditory display with subjective as well as objective evaluation.

Furthermore, I hope to contribute to establishing an auditory menu navigation theory from a long term perspective.

Hypotheses

Based on previous spindex research (Jeon & Walker, 2010), spindex is anticipated to be better than TTS only, objectively and subjectively. Target search time, number of errors, and required learning for the TTS + spindex condition will be lower than those of TTS alone in all input gesture types, at least in the visuals-off condition. Spindex cues will be more favored than plain TTS on perceived performance, subjective preference, and perceived workload evaluation in the visuals-on and visuals-off conditions.

To test these hypotheses empirically, I conducted an experiment (Chapter 5). In this experiment, undergraduate participants navigated auditory menus with TTS + spindex and TTS-only to examine whether adding spindex cues would improve navigation performance and subjective experience. Six groups of participants navigated an auditorily rendered song list menu using different input gestures – tapping, wheeling, and flicking – with the visuals-on and the visuals-off conditions. Details on methods (Chapter 5) and results (Chapter 6) follow.

CHAPTER 5

METHOD

Participants

One hundred and twenty two undergraduate students participated in this study for partial credit in psychology courses. They 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 handedness; mean number of years of touch screen device experience = 1.6). See Appendix A for questionnaire details.

Apparatus

Stimuli were presented using a Google Nexus One HTC1, a full touch screen smartphone. The mobile phone included a 3.75" resistive touch screen panel. The internal sound card was used for sound rendering. Participants listened to auditory stimuli using Sehnheiser HD 202 headphones plugged into the phone's audio jack, and adjusted for fit and comfort.

Stimuli

MP3 Song List Menu

An MP3 song list menu was created with 150 song titles gathered from the Billboard Hot 100 & Pop 100 (2009, 2008) (http://www.billboard.com/bbcom/index.jsp) and iTunes Top 100 (http://www.apple.com/itunes/top-100/songs/) (see Appendix B for

the entire list). Each visual menu (see Figure 1) was implemented in JAVA using the Android SDK programming tool for use as a plug-in for the Nexus One HTC1 smartphone.

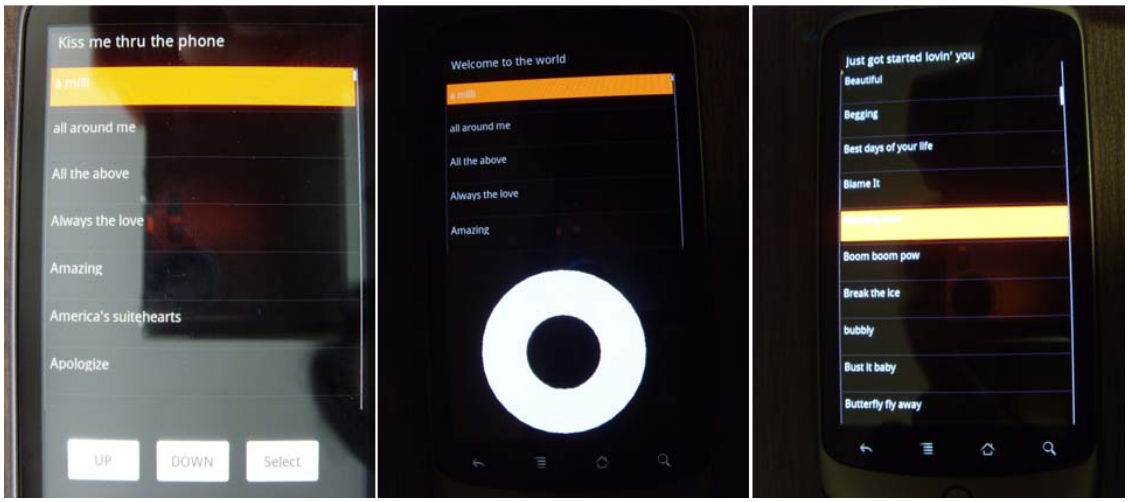


Figure 1. Visual menus for each input gesture style. From the left, the tapping, the wheeling, and the flicking condition. The target song title was visually displayed on the top of the screen in all conditions.

The menu items were presented in alphabetical order. In each type of input gesture, participants were able to scroll downward and upward in the menu by tapping on "up" or "down" button areas on the bottom of the screen (tapping condition), wheeling on a marked circular area at the bottom of the screen (wheeling condition), or flicking the list in the desired scrolling direction (flicking condition) on the touch screen device. In the tapping condition, there were seven lines of 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 zone (an orange bar). 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 moves the 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 decrease in the number of visible lines was necessary to accommodate the wheel area. The location of the selection zone was the same as for the tapping condition. Participants could select an item by touching the center circle of the wheel. The circular wheeling area was divided into four sections. Thus, sliding the finger clockwise one quarter of the circle to the right moves 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 selection zone was located in the fifth line. Menu position was moved by several items depending on strength of flicking. In all conditions, if participants reached the top or bottom of the menu, the menu list did not wrap around. All of the sounds were prerecorded as a separate file (16000Hz, 16-bit, Mono) for each menu item.

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 (http://www.research.att.com/~ttsweb/tts/demo.php). 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.

Spindex Cues

Since the attenuated spindex design has been shown to be the most preferred and simplest to implement (Jeon & Walker, 2010), it was used in this experiment. The

attenuated version of spindex contains cues that are attenuated by 20 dB from the first menu item in a letter category. Spindex cues were created by generating TTS files for each letter (e.g., “A”). Each spindex cue consisted of only one syllable, each pronouncing 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, shorter sounds were generated from the second item in that letter category (e.g., /es/ then /s/ for “S”, see Table 2). Spindex cues used in the list were presented before the TTS cues, such that, for example, the “a milli” target item would sound “a”-pause-“a milli”. An interval between the spindex and the TTS was 250ms as used in previous research (Jeon & Walker, 2010) (Figure 2). If participants would tap, wheel, or flick the appropriate area fast, the spindex cues were generated preemptively without a lag between items.

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

1st	**eɪ**	bi:	si:	di:	i:	**ef**	**dʒi:**	**eɪtʃ**	**aɪ**	**dʒeɪ**	**Kei**	el	em
2nd	**ɑ**	bi	si	di	i	**f**	**dʒi**	**h**	**i**	**dʒ**	**k**	el	em
1st	en	oʊ	pi ː	kju ː	ɑ ː (r)	**es**	ti ː	ju ː	vi ː	**dʌblju ː**	**eks**	**waɪ**	zi ː
2nd	en	oʊ	pi	kju	ɑ ː (r)	**s**	ti	ju	vi	**wa**	**s**	**yo**	zi

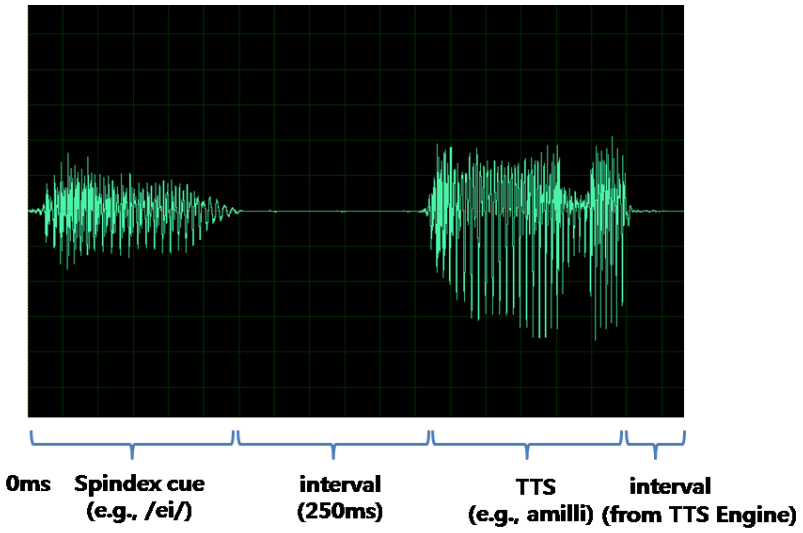


Figure 2. A structure for sound cues and intervals for a single song title.

Design & Procedure

A split-plot design was used in this experiment. There were two between-subjects variables and two within-subjects variables. The between-subjects variables include input gesture style (tapping, wheeling, and flicking) and visual type (on and off). The within-subjects variables involve auditory cue type (TTS-only and TTS + spindex) and block (1-3).

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. There was a short practice session (10-30 seconds) with plain TTS cues before the initial experiment block. In the experiment, one 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.

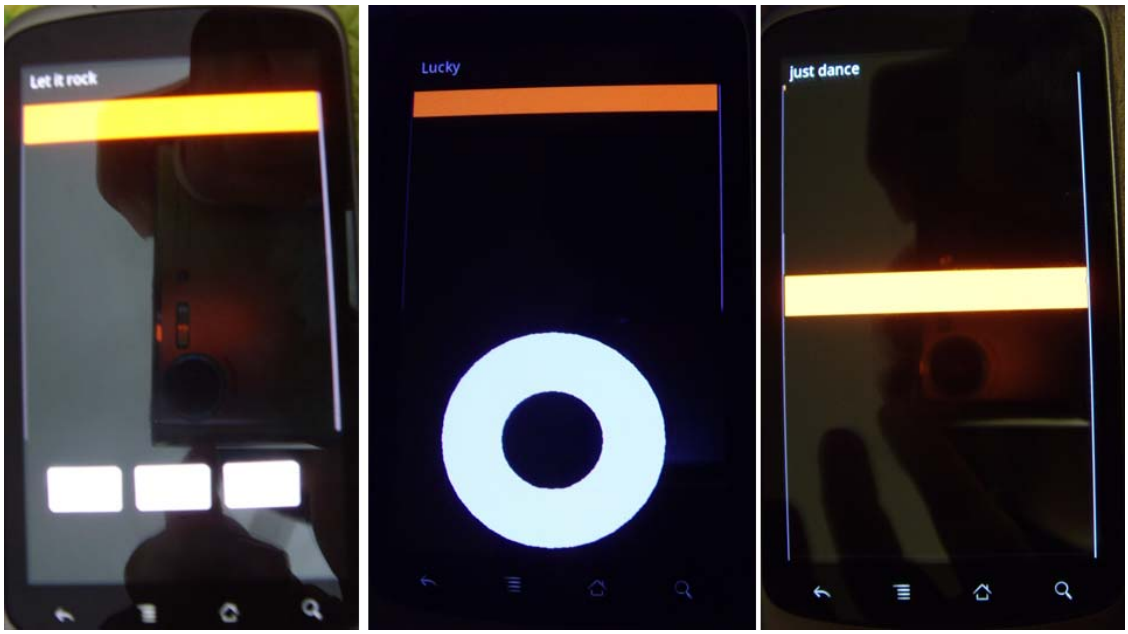


Figure 3. Menus for each input gesture style in the visuals-off condition: the tapping, the wheeling, and the flicking conditions from left to right.

After the informed consent procedure, participants were randomly assigned to one of the six groups (3 input gesture style x 2 visual type). 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 headphone and could adjust the headphones themselves for fit and comfort, as well as the volume level on the phone. Next, they practiced one or two trials with TTS cues to be familiar with how to control the device. Then, the experimental session began.

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 (Figure 3). When participants first touched the available area, the timer started. 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 wheeled around the circular area using their finger to navigate, and pressed the center circle selection area. In the flicking condition, they glided 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 (flicking condition) indicated the selection of the requested target and recorded the end time. 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 (e.g., Hart, 2006) on the desktop computer to report their perceived workload for the navigation task (see Appendix C for detailed questions). While completing the NASA TLX, participants were allowed to have their headset off. Then, they repeated the same procedure for the second condition (15 trials x 3 blocks and NASA TLX again). After finishing two 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 (see Appendix A for the questions in the questionnaires). Finally, participants provided comments on the study.

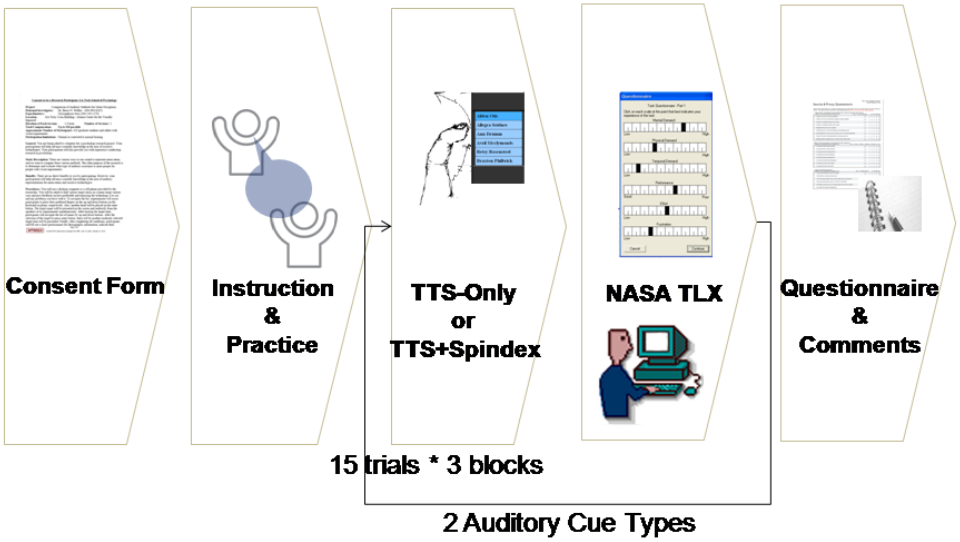


Figure 4. An experimental procedure for each input gesture style. Every participant was assigned one input gesture style and one visual mode with two auditory cue types.

CHAPTER 6

RESULTS

To look at representative objective and subjective evaluation results in one dimension, a 3 (Input gesture type) x 2 (Visual type) 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. A MANOVA found a significant positive effect of adding spindex, $F(2, 115) = 3.818$, $p < .05$, *Pillai's Trace* $= .062$, $\eta_p^2 = .06$. Both time to target and subjective workload showed consistent spindex enhancements in separate univariate test results and there was no interaction or trade-off between two dependent variables. Therefore, subsequent univariate tests (adding block to time to target as a variable) for each dependent measure are described in the following section.

Objective Evaluation

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, I will focus more on the mean time to target and learning rate in the objective evaluation analyses.

Navigation Efficiency & Learning Rate

The time to target results are depicted in Figures 5-11. Results were analyzed with a 3 (Input gesture type) x 2 (Visual type) x 2 (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 consistently showed, regardless of visual type and input gesture type (see Figure 6-7). 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 type 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), ($p < .05$) and the wheeling condition was significantly faster than the tapping condition (M = 23275, SD = 2694), ($p < .001$). However, this main effect was moderated by the interaction effect between input gesture type and visual type, $F(2, 116) = 23.82$, $p < .001$, $\eta_p^2 = .29$. This occurred because the flicking condition showed a more sharp increase of navigation time in the visuals-off condition than other input gesture styles (Figure 9). In the visuals-off condition, time to target of the flicking type (M = 24929) increased to the statistically same level of the tapping condition (M = 25094), $t(39) = .17$, $p = .87$. The interaction between block and visual type was also significant, $F(1.86, 232) = 4.41$, $p < .05$, $\eta_p^2 = .04$ (Figure 10). This interaction term reflects the fact that as block number increased, more learning occurred in the visuals-off condition than in the visuals-on condition. Although the interaction between auditory cue type and block was not significant, and there was no more learning effect between Block 2 (M = 20721) and

Block 3 (M = 20630) in the TTS condition, $t(121) = .38$, $p = .70$, there was still a learning effect between Block 2 (M = 20624) and Block 3 (M = 19977) in the TTS + spindex condition, $t(121) = 2.52$, $p < .05$ (Figure 11). This means that users can benefit more from adding spindex as their level of experience increases, than otherwise.

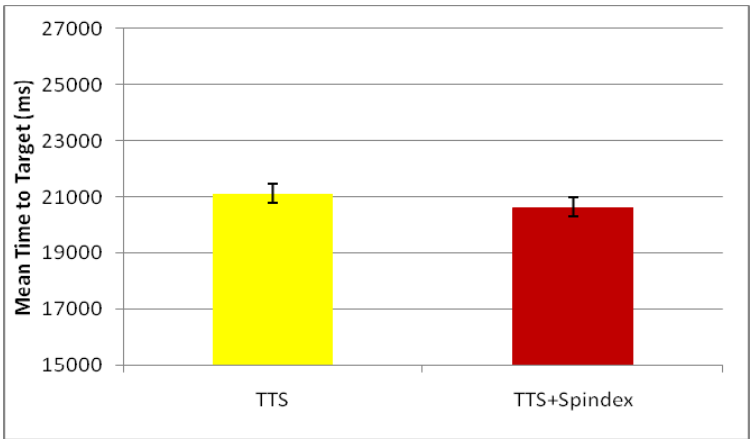


Figure 5. Time to target for auditory cue type. The TTS + spindex condition showed significantly lower navigation time than the TTS condition.

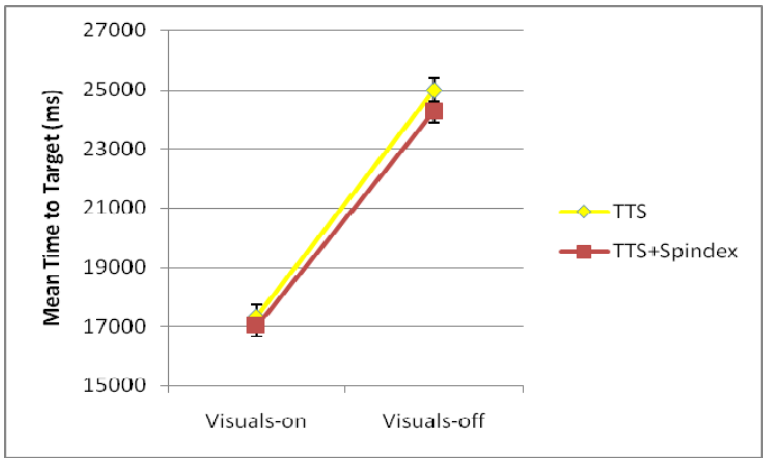


Figure 6. Time to target for visual type and auditory cue type. The enhancement effect of the spindex showed consistently in both visuals-on and visuals-off conditions.

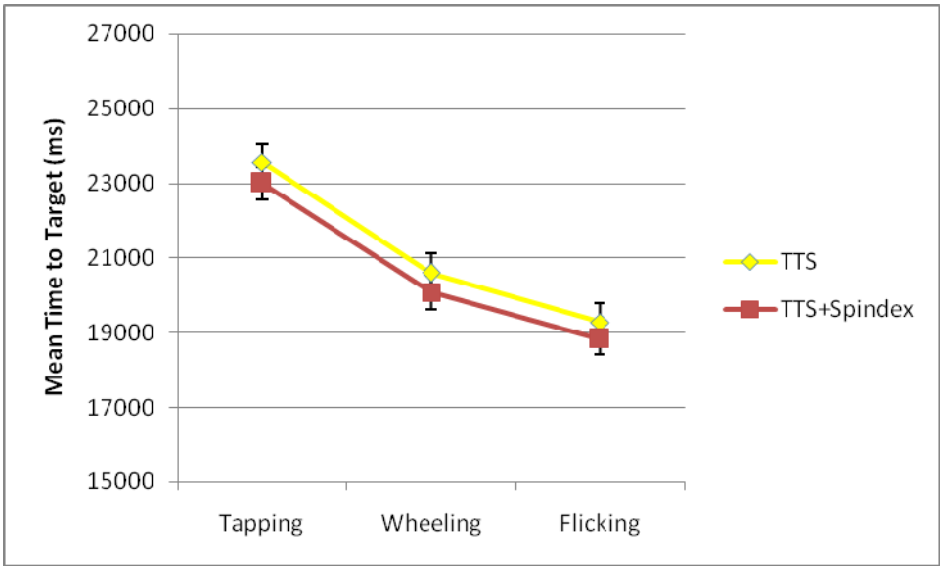


Figure 7. Time to target for input gesture type and auditory cue type. The enhancement effect of the spindex showed consistently across all input gesture types.

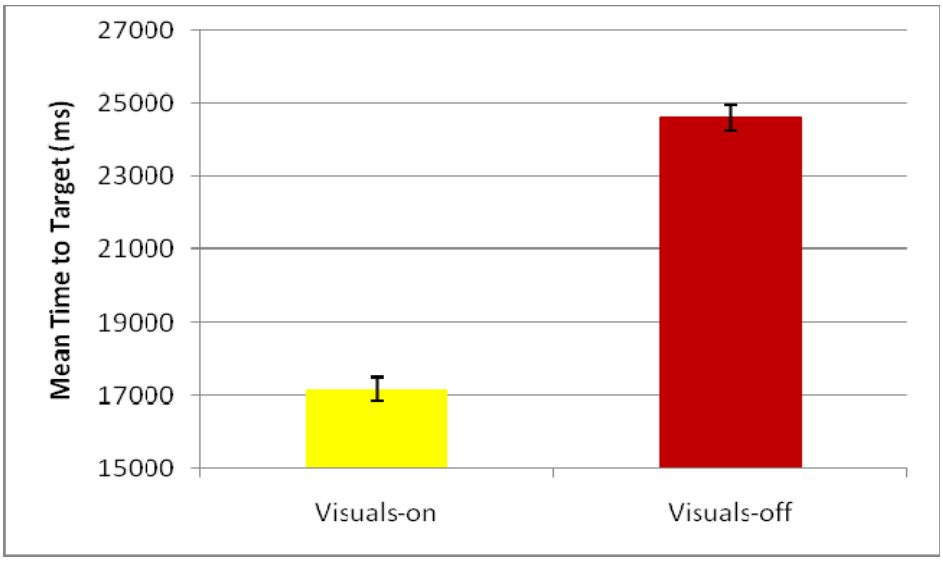


Figure 8. Time to target for visual type. The visuals-on condition showed significantly lower navigation time than the visuals-off condition.

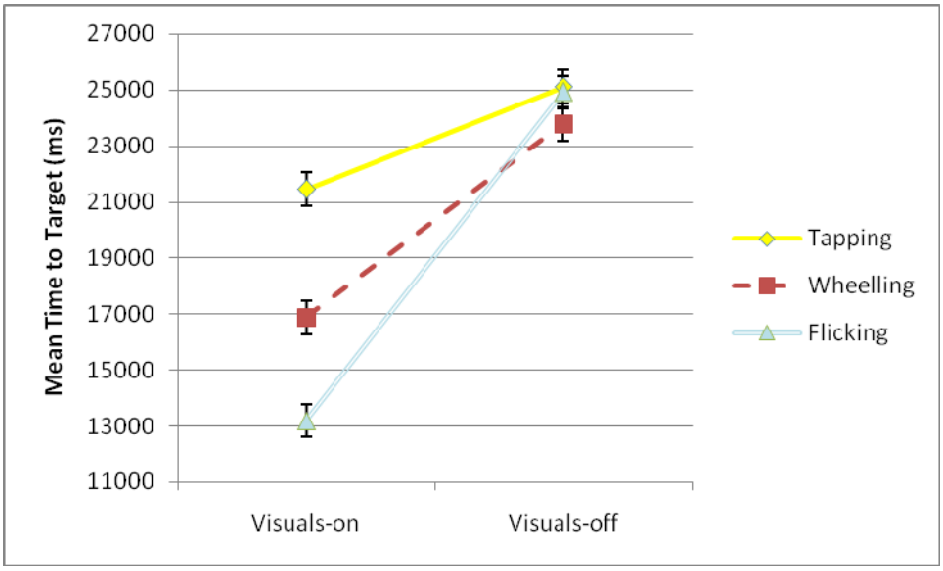


Figure 9. The interaction between visual type and input gesture type. The flicking condition showed a sharp increase in time to target in the visuals-off condition.

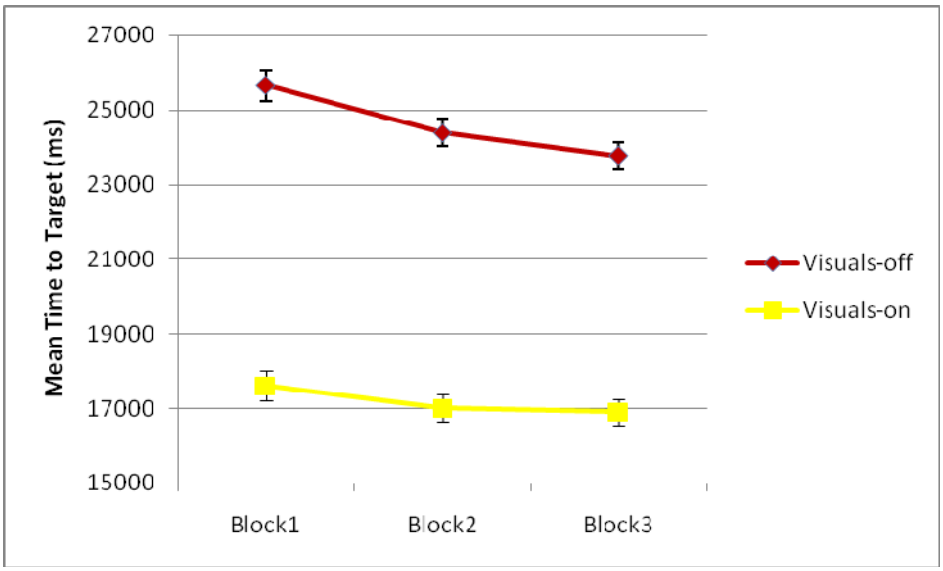


Figure 10. The interaction between block and visual type. A greater learning effect occurred in the visuals-off condition than in the visuals-on condition.

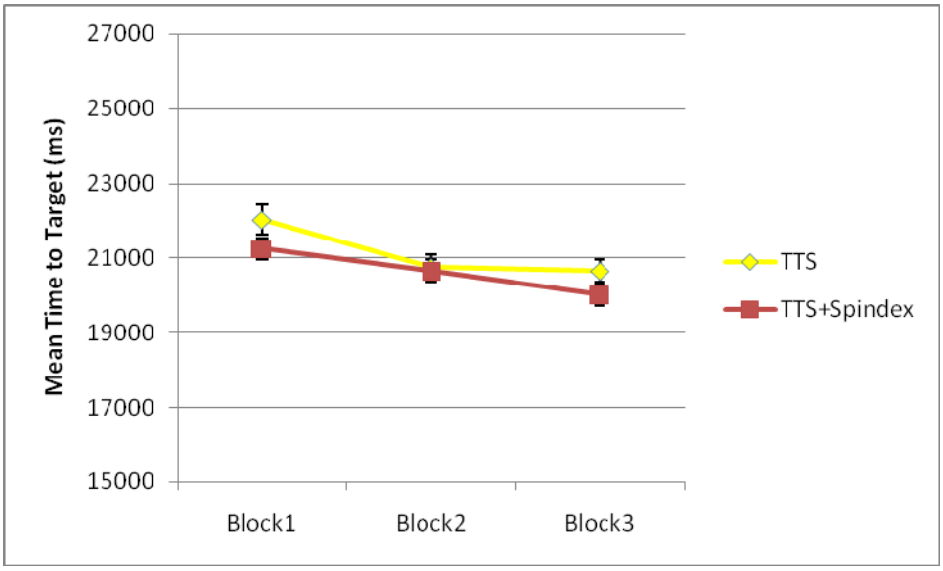


Figure 11. Time to target for block and auditory cue type. There was no interaction between block and auditory cue type. However, more learning took place between Block 2 and Block 3 in the TTS + spindex condition compared to the TTS condition.

Subjective Evaluation

Perceived Workload

Perceived workload scores (NASA-TLX) were also analyzed with a 3 (Input gesture type) x 2 (Visual type) x 2 (Auditory cue type) repeated measures analysis of variance (ANOVA). Perceived workload results are depicted in Figures 12-15. The analysis revealed that adding spindex cues to TTS (M = 51.63, SD = 18.56) reduces 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 consistently showed, regardless of visual type and input gesture type (see Figure 13-14). 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 type showed the significant main effect for 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 each other ($p > .05$). This main effect was moderated by the interaction between input gesture type and visual type, $F(2, 116) = 3.96$, $p < .05$, $\eta_p^2 = .06$. This occurred because the flicking condition showed a sharp increase of the workload in the visuals-off condition (Figure 15) just as was reported for navigation time. Workload scores of the flicking type ($M = 62.90$) in the visuals-off condition increased to the same level as the tapping condition ($M = 65.17$), $t(39) = .46$, $p = .65$.

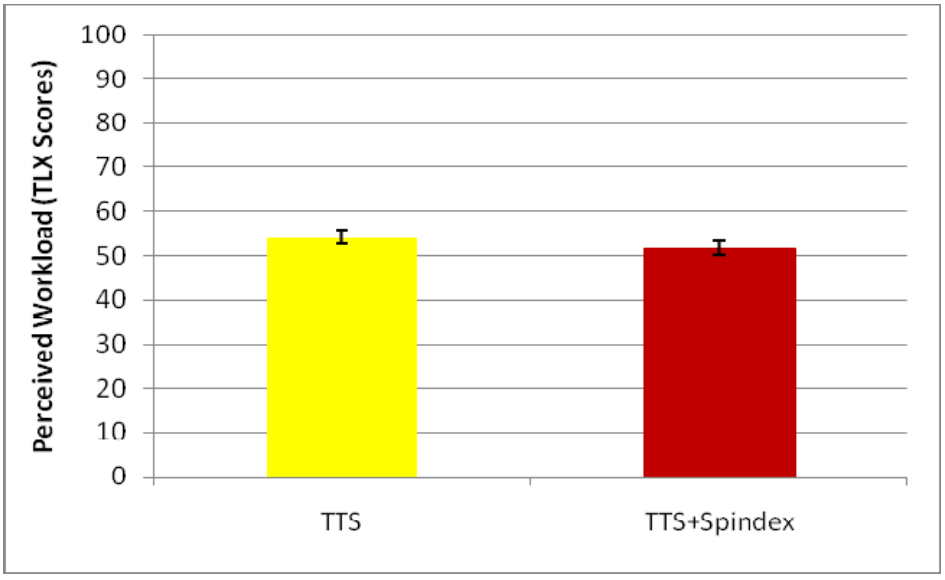


Figure 12. Perceived workload for auditory cue type. The TTS + spindex condition showed significantly lower workload scores than the TTS condition.

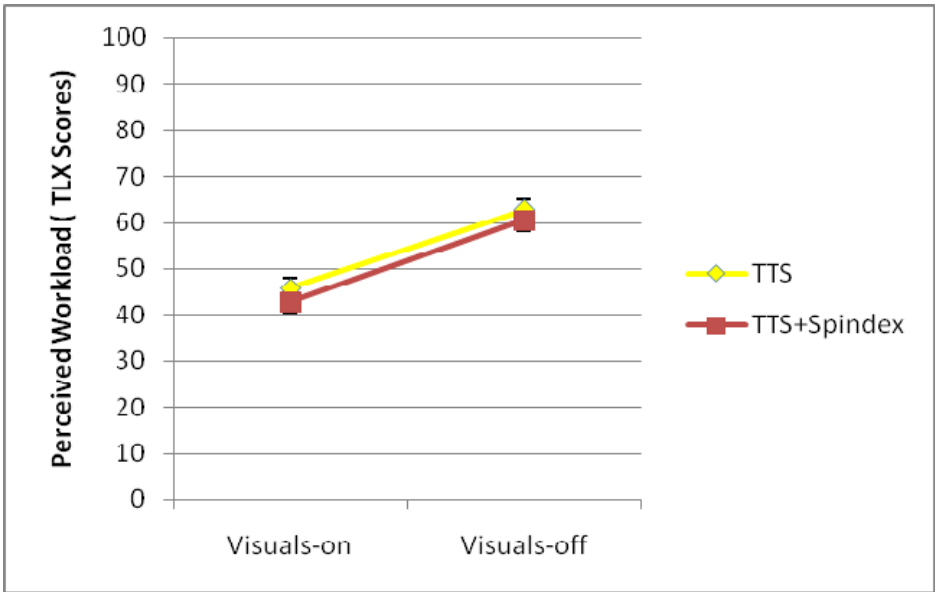


Figure 13. Perceived workload for visual type and auditory cue type. The spindex consistently reduced perceived workload both in visuals-on and visuals-off conditions.

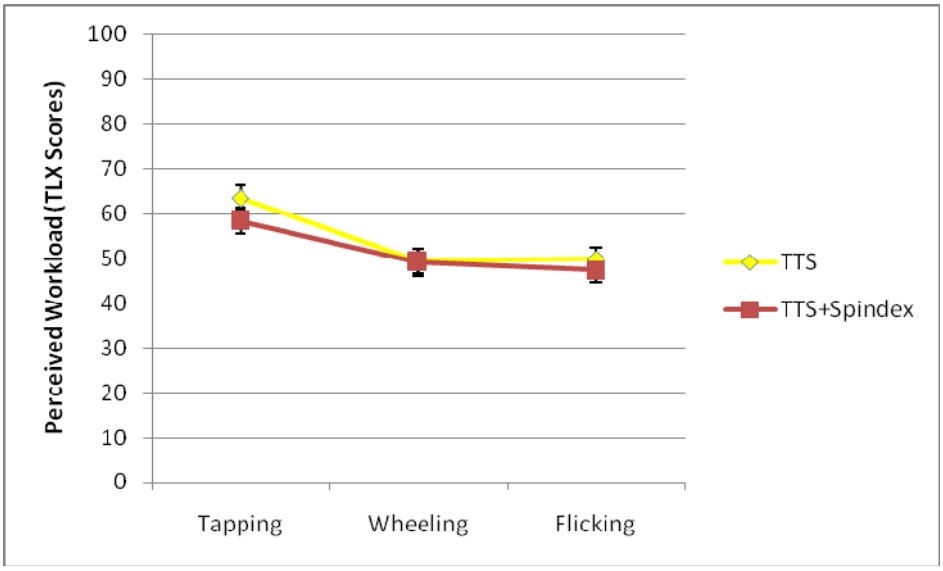


Figure 14. Perceived workload for input gesture type and auditory cue type. The spindex consistently reduced perceived workload across all input gesture styles.

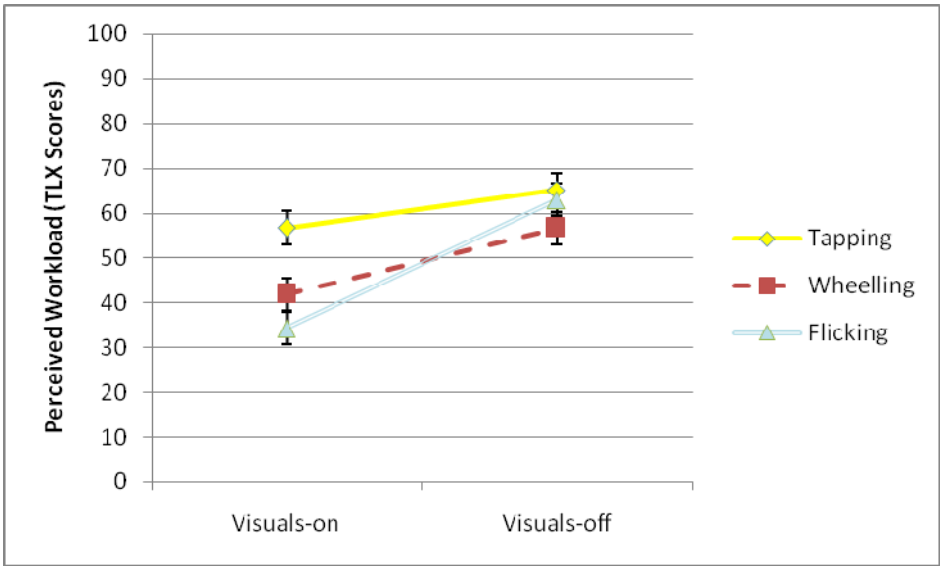


Figure 15. The interaction between visual type and input gesture type. The flicking condition showed a sharp increase in the visuals-off condition.

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 ($M = 4.79$, $SD = 2.65$), $t(121) = -5.58$, $p < .001$.

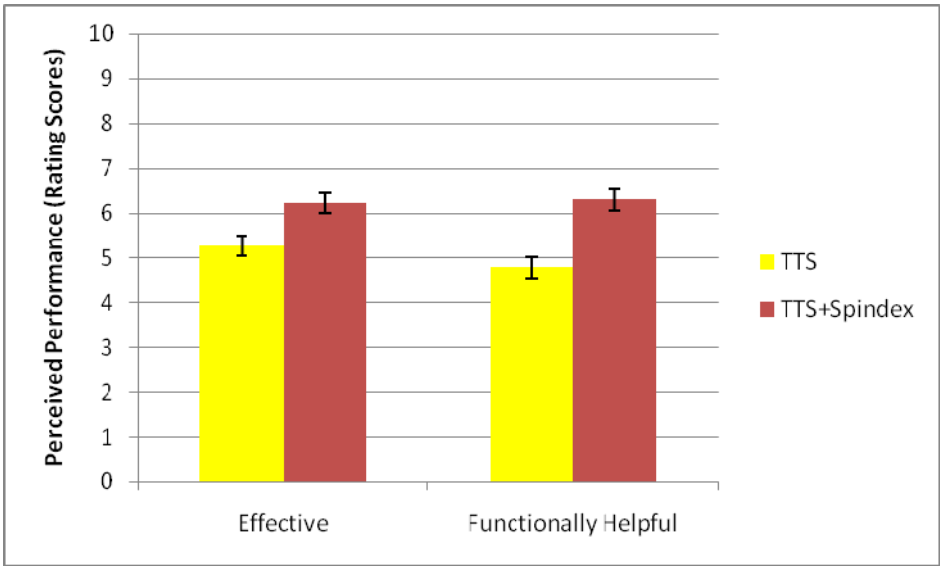


Figure 16. Perceived performance for auditory cue type. Participants rated the TTS + spindex condition significantly higher than the TTS condition on both perceived performance scales.

Subjective Preference

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 $t(121) = -0.29$, $p = .77$, and on the 'annoying' scale $t(121) = 0.30$, $p = .76$.

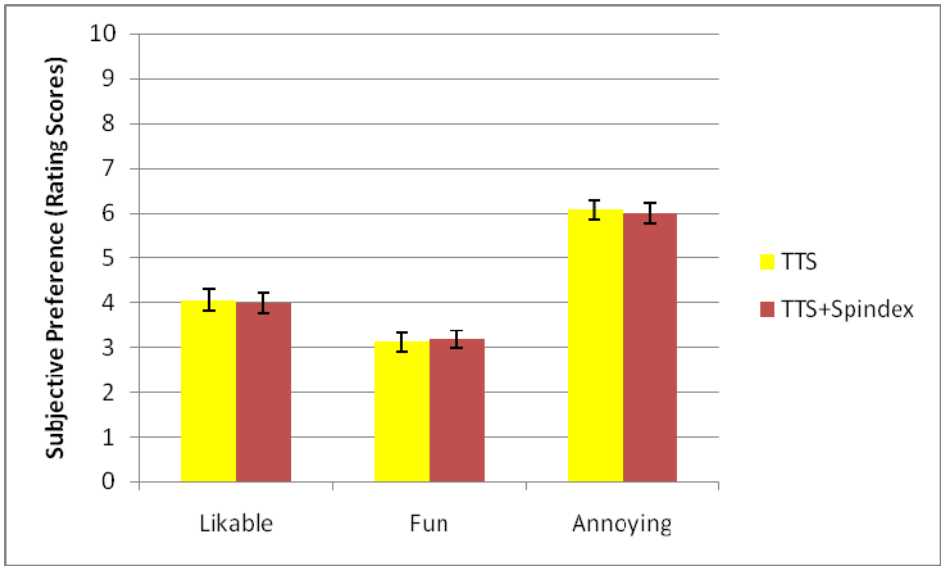


Figure 17. Subjective preference for auditory cue type. There was no difference on subjective preference scores between auditory cue types.

Navigation Behavior Pattern Analysis

In addition to the objective and subjective data analyses, I analyzed participants' navigation behavior patterns with obtained navigation time data according to the interaction between the input method and the output mode (i.e., auditory cue type). From theses analyses, where and how the spindex cues facilitated navigation efficiency was revealed more clearly.

Where to facilitate

First of all, I plotted a mean time to target graph as a function of distance from the top of the menu list to the target item (i.e., bin number of the target) (Figure 18). Overall, as expected, as the target distance increased the navigation time increased. Also, the disparity between the navigation times for the three input gesture styles also increased as the bin number increased.

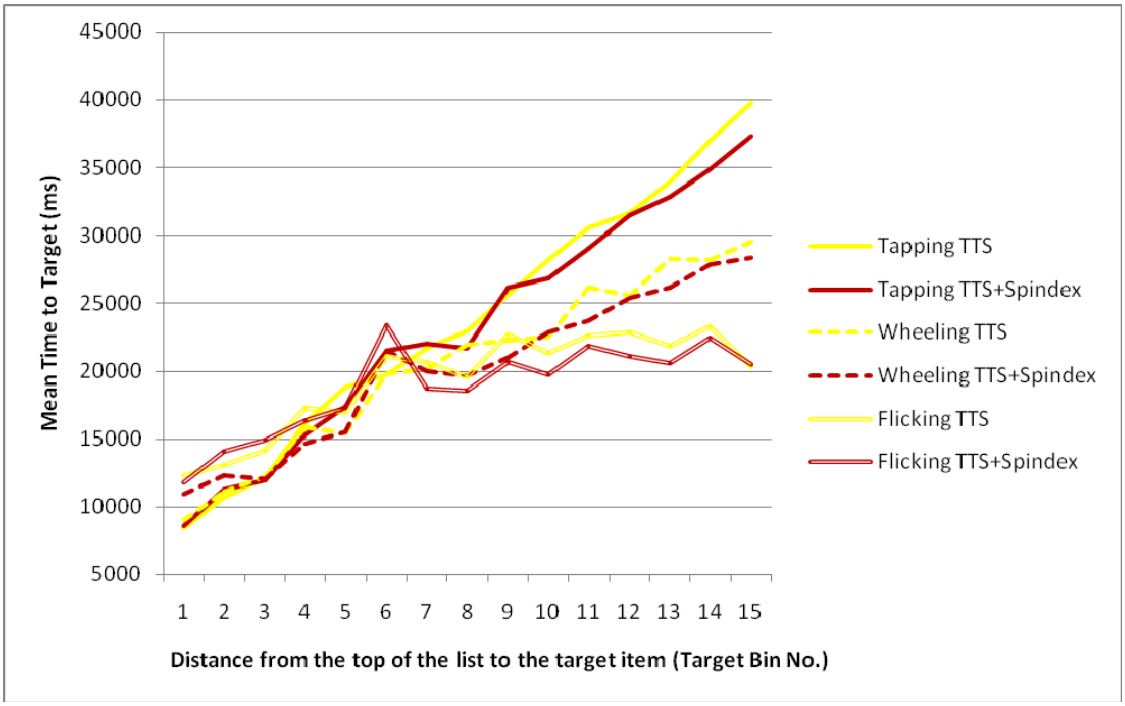


Figure 18. Time to target as a function of target distance.

Regression lines for each input gesture style 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 the fact that while one tapping moves only one item, one wheeling moves four items at once. In contrast to other input gesture styles, the flicking condition showed a power function increase as the increase of the distance to the target (TTS condition: $R^2 = .839$, $y = x^{0.23} + 12133$, TTS + spindex condition: $R^2 = .886$, $y = x^{0.26} + 11758$). In the flicking condition, participants could get to a distant point faster depending on flicking strength than in other

conditions in which 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. Moreover, the data from Bin 1 showed that the flicking had some starting cost, more than the other input gesture types. In all input gesture styles, the slope of the TTS + spindex condition was less steep than that of the TTS condition. Note that in near bins (e.g., Bin 1, 2, 3), a spindex effect did not show well, but as the target distance increases (e.g., after Bin 7), the spindex effect appeared clearly. The increase in navigation time of the spindex condition in the Bin 6 and 7 was due to the fact 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 that participants had to listen to the TTS part more in those zones than in other bins.

How to facilitate

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, I looked into how navigation behaviors were changed in one specific trial as a function of the number of input behaviors for each input gesture style. For these analyses, I selected a trial with a relatively distant target which showed clearer spindex effects.

Because the tapping and the wheeling involve 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 her location in an early stage (Figure 19 and 22). In contrast, in the TTS + spindex

condition, the participant paused fewer times until they got to the target zone (Figure 20 and 23). This behavioral change seems to be in accordance with the two-stage menu navigation strategy. This is discussed more in Chapter 7. In both cumulative figures of the tapping and the wheeling trial (Figure 21and 24), again spindex benefits increase as the number of input actions increases.

In the flicking condition, participants' common strategy was to flick strongly in an early stage (non-target zone), then flick softly several times near the target zone. In the TTS condition, there seemed more soft flicks than in the TTS + spindex condition. Whereas in the TTS condition of the tapping and wheeling condition participants needed more breaks for the status check between inputs, in the flicking condition, needs for the status check resulted in more flicking times. This is supported by analysis with a 2 (Visual type) x 2 (Auditory cue type) x 3 (Block) repeated measures analysis of variance (ANOVA) for the number of flicks. The results revealed a statistically significant difference in auditory cue type and visual type 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 ($M = 75.76$, $SD = 28.46$), $F(1, 38) = 4.29$, $p < .05$, $\eta_p^2 = .10$. Also, the visuals-on condition ($M = 46.24$, $SD = 30.73$) led to significantly fewer flicks than the visuals-off ($M = 98.01$, $SD = 30.75$), $F(1, 38) = 28.29$, $p < .001$, $\eta_p^2 = .43$. In addition, block showed a significant practice effect for the number of flicks, $F(2, 76) = 5.31$, $p = .007$, $\eta_p^2 = .12$.

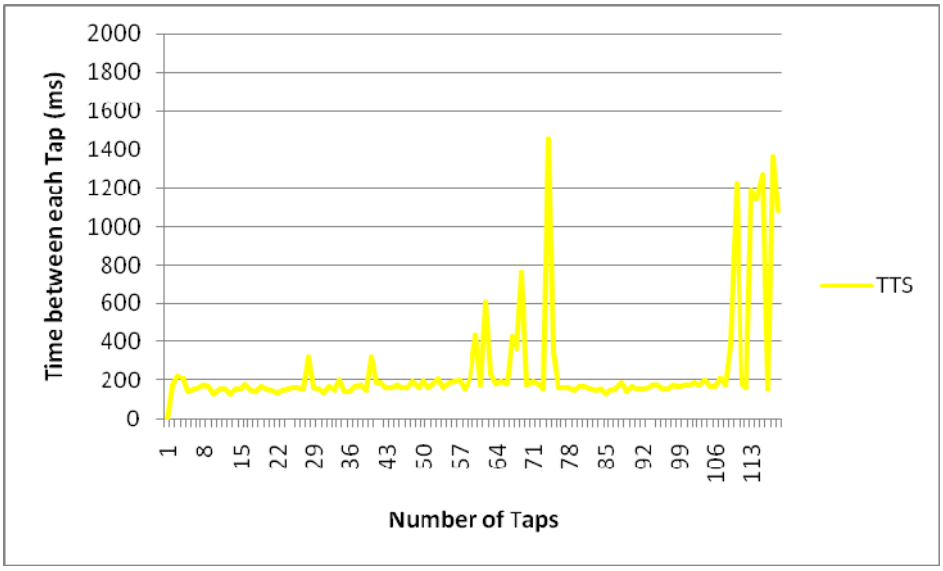


Figure 19. Time between each tap as a function of the number of taps (Participant A in the visuals-off TTS condition Block 1 Trial 7 Target No. 118).

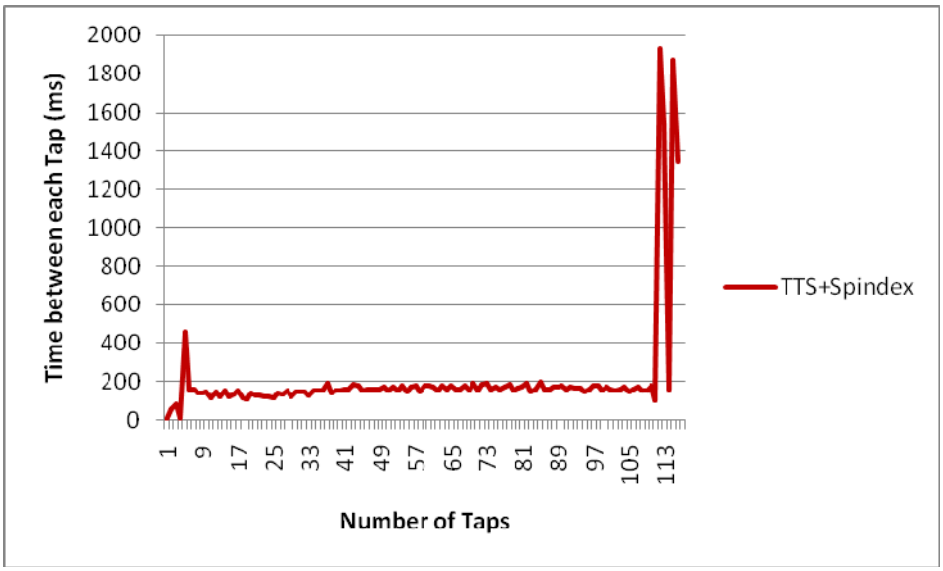


Figure 20. Time between each tap as a function of the number of taps (Participant A in the visuals-off TTS + spindex condition Block 3 Trial 14 Target No. 113).

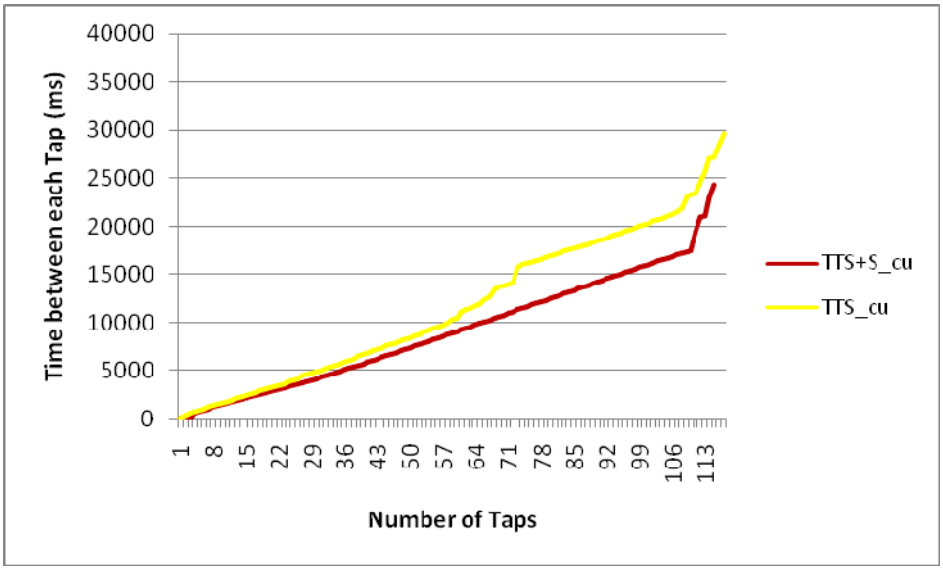


Figure 21. Cumulative time between each tap as a function of the number of taps (Participant A in the visuals-off condition).

.

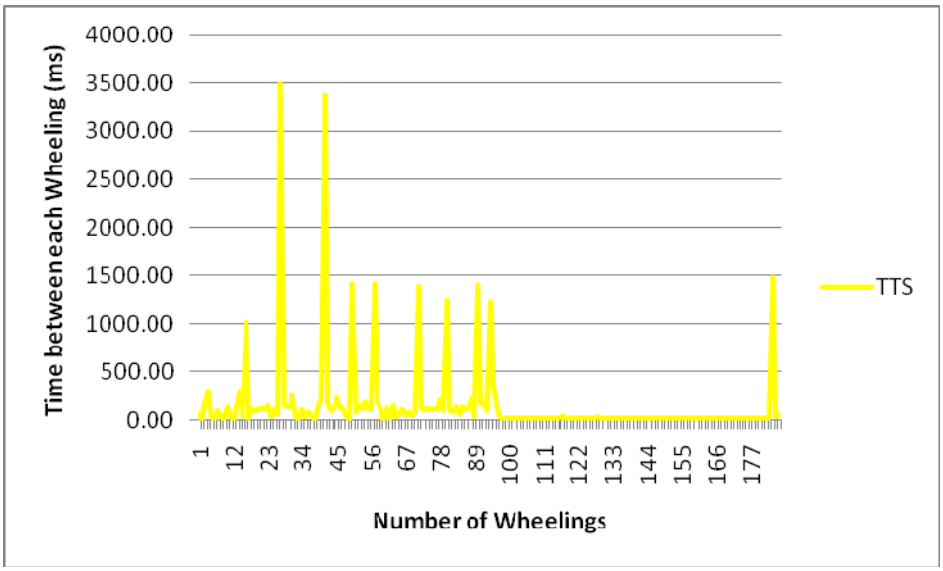


Figure 22. Time between each wheeling as a function of the number of wheelings (Participant B in the visuals-off TTS Condition Block 1 Trial 4 Target No. 112).

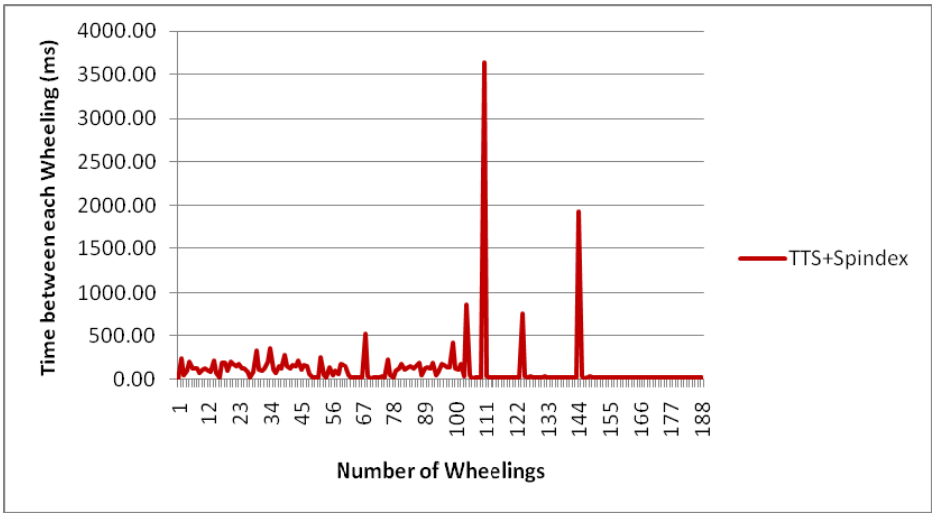


Figure 23. Time between each wheeling as a function of the number of wheelings (Participant B in the visuals-off TTS + spindex condition Block 1 Trial6 Target 114).

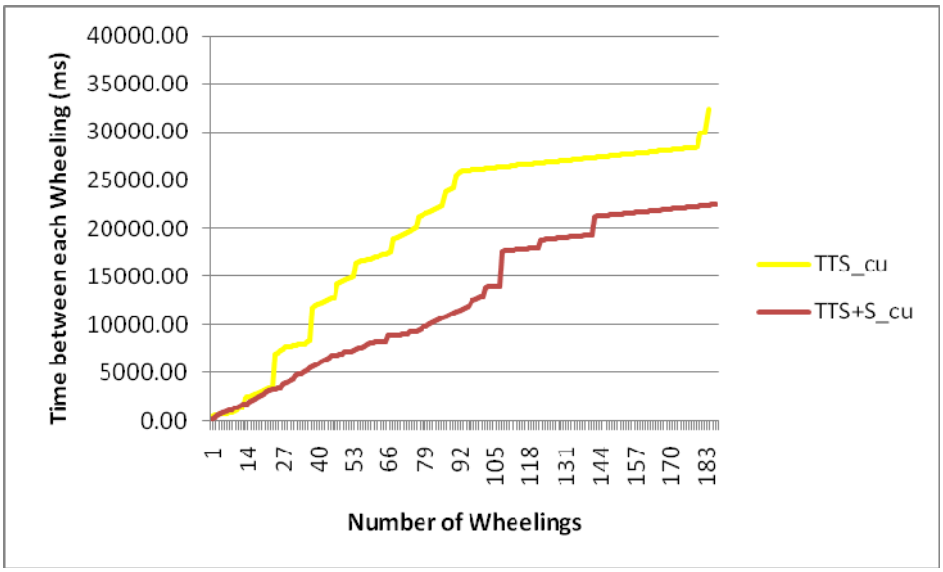


Figure 24. Cumulative time between each wheeling as a function of the number of wheelings (Participant B in the visuals-off condition).

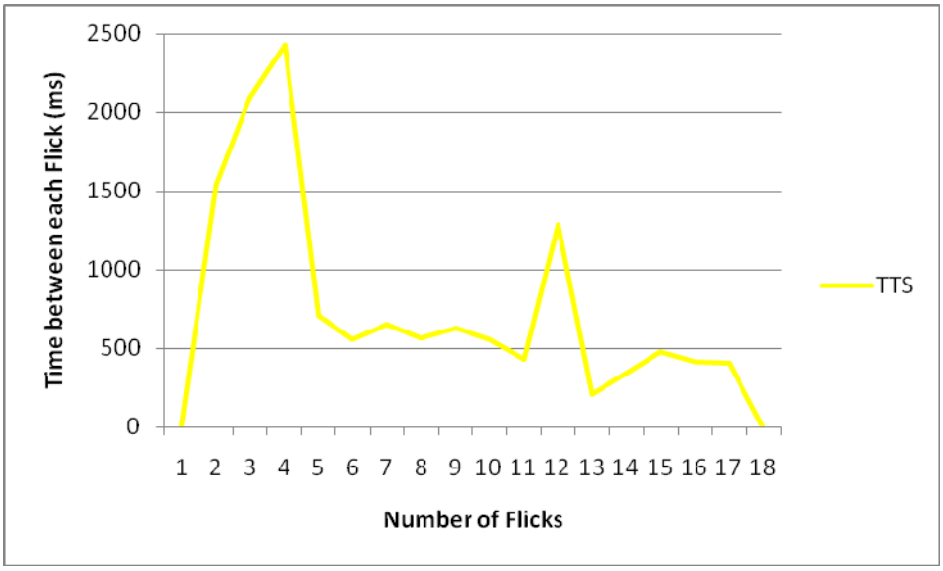


Figure 25. Time between each flick as a function of the number of flicks (Participant C in the visuals-off TTS condition Block Block 1 Trial 5 Target 109).

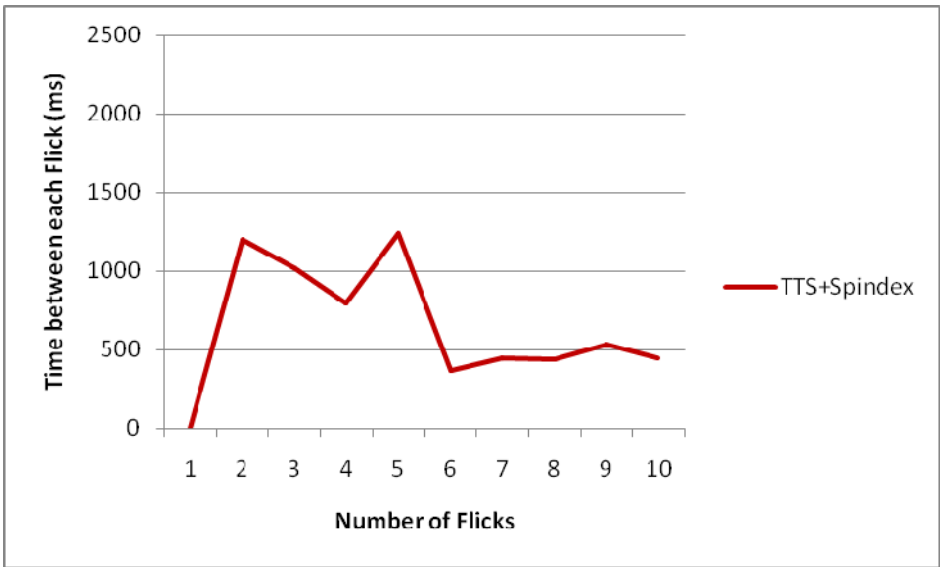


Figure 26. Time between each flick as a function of the number of flicks (Participant C in the visuals-off TTS + spindex condition Block 1 Trial 10 Target 110).

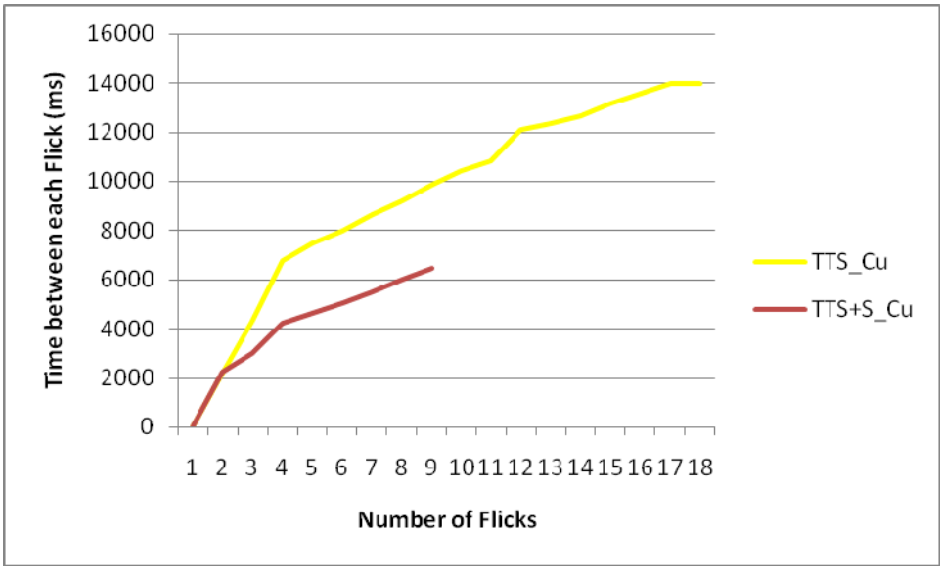


Figure 27. Cumulative time between each flick as a function of the number of flicks (Participant C in the visuals-off condition).

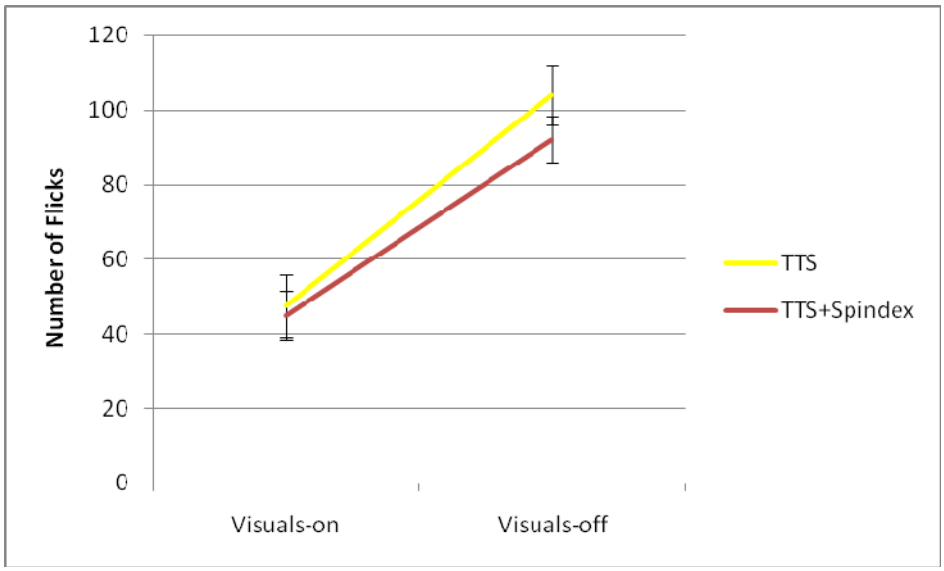


Figure 28. Number of flicks for visual type. The TTS + spindex condition required fewer number of flicks than the TTS condition in both visual types.

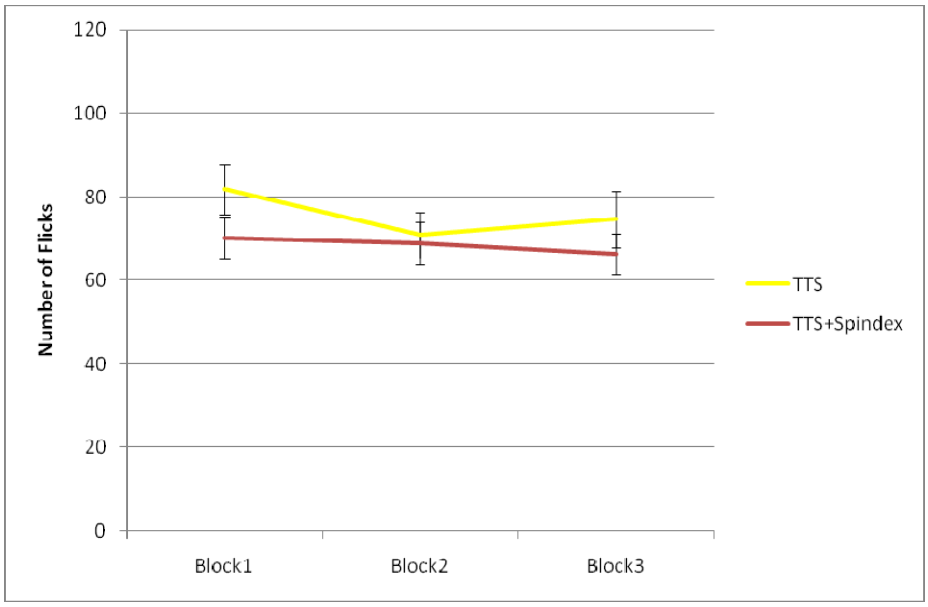


Figure 29. Number of flicks for block. The TTS + spindex condition reduced the number of flicks more consistently as block increased than the TTS condition.

CHAPTER 7

DISCUSSION

Fairly recently, the spindex, a new type of non-speech auditory cue was introduced and showed promise for performance and preference in one-dimensional auditory menu navigation in several studies (Jeon, et al., 2009; Jeon & Walker, 2010). Correspondingly, results in the present study strongly supported the benefits of adding spindex cues to speech menus of a touch screen mobile device in various input gesture styles.

In the experiment, the TTS + spindex condition showed better 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 gesture types. In terms of the universal design, the enhancement of the spindex even in the visuals-on condition showed that this improvement contributed to not only visually impaired people but also sighted people. Also, because there was no difference in error rates between the two auditory cue types, there was no speed accuracy trade-off. Therefore, all of these results generally satisfy the hypotheses of this study. From the results, in addition to spindex benefits, the unique characteristics of each input gesture style were also identified. For instance, the tapping condition showed higher workload ratings than others because tapping required more physical movements than other input gesture styles. On the other hand, in the visuals-off condition, the flicking showed a sharp increase both in navigation time and workload scores. This tells us that flicking is a more visually-demanding task than the others.

Only subjective preference ratings (likable, fun, and annoying) showed no difference between the TTS condition and the TTS + spindex condition. It may be a good sign that adding spindex showed no higher annoyance even though the overall auditory display time increased. Also, because the ratings referred to the 'TTS + spindex' instead of only to the 'spindex', the results should contain the TTS portion in both conditions. These subjective preference results could be different in a dual-task paradigm (Jeon et al., 2009) or if the participants were visually impaired (Jeon & Walker, 2010).

Going one step further, with the navigation time data, where and how the spindex benefits occurred was revealed according to input gesture styles. As can be seen in Figure 18, the spindex might not have changed the behavioral function per se, but slightly changed the slope of the function. The spindex effect in navigation efficiency increases as the target distance increases. Again, this is due to the fact that even small per-item enhancements lead to important and noticeable navigation time efficiency in the list type menu search. In addition, more micro-level analysis showed how the spindex worked in one trial. In the TTS condition of the tapping and wheeling types, participants frequently paused in between taps and wheelings 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 softly in the target zone. Nevertheless, adding the spindex made them flick less, overall.

We can infer that these three behaviors correspond to the two-stage navigation model: *rough navigation* and *fine navigation*. As mentioned in Chapter 2, in the *rough navigation* stage, users exclude non-targets until they approach the alphabetical area

including 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 non-targets. It is enough for them to obtain only enough information to decide whether they are in the target zone or not. After users perceive that they reach the target zone, they then do need the detailed information to compare it with the target. The spindex-enhanced auditory menu can contribute significant per-item speedups in the *rough navigation* stage, and then, the TTS phrase still supports detailed item information in the *fine navigation*. Figures 19 to 27 show time consumption in the *rough navigation* of the TTS condition was much more than that in the *fine navigation* of the TTS + spindex condition.

In addition to a functional approach to navigation efficiency, the spindex effect in terms of perceived workload can be explained 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, but finding a red "O" among many white characters is easier because the oddball color is automatically processed without the full use of attention (Treisman & Gelade, 1980). 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., non-targets) unified (e.g., /si/, /k/, /cha/, /t□i/,… of the "C" → /si/), people can easily '*filter out*' non-targets 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 instead of a deep and linguistic (or semantic) level. It may require merely pre-attentive and automatic processing. Reflecting exactly the same notion, one participant commented that "the softer voice [spindex cues]

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 felt less 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 temporal resolution, such as in a monitoring task (Kramer, 1994) or detecting task for the correct auditory signal in streaming (Walker & Kramer, 2004). Therefore, if adding the spindex cues to the TTS can change an auditory search task (which needs active role of attentional processing) into a monitoring task (which does not require that effort), the psychological 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 thus, might be blurred.

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 & Elkerton, 1989) suggested that visual indexes could decrease visual search time with list menus. I would explain that the spindex is a successful translation of the index from the visual display into the auditory display realm.

Despite these positive results, the spindex can be still improved. Although users can benefit from adding spindex cues in the *rough navigation*, in the *fine navigation* it

takes more time for them to hear both the spindex part and TTS part for fine tuning. 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]." There are two plausible solutions for this issue. 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 cue and 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 weak, the speech menu system could speak out only the item itself. On the other hand, if the user input is faster or stronger, the device would generate spindex cues only. Then, there would be less or no time sacrifice due to the spindex and the interval even in the *fine navigation*.

From the perspective of practical applications for touch screen devices, the spindex has several advantages. First, the spindex does not require any major change of the programming arhitecture 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 better user experience in their navigation task. Especially with the new input gesture styles of touch screen mobile devices (e.g., fliking), user experience should be fun, engaging, and creative (Russell &

Bryan, 2009) because to a certain user population, such as the “thumb generation”, the mobile phone serves as an entertainment object (Jeon, Na, Ahn, & Hong, 2008). Finally, it is also encouraging that using spindex cues requires little or no practice to benefit from them, which indicates a low threshold for new users. These advantages can significantly increase the possibility of applying spindex cues in real devices.

For constructing a comprehensive and robust auditory menu navigation theory, this research can be extended to the point where we can figure out where and what the optimal combinations of each non-speech auditory cue (e.g., earcons, spearcons, auditory scrollbars, and spindexes) would be in every case. In addition to this ‘gradual disclosure’ of auditory contents with respect to a ‘time’ domain in a ‘target search’ behavior, researchers can also examine other behavioral patterns such as ‘browsing’ or ‘exploring’ the ‘streaming’ auditory menus in a ‘spatial’ domain without a specific target. Moreover, future research should be carried out in real mobile contexts such as walking, jogging, or driving to gain more practical relevance (Fisk & Kirlik, 1996).

CHAPTER 8

CONCLUSION

In this study, I examined whether the benefits of using spindex cues in auditory menus could be generalized to mobile touch screen devices, with respect to specialized input gesture styles such as tapping, wheeling, and flicking.

The results showed that spindex cues do produce the objective and subjective usability improvements irrespective of the input gesture style and visual mode. Implications from these findings could contribute to forming a fundamental theory of auditory menu navigation. Moreover, researchers and designers of auditory interfaces could use the usability metrices and practical results of this study for implementation of auditory user interfaces and further design evaluation.

APPENDIX A. QUESTIONNAIRE

Questionnaire for Speech Index Study

Participant #:_____

Demographics

1. Age: _____
2. Gender: M / F
3. Handedness
4. Experience on Touch Screen Device
   - name of device ( ) – year of use ( )
   - name of device ( ) – year of use ( )

1. Please circle your level of performance for each of the followings.

[Perceived Performance of the auditory cues]

How effective was this sound	0 1 2 3 4 5 6 7 8 9 10 Not at all effective very effective
How functionally helpful was this sound	0 1 2 3 4 5 6 7 8 9 10 Not at all helpful very helpful

2. Please circle your level of likability for each of the followings.

[Likability of the auditory cues]

How likable was this sound	0 1 2 3 4 5 6 7 8 9 10 Not at all Likable Very Likable
How fun was this sound	0 1 2 3 4 5 6 7 8 9 10 Not at all Fun Very
How annoying was this sound	0 1 2 3 4 5 6 7 8 9 10 Not at all Annoying Very Annoying

3. Please write your comments/suggestions to help us with this research?

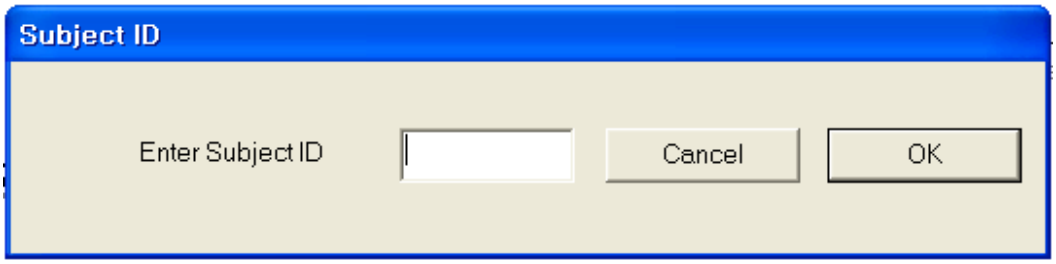
_____

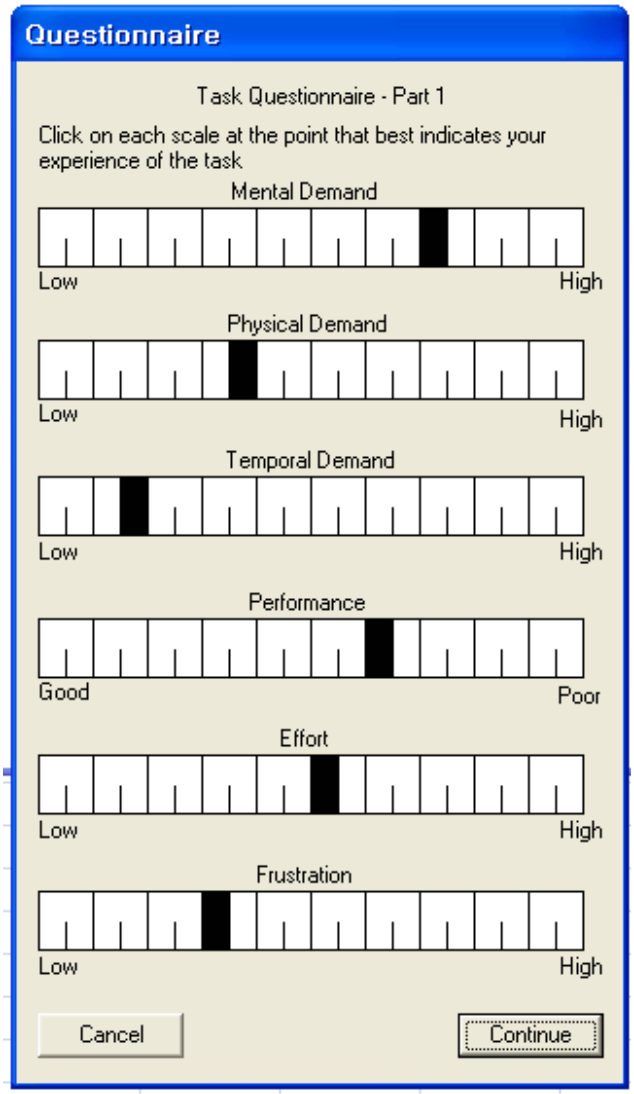
_____

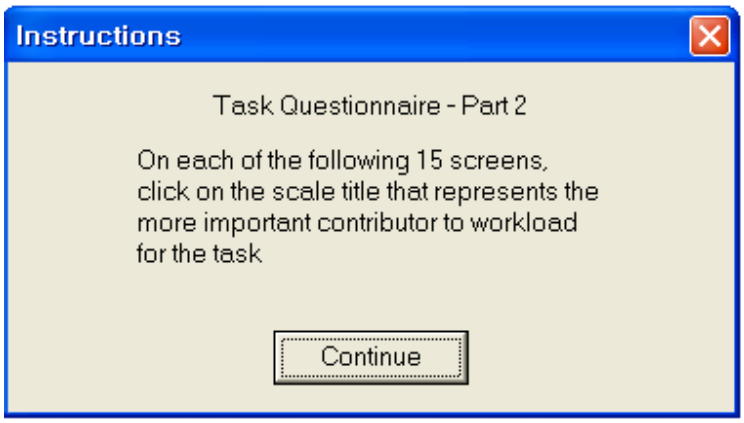
APPENDIX B. A SONG LIST (150 SONGS)

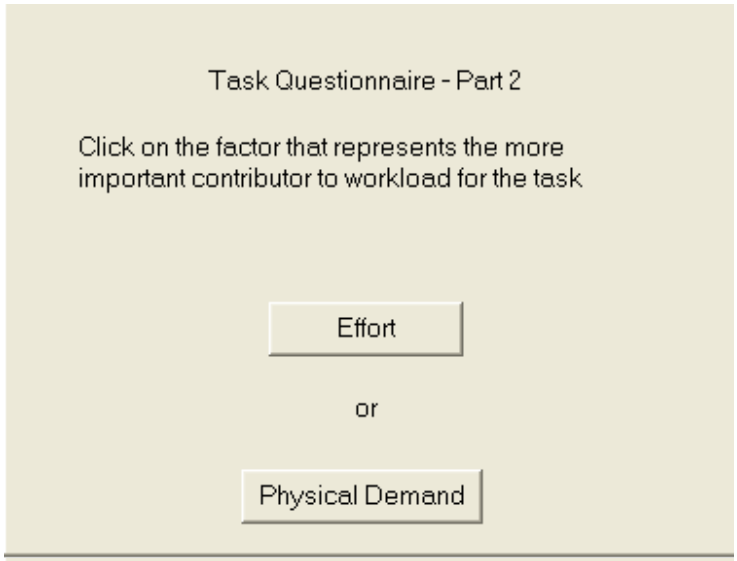
a milli
all around me
All the above
Always the love
Amazing
America's suitehearts
Apologize
Beautiful
Begging
Best days of your life
Blame It
Bleeding love
Boom Boom Pow
Break the ice
bubbly
Bust it baby
Butterfly fly away
Bye bye
Candles
Careless whisper
Chicken Fried
circus
Come on get higher
crazier
Damaged
Day and Nite
Dead And Gone
Disturbia
Don’t stop the music
Don’t trust me
Don't forget
don't stop believing
Elevator
Feels like tonight
Forever
four minutes
Funny the way it is
gives you hell
Goodbye
Gotta be somebody
Halle Berry
halo
Heartless
Here comes goodbye
Hoedown Throwdown
Hot and Cold
How do you sleep?
I do not hook up
I hate this part
I know you want me
I love college
I run to you
I saw god today
I told you so
If Today was your last day
If U seek amy
I'm still a guy
I'm yours
In love with a girl
Independent
It happens
It won't be like this for long
It's America
It's not my time
Jai Ho
just dance
Just got started lovin' you
Kids
Killa
Kiss a girl
Kiss me thru the Phone
Knock you down
Know your enemy
La La land
last name
Leaving
Let it rock
Let's get crazy
Lollipop
Love Game
Love in this club
love is a beautiful thing
Love song
love story
Low
Lucky
Mad
My life would suck without you
No air
No one
Not meant to be
One in every crowd
One two three four
Our song
Paralyzer
Party people
People are crazy
Picture to burn
Please don't leave me
Pocketful of sunshine
Poker Face
Prom Queen
Realize
Right Round
Rockin' that Thang
Say
second chance
See you again
Shake it
She got it
She's country
Show me what I'm looking for
Sideways
Single ladies
Sissy's song
So what
sober
Sorry
Soulmate
Starstruckk
Stop and stare
Sugar
Superstar
Tattoo
Teardrops on my guitar
That's not my name
The boss
the climb
The way that I love you
Then
Thinking of you
Touch my body
Turn my swag on
Turnin me on
Untouched
Use somebody
viva la vida
Waking up in Vegas
We made you
Welcome to the world
What you got
Whatever it is
White horse
With you
Womanizer
You belong with me
You can get it all
You found me
You'll always find your way back home
You're gonna miss this

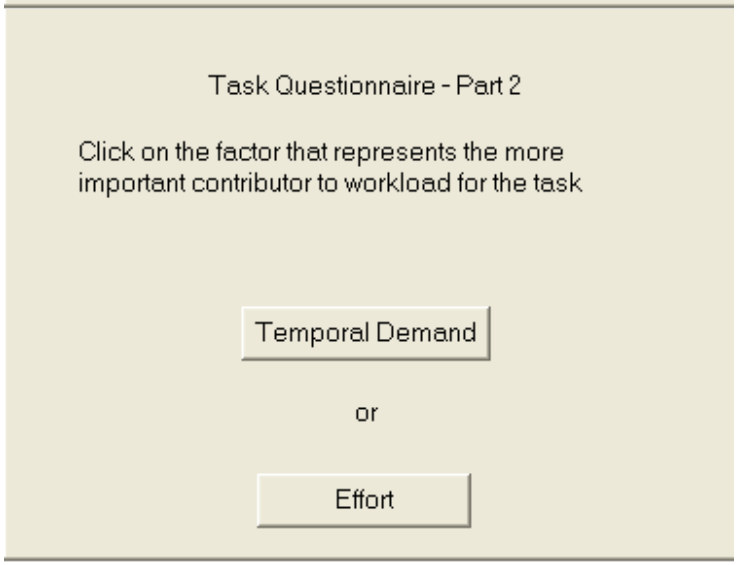
APPENDIX C. ELECTRONIC NASA TLX SCREENSHOTS FOR PERCEIVED WORKLOAD

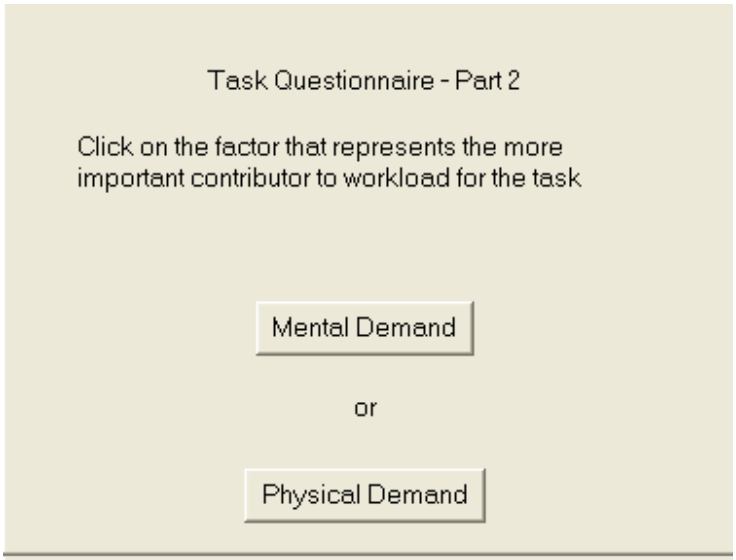


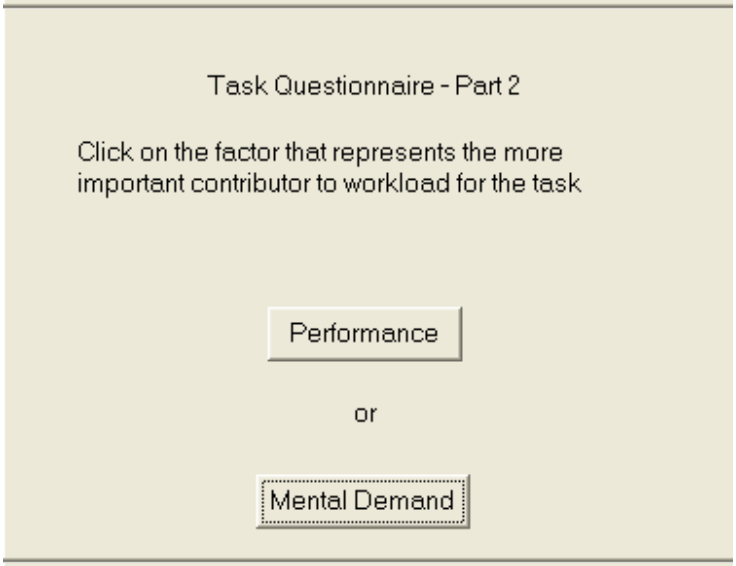


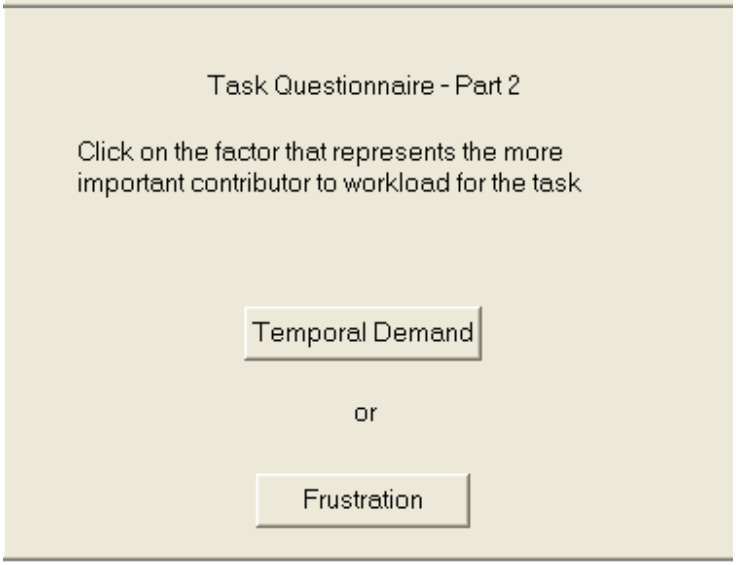

Instructions
Task Questionnaire - Part 2
On each of the following 15 screens,
click on the scale title that represents the
more important contributor to workload
for the task
Continue

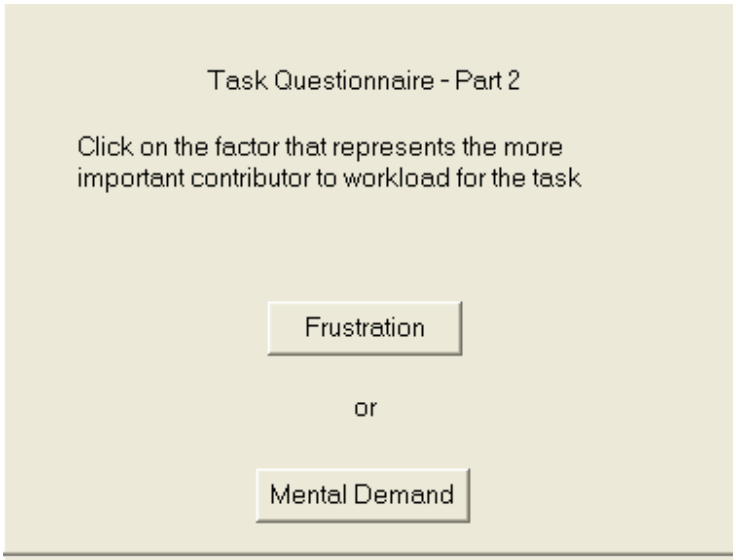

Task Questionnaire - Part 2
Click on the factor that represents the more
important contributor to workload for the task
Effort
or
Physical Demand

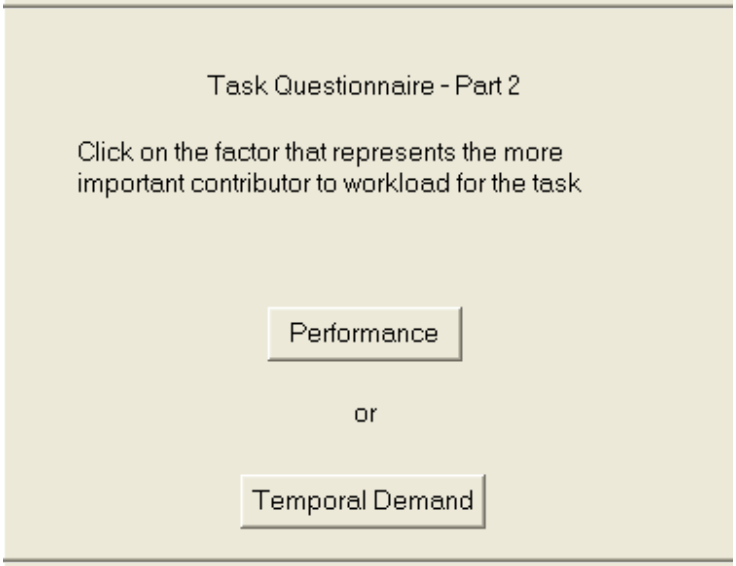

Task Questionnaire - Part 2
Click on the factor that represents the more
important contributor to workload for the task
Temporal Demand
or
Effort

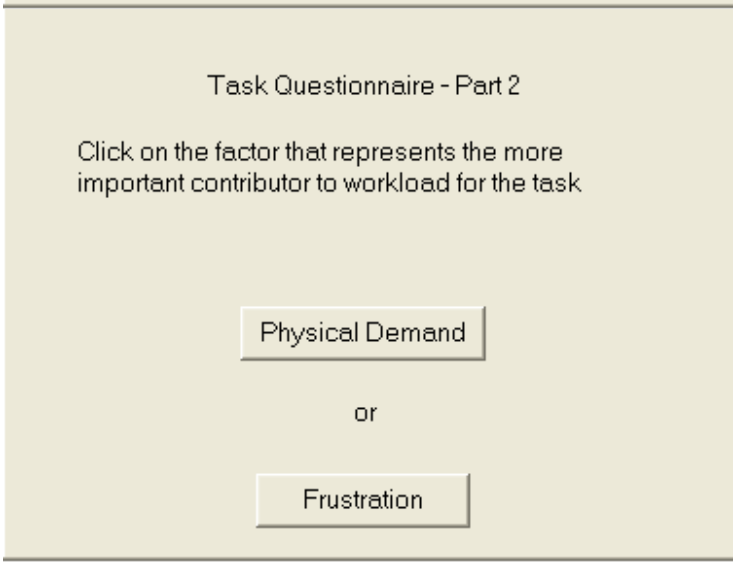

Task Questionnaire - Part 2
Click on the factor that represents the more important contributor to workload for the task
Mental Demand
or
Physical Demand

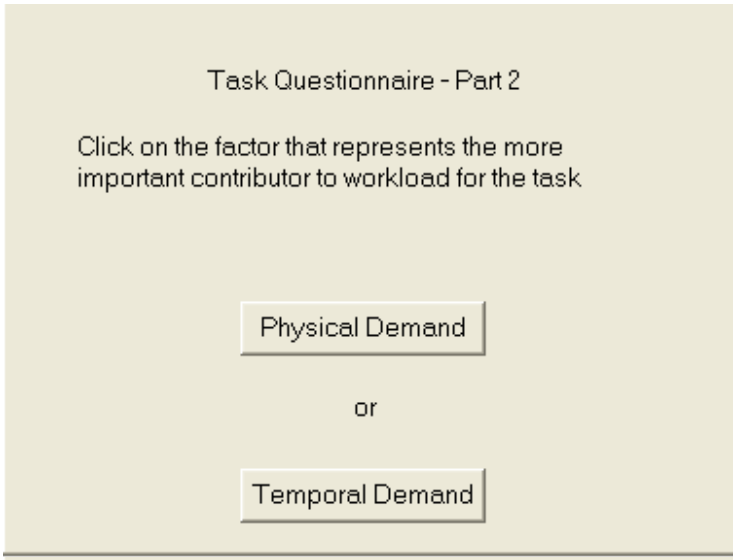

Task Questionnaire - Part 2
Click on the factor that represents the more important contributor to workload for the task
Performance
or
Mental Demand

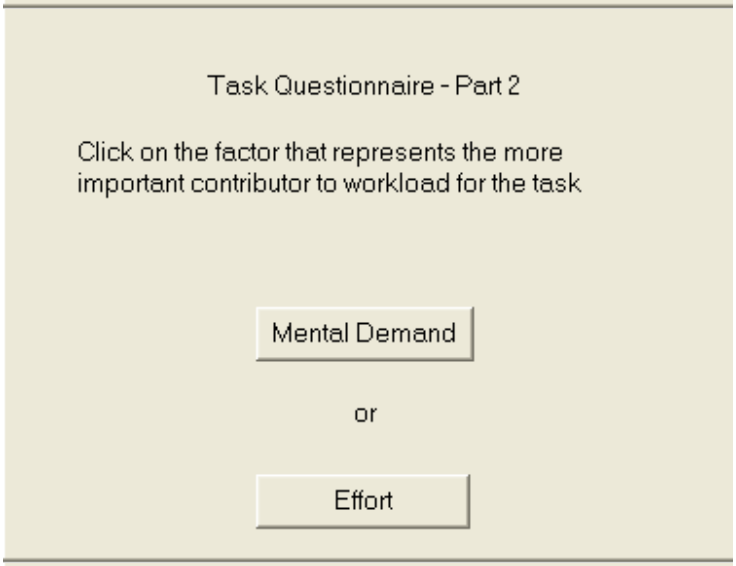

Task Questionnaire - Part 2
Click on the factor that represents the more important contributor to workload for the task
Temporal Demand
or
Frustration

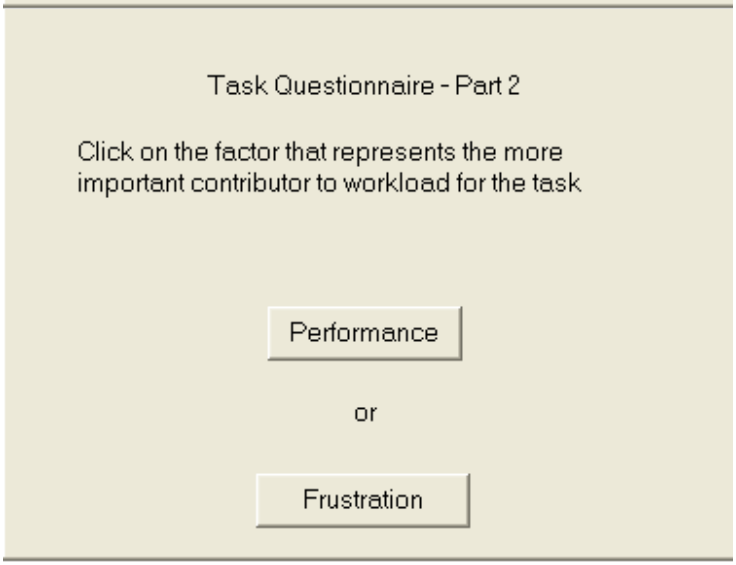

Task Questionnaire - Part 2
Click on the factor that represents the more important contributor to workload for the task
Frustration
or
Mental Demand

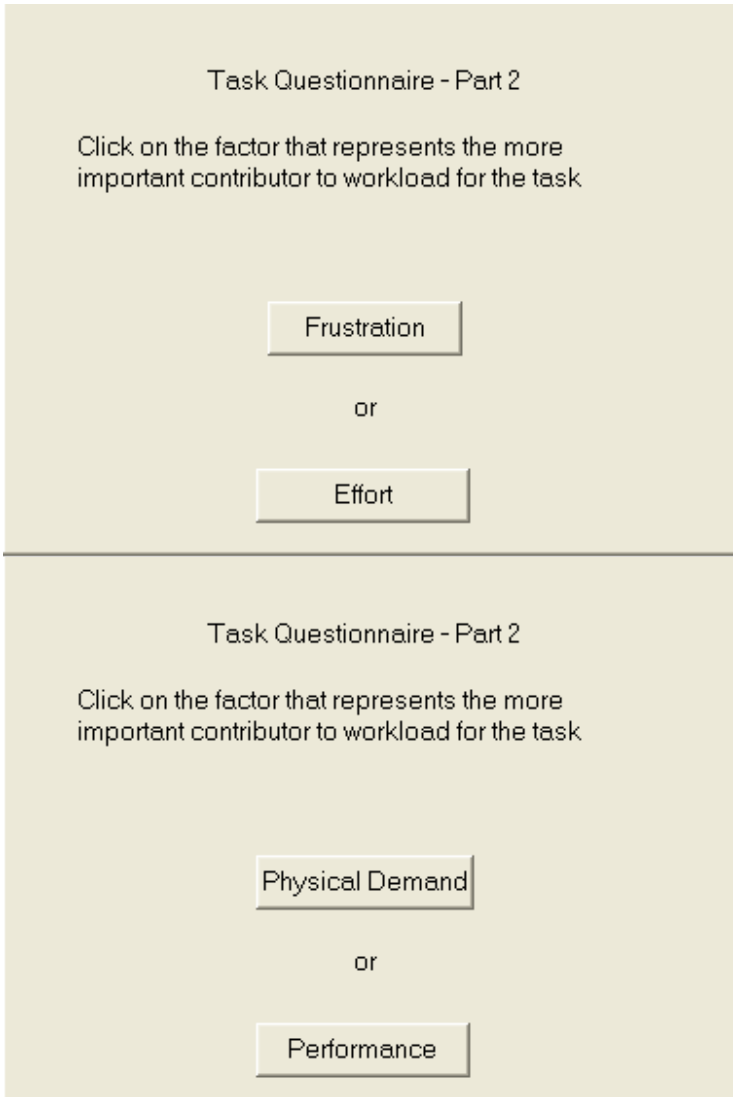

Task Questionnaire - Part 2
Click on the factor that represents the more important contributor to workload for the task
Performance
or
Temporal Demand

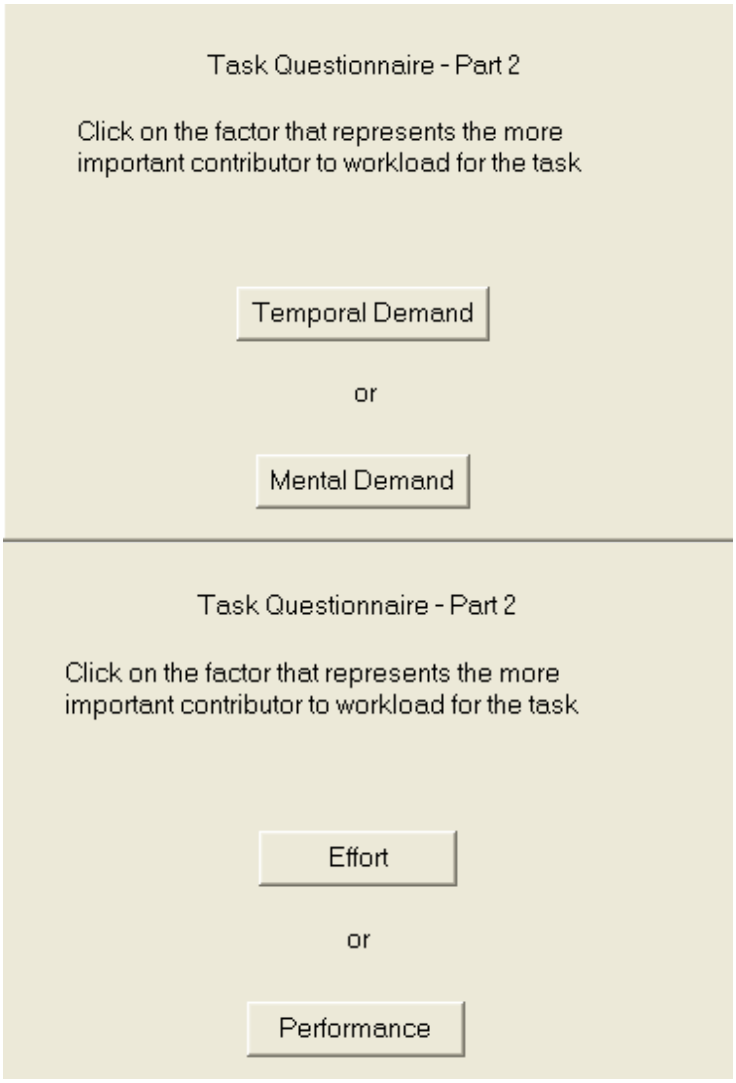

Task Questionnaire - Part 2
Click on the factor that represents the more important contributor to workload for the task
Physical Demand
or
Frustration


Task Questionnaire - Part 2
Click on the factor that represents the more important contributor to workload for the task
Physical Demand
or
Temporal Demand


Task Questionnaire - Part 2
Click on the factor that represents the more important contributor to workload for the task
Mental Demand
or
Effort


Task Questionnaire - Part 2
Click on the factor that represents the more important contributor to workload for the task
Performance
or
Frustration


Task Questionnaire - Part 2
Click on the factor that represents the more important contributor to workload for the task
Frustration
or
Effort
Task Questionnaire - Part 2
Click on the factor that represents the more important contributor to workload for the task
Physical Demand
or
Performance


Task Questionnaire - Part 2
Click on the factor that represents the more important contributor to workload for the task
Temporal Demand
or
Mental Demand
Task Questionnaire - Part 2
Click on the factor that represents the more important contributor to workload for the task
Effort
or
Performance

REFERENCES

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

Arons, B. (1997). SpeechSkimmer: A system for interactively skimming recorded speech. *ACM Transactions on Computer-Human Interaction, 4*(1), 3-38.

Beck, D., & Elkerton, J. (1989). Development and evaluation of direct manipulation list. *SIGCHI Bulletin, 20*(3), 72-78.

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

Brewster, S. 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. *Personal and Ubiquitous Computing, 6*(3), 188-205.

Brewster, S. A. (2008). Chapter13: Nonspeech auditory output. In A. Sears & J. Jacko (Ed.), *The human computer interaction handbook* (pp. 247-264). New York: Lawrence Erlbaum Associates.

Brewster, S. A., & Cryer, P. G. (1999). Maximising screen-space on mobile computing devices. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'99)* (pp. 224-225), Pittsburgh, PA, USA.

Brewster, S. A., Leplatre, G., & Crease, M. G. (1998). Using non-speech sounds in mobile computing devices. *Proceedings of the 1st Workshop on Human Computer Interaction with Mobile Devices*, Glasgow, UK.

Brewster, S. A., Lumsden, J., Bell, M., Hall, M., & Tasker, S. (2003). Multimodal 'eyes-free' interaction techniques for wearable devices. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI2003)* (pp. 473-480), Florida, USA.

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

Brewster, S. A., Wright, P. C., & Edwards, A. D. N. (1992). A detailed investigation into the effectiveness of earcons. *Proceedings of the 1st International Conference on Auditory Display (ICAD94)* (pp. 471-478), Santa Fe, USA.

Davison, B. D., & Walker, B. N. (2008). AudioPlusWidgets: Bringing sound to software widgets and interface components. *Proceedings of the International Conference on Auditory Display (ICAD2008)*, Paris, France.

Edwards, A. D. N. (1989). Soundtrack: An auditory interface for blind users. *Human-Computer Interaction, 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'. *Interactive with Computers, 10*, 401-409.

Fisk, A. D., & Kirlik, A. (1996). Practical relevance and age-related research: Can theory advance without practice? In W. A. Rogers, A. D. Fisk, & N. Walker (Eds.),

Aging and skilled performance: Advances in theory and application (pp. 1-15). Mahwah, NJ:Erlbaum.

Fritz, M. (2000). Keys to the kiosk -- The temptations of touch-screen. *Emedia, 13*, 28-39.

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

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

Goose, S., & Moller, C. (1999). A 3D audio only interactive web browser: Using spatialization to convey hypermedia document structure. *Proceedings of the seventh ACM international conference on Multimedia (Part 1) (MULTIMEDIA'99)* (pp. 363-371), FL, USA.

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting,* San Francisco, USA.

Helle, S., Leplatre, G., Marila, J., & Laine, P. (2001). Menu sonification in a mobile phone -- a prototype study. *Proceedings of the International Conference on Auditory Display (ICAD2001)*, Espoo, Finland.

Jeon, M., Davison, B. K., Nees, M. A., Wilson, J., & Walker, B. N. (2009). Enhanced auditory menu cues improve dual task performance and are preferred with in-vehicle technologies. *Proceedings of the the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI09)* (pp. 91-98), Essen, Germany.

Jeon, M., Na, D., Ahn, J., & Hong, J. (2008). User segmentation & UI optimization through mobile phone log analysis. *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI2008)* (pp. 495-496), Amsterdam, Netherlands.

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

Jeon, M., Yalla, P., & Walker, B. N. (under review). Audiotry scrollbar and spindex (Speech Index) improve auditory menu acceptance and performance. *ACM Transactions on Accessible Computing*.

Klante, P. (2004). Auditory interaction objects for mobile applications. *Proceedings of the 7th International Conference on Work with Computing Systems (WWCS2004),* Kuala Lumpur, Malaysia.

Kramer, G. (1994). An introduction to auditory display. In G. Kramer (Ed.), *Auditory display: Sonification, audification, and auditory interfaces* (pp. 1-77). MA: Addison-Wesley.

Lee, J., & Spence, C. (2008a). Feeling what you hear: Task-irrelevant sounds modulate tactile perception delievered via a touch screen. *Journal on Multimodal User Interfaces, 2*(3-4), 1783-7677.

Lee, J., & Spence, C. (2008b). Spatiotemporal visuotactile interaction. *EuroHaptics 2008, LNCS 5024*, 826-831.

Leplatre, G., & Brewster, S. A. (2000). Designing non-speech sounds to support navigation in mobile phone menus. *Proceedings of the International Conference on Auditory Display (ICAD2000)* (pp. 190-199), GA, USA.

Marila, J. (2002). Experimental comparison of complex and simple sounds in menu and hierarchy sonification. *Proceedings of the International Conference on Auditory Display (ICAD2000)*, Kyoto, Japan.

Morley, S., Petrie, H., & McNally, P. (1998). Auditory navigation in hyperspace: Design and evaluation of a non-visual hypermedia system for blind users. *Proceedings of the Annual ACM Conference on Assistive Technologies (ASSETS98)*, Marina del Rey, CA, USA.

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

Mynatt, E., & Weber, G. (1994). Nonvisual presentation of graphical user interfaces: Contrasting two approaches. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI94)* (pp, 166-172), Boston, MA, USA.

Nees, M. A., & Walker, B. N. (2009). Auditory interfaces and sonification. In C. Stephanidis (Ed.), *The universal access handbook* (pp. 507-521). New York: Lawrence Erlbaum Associates.

Norman, D. A. (2004). *Emotional design*. New York: Basic Books.

Norman, D. A. (2007). *The design of future things*. New York: Basic Books.

Oh, J. W., Park, J. H., Jo, J. H., Lee, C., & Yun, M. H. (2007). Development of a kansei analysis system on the physical user interface. *Proceedings of the Korean Conference on Human Computer Interaction, Kangwon, Korea.*

Palladino, D., & Walker, B. N. (2007). Learning rates for auditory menus enhanced with spearcons versus earcons. *Proceedings of the International Conference on Auditory Display (ICAD2007)* (pp. 274-279), Montreal, Canada.

Palladino, D., & Walker, B. N. (2008a). Efficiency of spearcon-enhanced navigation of one dimensional electronic menus. *Proceedings of the International Conference on Auditory Display (ICAD2008)*, Paris, France.

Palladino, D., & Walker, B. N. (2008b). Navigation efficiency of two dimensional auditory menus using spearcon enhancements. *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society (HFES2008)* (pp. 1262-1266), NY, USA.

Pitts, M. J., Williams, M. A., Wellings, T., & Attridge, A. (2009). Assessing subjective response to haptic feedback in automotive touchscreens. *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI 09)* (pp. 11-18), Essesn, Germany.

Raman, T. V. (1997). *Auditory User Interfaces: Toward the Speaking Computer*. Boston: Kluwer Academic Publishers.

Russell, D. C., & 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. *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting (HFES 2009)* (pp. 980-984), San Antonio, TX.

Sanders, M., S. & McCormick, E. J. (1993). Chapter11: Controls and data entry devices. In M. S. Sanders & E. J. McCormick (Ed.), *Human factors in engineering and design* (pp. 334-382). New York:McGraw-Hill, Inc.

Sawhney, N., & Schmandt, C. (2000). Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments. *ACM Transactions on Computer-Human Interaction, 7*(3), 353-383.

Simms, C. (2008, June 23). HP TouchSmart IQ505a. CNET [On-line]. Available: http://www.zdnet.com.au/reviews/hardware/lifestyle/soa/HP-TouchSmart-IQ505a/0,2000065624,339290030,00.htm.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Vargas, M. L. M., & Anderson, S. (2003). Combining speech and earcons to assist menu navigation. *Proceedings of the 2003 International Conference on Auditory Display (ICAD2003)* (pp. 38-46), Bonston, MA, USA.

Walker, B. N., & Kogan, A. (2009). Spearcons enhance performance and preference for auditory menus on a mobile phone. In C. Stephanidis (Ed.), *Universal Access in HCI, Part II, HCII 2009, Lecture Notes in Computer Science 5615. Berlin: Springer-Verlag. (pp. 445-454)*.

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

Walker, B. N., Nance, A., & Lindsay, J. (2006). Spearcons: Speech-based earcons improve navigation performance in auditory menus. *Proceedings of the International Conference on Auditory Display (ICAD2006)* (pp. 95-98), London, UK.

Wilson, J., Walker, B. N., Lindsay, J., Cambias, C., & Dellaert, F. (2007). SWAN: System for wearable audio navigation. *Proceedings of the 11th International Symposium on Wearable Computers (ISWC2007)*, Boston, MA, USA.

Yalla, P., & Walker, B. N. (2008). Advanced auditory menus: Design and evaluation of auditory scrollbars. *Proceedings of the Annual ACM Conference on Assistive Technologies (ASSETS'08)*, Halifax, Nova Scotia, Canada.

Zhao, S., Dragicevic, P., Chignell, M., Balakrishnan, R., & Baudisch, P. (2007). earPod: Eyes-free menu selection using touch input and reactive audio feedback. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI07)* (pp. 1395-1404), San Jose, CA. USA.