Interacting with Computers Advance Access published February 6, 2013

doi:10.1093/iwc/iws021

Shared Input Multimodal Mobile Interfaces: Interaction Modality Effects on Menu Selection in Single-Task and Dual-Task Environments†

SHENGDONG ZHAO[1,*], DUNCAN P. BRUMBY[2], MARK CHIGNELL[3], DARIO SALVUCCI[4], AND SAHIL GOYAL[5]

[1]*Department of Computer Science, National University of Singapore, 13 Computing Drive, Computing 2, #01-04, Singapore 117417*
[2]*UCL Interaction Centre, University College London, Gower Street, London WC1E 6BT, UK*
[3]*Knowledge Media Design Institute (KMDI), University of Toronto, 27 King's College Circle, Toronto, Ont., Canada M5S 1A1*
[4]*Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104, USA*
[5]*National University of Singapore, 13 Computing Drive, Computing 2, #01-04, Singapore 117417*
**Corresponding author: zhaosd@comp.nus.edu.sg*

Audio and visual modalities are two common output channels in the user interfaces embedded in today's mobile devices. However, these user interfaces are typically centered on the visual modality as the primary output channel, with audio output serving a secondary role. This paper argues for an increased need for shared input multimodal user interfaces for mobile devices. A shared input multimodal interface can be operated independently using a specific output modality, leaving users to choose the preferred method of interaction in different scenarios. We evaluate the value of a shared input multimodal menu system both in a single-task desktop setting and in a dynamic dual-task setting, in which the user was required to interact with the shared input multimodal menu system while driving a simulated vehicle. Results indicate that users were faster at locating a target item in the menu when visual feedback was provided in the single-task desktop setting, but in the dual-task driving setting, visual output presented a significant source of visual distraction that interfered with driving performance. In contrast, auditory output mitigated some of the risk associated with menu selection while driving. A shared input multimodal interface allows users to take advantage of multiple feedback modalities properly, providing a better overall experience.

STUDY HIGHLIGHTS

- We propose a shared input multimodal user interfaces for mobile devices
- We tested it on a single task desktop and a dual task driving setting
- Time taken was shorter when visual feedback was provided in the single-task desktop
- In dual-task driving setting, visual output is a major source of distraction
- Auditory output mitigated some risk associated with menu selection while driving

Keywords: earPod; eyes-free; shared-input multimodal interfaces

Editorial Board Member: Eva Hornecker

Received 29 December 2011; Revised 11 October 2012; Accepted 13 November 2012

1. INTRODUCTION

In today's technology-rich world, devices are increasingly powerful and multi-functional. For instance, a single handheld

†This submission contains original work, which has not been published previously and it is not under consideration for publication elsewhere.

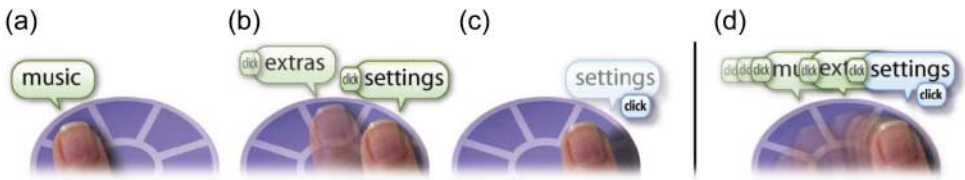

Figure 1. Using *earPod*. (a, b) Sliding the thumb on the circular touchpad allows discovery of menu items; (c) the desired item is selected by lifting the thumb; (d) faster finger motions cause partial playback of audio. Size of the touchpad has been exaggerated for illustration purposes.

device can act as a cell phone, music player, digital camera, GPS navigation system and personal digital assistant. With increased functionality to support day-to-day tasks, people increasingly turn to such devices in any number of different contexts and settings. An important distinction that can be made between usage scenarios is between those where the user is in a relatively isolated and static environment and those where the user is on the move.

Most interfaces today rely on the visual modality to present information to users, which works well in a relatively isolated and static environment when visual attention is available; however, for users on the move interacting with visual interfaces creates competition for limited visual resources. For example, interaction with an iPod while driving may be distracting and constitutes a potential safety hazard (Salvucci *et al.*, 2007). Multiple resource theory (Wickens, 2002) suggests that using auditory output for the secondary task may alleviate interference in a dual-task setting where the primary task is visually demanding.

One of the challenges of mobile device design is how to support the diverse range of usage contexts in which the device will likely be used. One solution to this problem is to offer the user a choice of different output modalities so that they might then choose the most appropriate interaction method for a specific context of use. For instance, calls can be made on a mobile phone both via voice commands and via pressing digits on a keyboard. This gives the user a choice of two distinct ways of interacting with the device depending on their situation and preference of use.

In this paper we build on this idea of providing users with a choice of interfaces. Instead of using two different (manual versus voice) interfaces, we propose to use two related interfaces with a shared input mechanism. The two interfaces differ only in their output modalities, resulting in a shared input multimodal interface that can be independently operated using either audio or visual feedback. We apply this new interface design approach to the design of mobile menus by extending a touch-based auditory menu technique called *earPod* into an integrated interface that has both audio and visual interfaces (Zhao *et al.*, 2007) (Figs 1 and 2).

The original *earPod* technique is designed for an auditory device controlled by a circular touchpad whose output is experienced via a headset (Fig. 2) as is found, for example, on an Apple iPod. Figure 3 shows how the touchpad area is

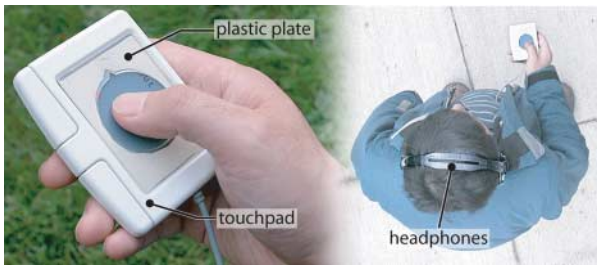

Figure 2. Our *earPod* prototype uses a headset and a modified touchpad.

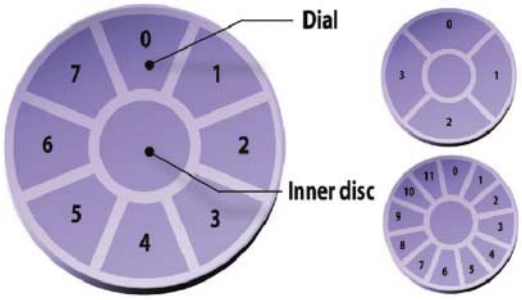

Figure 3. The functional areas of *earPod*'s touchpad. Up to 12 menu items can be mapped to the track. The inner disc is used for canceling a selection.

functionally divided into an inner disc and an outer track called the dial. The dial is divided evenly into sectors, similar to a Pie (Callahan *et al.*, 1988) or marking menu (Kurtenbach, 1993; Zhao *et al.*, 2006, 2004). How the *earPod* technique is used for menu selection is illustrated in Fig. 1. When a user touches the dial, the audio menu responds by saying the name of the menu item located under the finger (Fig. 1a). Users may continue to press their finger on the touch surface or initiate an exploratory gesture on the dial (Fig. 1b). Whenever the finger enters a new sector on the dial, playback of the previous menu item is aborted. In addition to speech playback of menu items, we use non-speech audio to provide rapid navigational cues to the user. Boundary crossing is reinforced by a click sound, and then the new menu item is played. Once a desired menu item has been reached, users select it by lifting the touching finger, which is confirmed by a 'camera-shutter' sound (Fig. 1c). Users can abort item selections by releasing the operating finger on the

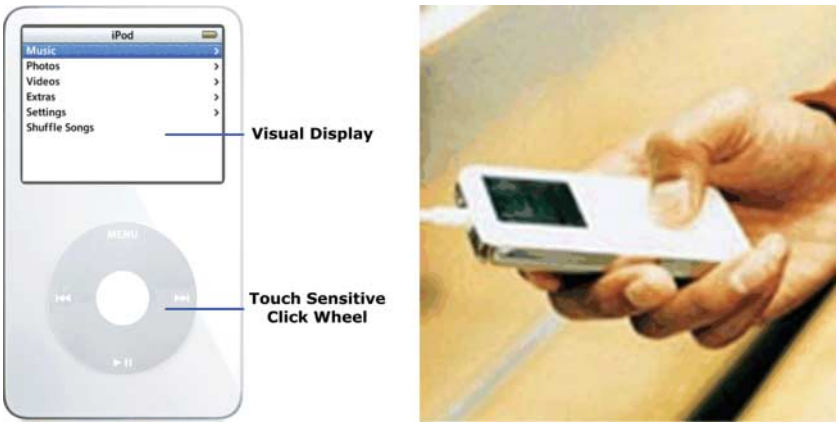

Figure 4. The iPod visual menu (left) and its interaction technique (right).

center of the touchpad. If a selected item has submenus, users repeat the above process to drill down the hierarchy, until they reach a desired leaf item. Users can skip items rapidly using fast dialing gestures (Fig. 1d). All speech sounds used in *earPod* are human voices recorded in CD quality (16 bits, 44 KhZ) using professional equipment.

A previous study (Zhao *et al.*, 2007) has shown *earPod* to be an effective eyes-free menu selection technique, with comparable performance with iPod style linear menus (Fig. 4) for menu selection tasks in the single-task desktop context. This suggests that *earPod* is a compelling technique for eyes-free usage. However, today's mobile devices can be used in many different contexts and settings (e.g. in a static environment or in changing environments). A technique optimized for a particular scenario may work poorly for other scenarios, reducing its overall usefulness. In order to design menu techniques that can work well for different scenarios, we expand the design space of *earPod* (Zhao *et al.*, 2007) into a family of menu techniques that differ in modality of feedback and menu style. We then investigate how different points in this design space (modality: visual, audio and audio–visual feedbacks; menu layout style: linear and radial) affect user performance and preference for menu selection in single- and dual-task environments. Finally, on the basis of these results, we draw out design recommendations and guidelines for the design of shared input multimodal mobile interfaces that are suitable for both single- and dual-task contexts.

This paper attempts to address the following research questions related to the design of a shared input multimodal mobile menu.

1. In what situation should each interface in a shared input multimodal menu be used?
2. What is the benefit (if any) of allowing users to choose which modality to use for mobile input?
3. Should the two modality of feedback exist simultaneously, or be provided separately?
4. How do different menu styles affect the design of shared input multimodal interfaces?

To answer these questions, we performed two rounds of experiments, which evaluated the alternative output modalities under the single-task desktop and dual-task driving conditions, respectively. The results, along with design recommendations for both desktop and driving scenarios and general discussions of interface design for mobile and ubiquitous computing, are presented and discussed in later sections of this paper.

2. RELATED WORK

The research literature on hierarchical menu layout and on output modality in hierarchical menus is reviewed in the following sections. We begin by reviewing previous work that has explored the design space for difference menu layout conventions and that has examined how various output modalities can be used to support user interactions with the system.

2.1. Menu layout

Many menus have been developed for diverse applications and platforms. They can be classified under different systems but this paper focuses on two contrasting menu styles, namely, linear style menus and radial style menus. Linear style menus lay out their items linearly where the cost (effort) to access each item is different (Fig. 5, right); radial style menus lay out the items radially in a polar coordinate system where there is a constant distance from each item to the center of the circle in which the menu is embedded (Fig. 5, left). Items in linear menus are also relative to each other in the sense that they have to be traversed sequentially in order to reach the target item. In contrast, items in radial menus have absolute locations in the sense that, with sufficient skill and knowledge, users can go directly to the target item without having to traverse through other items on the way.

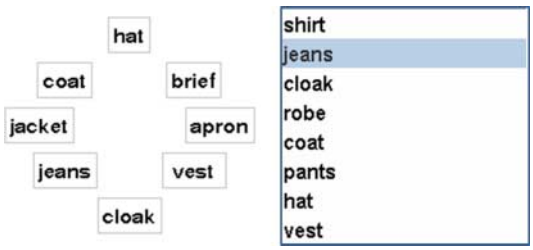

Figure 5. Screenshots of the visual radial interface (left) and visual linear interface (right).

Radial menus also have the advantage of allowing items to be placed in a meaningful location, for example, 'Open' and 'Close' can be placed in two opposite directions and 'Previous' and 'Next' can be placed in a way that reflects the semantics of those words. In this paper, the linear versus radial terminology will be used throughout for distinguishing between menu types. However, it should be kept in mind that linear menus are also relative and that radial menus are also absolute.

Callahan *et al.* (1988) summarized the strengths and weaknesses of each menu style. Linear style menus are easier for arranging items and they are more flexible in the number of choices in a single menu/submenu and are more familiar to users (Sears and Shneiderman, 1994). However, because items are arranged sequentially, access time to each item is uneven: depending on the initial placement of the cursor, items closer to the cursor are quicker to select than items further away. Radial style menus, on the other hand, lay out items at equal-distance from the center and require constant access time and have better performance than linear style menus (Callahan *et al.*, 1988; Kurtenbach and Buxton, 1994). However, placing labels in a circular layout requires more space (Fig. 5, left) and the number of items allowed in one circular array is typically limited to no more than 12 items because of performance concerns (Kurtenbach and Buxton, 1993; Zhao and Balakrishnan, 2004; Zhao *et al.*, 2006).

2.2. Output modality and menu design

Interfaces typically require some form of output feedback to guide and inform users. Visual output is very common in current graphic user interface. Haptic output is another possibility. Haptic output is by nature a 'private' display and can communicate information even in noisy environments (Wagner *et al.*, 2004; Luk *et al.*, 2006). However, most users are not familiar with haptic-based languages such as the Braille alphabet, making it difficult to use haptics to communicate rich amount of information effectively.

Auditory output, on the other hand, may utilize both speech and non-speech audio, allowing communication of information containing rich semantics to users with less learning. Sound can travel in space and is omnidirectional, making it particularly suitable for delivering important messages like alarms and alerts.

There is a large body of work on audio-based icons, for example, using much shorter non-speech audio segments to represent equivalent messages using speech, e.g. (Brewster *et al.*, 2003a). This is similar to the use of space-efficient icons to represent labels in graphical interfaces, an approach first introduced by Gaver and Smith (1991) and Gaver (1989), who designed an auditory system called Sonic Finder for the Apple Macintosh computer using everyday sound to represent objects, tasks and events. These symbolic sounds, which are analogous to what they represent, are called auditory icons.

Motivated by the needs of blind users, e.g. Mynatt (1995) discussed methods for creating auditory equivalents of desktop user-interface elements. Other studies have examined the use of auditory interfaces for specific tasks. Arons (1997) described the SpeechSkimmer system for efficiently browsing through recorded speech. Minoru and Schmandt (1997) and Roy and Schmandt (1996) also described systems for navigating through audio information. Other systems that demonstrated interactive methods for interacting with audio were described by Schmandt 1998 and Schmandt *et al.* (2004). Sawhney and Schmandt (2000) described the Nomadic Radio system for using an audio interface to access communication and information services while on the move. Stifelman *et al.* (1993, 2001) examined the problem of note taking and annotation using an auditory interface.

However, relatively fewer studies have focused on the interaction design of auditory menus. One notable exception is Brewster (1998)'s work, which investigated the effectiveness of non-speech audio navigational cues in voice menus. Brewster (1998) found that distinctive non-speech audio elements (*earcons*) were a powerful method of communicating hierarchy information, as it is difficult to match auditory icons with suitable iconic sounds for events in an interface. Not every auditory icon will correspond to a sound-producing event which can be easily mapped to corresponding interface actions or states in the real world. However, in this work, Brewster used earcons to indicate the current tool and the tool changes, and not as a form of audio feedback for an imminent tool selection by the user.

Dingler *et al.* (2008) investigated the learnability of sonification techniques such as auditory icons, earcons, speech and spearcons for representing common environmental features. Spearcons are speech stimuli that have been greatly sped up (Walker *et al.*, 2006). They found that speech and spearcons were easily learnable compared with earcons and auditory icons. In fact, earcons were much more difficult to learn than speech. With this study in mind, we used speech as the audio feedback for *earPod*.

The emergence of mobile computing has inspired researchers to rethink the interaction model of audio command selection. Pirhonen *et al.* (2002) investigated the use of simple gestures and audio-only feedback to control music playback in mobile

devices. Brewster *et al.* (2003b) also investigated the use of head gestures to operate auditory menus. Both techniques have demonstrated effectiveness in the mobile environment. However, they have only been investigated with a very limited number of commands. For example, the head gesture menu that Brewster *et al.* (2003b) created used only four options, which is insufficient for the wide range of functionality that exists in many devices.

One prior technique that is similar to *earPod* is Rinnot's Sonic Texting (Rinnot, 2005), which is a mobile text input method that leverages touch input and auditory output. It uses a two-level radial menu layout to organize the alphabets, and uses a keybong joystick for input. Although it is similar to *earPod* in terms of the use of a radial menu layout and auditory feedback, the *earPod* technique significantly differs from it in the following aspects. Sonic texting uses joystick instead of touchpad for gestural input. The type of gestures supported for selection is quite different. Sonic texting uses the back-and-forth movement of the joystick to select text, while in *earPod*, finger gliding, tapping and lifting are the primary gestures for interaction. Sonic texting is designed for text input, which is optimized for a specific set of alphabets while *earPod* is a general menuing technique that can support multiple hierarchies of menu items. Sonic texting has not been formally evaluated, making it difficult to access the viability of sonic texting as a mobile text input method. As mentioned by the author, sonic texting's goal is not to maximize word-per-minute efficiency, but to create an engaging audio-tactile experience. In the informal evaluation session it was used, it is regarded by many users as a game or musical instrument rather than a mobile textual input method.

Nigay and Coutaz (1993) have proposed a design space for multimodal systems. Their taxonomy defined three dimensions of multimodal system: level of abstraction, concurrency and fusion. According to their definition, the difference between multimodal and multimedia lies in whether or not the system can interpret the meaning of the different channel of modalities. However, their analysis mostly focused on input modalities. No examples were given on how to apply this principle to systems with multiple output modalities. For example, assuming that there is a system that plays a movie clip to the audience and within the clip, there are both audio and video playbacks. This system, according to our conventional definition, should be classified as a multimedia system. However, if this system uses text-to-speech instead of raw recordings to play the audio dialogues, according to Nigay and Coutaz (1993)'s taxonomy, the system knows the meaning of the output, which will be classified as a multimodal system. However, this can be a bit of misleading since the text-to-speech is not used to provide any system feedbacks about the user input, but to provide content to users.

We would like to amend the definition of Nigay and Coutaz's (1993) taxonomy to include systems with multiple channel of output. If the multiple channel of output is to provide system feedback about user input, then it is considered multimodal; otherwise, it is considered multimedia. This is regardless whether or not the system understands the meaning of the output or not.

It is not a simple task to decide which modality is the best as each has its own advantages and disadvantages. Salmen *et al.* (1999), who weighed the pros and cons of using audio, visual and dual modality in a driving scenario, found that audio modality is good as drivers do not have to refer to the screen and can focus on the road. However, should the list of the audio instructions be too long, drivers may have difficulties recalling the full set of instructions as oral presentation typically takes three times longer to process compared with reading. Thus, visual modality may be useful in instances whereby drivers might want to get information faster but again, there are many issues with this, such as whether scrolling down a page or paging a text is better. Finally, the combination of visual and audio modality may seem like the perfect solution but, unfortunately, if used simultaneously, there may be some issues such as too many different audio and visual commands to remember which may lead to frequent mix up (Salmen *et al.*, 1999).

Within our 'shared input exclusive multimodal system', there are two single modal interfaces, which are designed to function independently for different scenarios. However, the two types of interfaces share the same input mechanism as well as the same mental model. This will train the user with both interfaces because they differ only in the output modalities.

In multimodal user interface design, input may be provided using multiple modalities. For instance, phone numbers may be input by dialing a cellular phone or via voice commands or via pressing digits on a keypad. Different feedback modalities can co-exist where different modalities may be used depending on the context of use. For instance, previous studies have examined and compared unimodal, bimodal and trimodal feedback conditions (Akamatsu *et al.*, 1995; Vitense *et al.*, 2002). Jacko *et al.* (2003) examined multimodal feedback (for persons who are older and possess either normal or impaired vision) in drag and drop tasks relating to daily computer use. In addition, the game industry has demonstrated successful union of three types of feedback (audio, haptic and visual) that can provide players with an 'immersive' experience in a simulation game (Jacko *et al.*, 2003).

Sodnik *et al.* (2008) compared the use of auditory versus visual interfaces for interaction with a mobile device while driving. The proposed auditory interfaces consisted of spatialized auditory cues for menu selection. Though their results indicated that the task completion rate was same for both audio and visual, they found that the driving performance was better and the perceived cognitive load was lower while interacting using auditory interfaces. Not surprisingly, users were distracted by the visual interface while driving and preferred the audio interface. Pfleging *et al.* (2011) present a prototype to combine speech and multi-touch gestures for multimodal input in an automotive environment.

Table 1. The 3 × 2 design space of modality versus menu style for menu interfaces.

	Menu style	
Modality	Radial	Linear
Audio	Audio radial (*earPod*)	Audio linear
Visual	Visual radial	Visual linear (iPod)
Dual	Dual radial	Dual linear

There are many scenarios in which a user might desire or prefer eyes-free interaction (Yi *et al.*, 2012). Apart from contexts where eyes-free interaction may be less demanding, Yi *et al.* (2012) noted that users' are willing to use eyes free as a form of social acceptance, or even a form of self-expression.

The purpose of the present study was to compare the effectiveness of alternative modalities of audio, visual and audio–visual feedback for menu selection tasks in single- and dual-task scenarios.

3. DESIGN SPACE OF MENU SELECTION

If we consider modality and menu style as two dimensions in a design space, a simple analysis (Table 1) reveals that there are a number of design alternatives. If we label an interface first by its primary feedback modality, followed by its menu style, the popular iPod will fit within the 'visual linear' category, whereas the *earPod* (Zhao *et al.*, 2007) will reside within the 'audio radial' cell. The two alternative interfaces here are the 'audio linear' and 'visual radial'. Additionally, since audio and visual feedback can co-exist and are not mutually exclusive (unlike menu style), there is a third possible choice in the modality dimension, which is the audio–visual or dual modality. This gives us a 3 × 2 matrix of six design possibilities. These alternative designs cover a variety of interesting properties and thus warrant further investigation in a multitasking context.

3.1. Visual linear (iPod) and audio radial (*earPod*)

The two interfaces (*earPod* and iPod-like menu) differ in two aspects, namely the modality of feedback and the menu style for presenting and navigating the menu items. For modality, the iPod-like menu primarily relies on a visual display to present the menu options and navigational cues, while *earPod* does these entirely using audio. In terms of menu style, the iPod uses linear menus (Fig. 4) where items are placed linear to each other and there is no one-to-one mapping between specific input areas to menu items; *earPod*, on the other hand, adopts a radial menu layout where each menu item is directly mapped to a physical location on the touchpad (Fig. 5, left shows an example of radial layout for the visual interface) and allows expert users to access any items in the list in constant time.

3.2. Audio linear

In Table 1, audio linear is the cell next to iPod-like visual menu. It provides spoken-word auditory feedback to the user as they scroll up or down a menu list. In some respects, the interface is similar to that used in the popular Apple iPod digital music player, except that in the absence of a visual display, it provides auditory feedback on users' actions. Moreover, one could imagine that such an audio linear interface could easily be integrated with the existing iPod interface, which is currently available through an open source solution from rockbox.org [Rockbox]. However, linear menus are often slower than radial ones (Callahan *et al.*, 1988); therefore, it could be even slower to operate than the visual linear interface. It will be informative to systematically evaluate it against the other cells in the design space.

3.3. Visual radial

As discussed above, this interface has a radial input area that supports a radial menu layout; that is, where specific spatial regions on the input device have a one-to-one mapping with items in the menu. Figure 5(left) shows our design of the visual radial interface. Although the interaction method differs, its appearance looks similar to that of marking menu (Kurtenbach, 1993; Zhao *et al.*, 2007). Note that this is in contrast to the linear menu layout, where the input device supports a vertical scroll of a focus point through the menu (see, Fig. 5, right). We might speculate that the performance advantages of the *earPod* interface discussed earlier (Zhao *et al.*, 2007), may, in part, be due to the radial menu layout used. We aim to more carefully evaluate the potential performance benefit of using a radial menu layout for selecting items from reasonably sized static menus.

3.4. Dual (audio–visual) linear and dual (audio–visual) radial

By providing both audio and visual feedback simultaneously, the interface may possibly combine the best of both worlds—namely, they can be operated using either modality, thereby giving users a choice of which modality to attend to in different circumstances. For example, if the device is operated inside one's pocket, the visual feedback can be ignored. If the device is in a noisy environment, the visual feedback prevails and the audio feedback becomes less useful. Since both channels of feedback use the same menu style, the training received in either modality can be used in the other. However, simultaneously providing both modalities might waste resources (such as battery power), and one source of feedback has the potential to be distracting if one modality of feedback is preferred: for example, a user who prefers visual feedback could be annoyed by the simultaneous audio feedback.

Modality (audio, visual and dual) and menu style (linear versus radial) are two dimensions for describing an interesting

design space of menu selection. To disentangle the individual effects of the two design dimensions and to further explore the properties of the other four design alternatives relative to the iPod and *earPod* interfaces in the baseline desktop conditions, we decided on the following 3 × 2 experimental design that employed all six interfaces from Table 1.

4. EXPERIMENT 1—SINGLE-TASK DESKTOP SETTING ENVIRONMENT

The aim of Experiment 1 was to systematically evaluate the design space for menu selection tasks along the dimensions outlined above, namely, feedback modality and layout style. The study had participants attempt to locate and select a pre-defined item from an eight-item menu as quickly and accurately as possible. In terms of user performance in selection time, based on earlier work, we would expect that the radial layout would support faster target selections because it allows direct access to menu items compared to the linear layout (Callahan *et al.*, 1988; Kurtenbach and Buxton, 1994). We also expected the visual output modality to be faster because auditory information is regarded as serial and temporal in nature while visual information can be scanned relatively quickly (Zhao *et al.*, 2007).

4.1. Method

4.1.1. Participants

Twelve right-handed participants (three females) ranging in age from 18 to 29 years (mean 22), recruited from the University of Toronto volunteered for the experiment.

4.1.2. Materials

A menu selection task was used that required participants to select a target item from a menu. Each menu contained eight items, all of which belonging to the same natural category. Materials were developed from examples of natural categories taken from KidsClick! (2010; http://sunsite.berkeley.edu/KidsClick!/) and Wikipedia (2010; http://en.wikipedia.org/wiki/Portal:Contents/Categories]. Across the set of materials, there were eight categories, describing types of Clothing, Fish, Instrument, Job, Animal, Color, Country and Fruit. For each of these category types, there were eight items. For instance, Carp, Cod, Eel, Haddock, Pollock, Redfish, Salmon and Sardine were used as types of Fish. Each item was a single word and no word appeared more than once in the database.

The experimental software ran on a Compaq Presario V2000 laptop with 2 GB of RAM running Microsoft Windows XP. Input was controlled by a Cirque EasyCat USB external touchpad. The touchpad was made circular by placing thin plastic overlay to the touchpad area. Both radial and linear menu layout styles were implemented on the circular touchpad. With the linear design, the movement of the thumb around the circular touchpad allows the user to scroll through the list of items in the menu (i.e. much like the interaction technique used for the Apple iPod). In contrast, the radial design subdivides the circular touchpad into discrete regions. In this way, items in the menu are located at particular locations. In both cases, the participant selected the currently highlighted item in the menu by lifting their thumb off the touchpad. A short click sound provided feedback that a selection had been made.

In addition to different layout styles, different output modalities could be used as the participant explored the menu (audio or visual). For visual output, menu items were arranged in a vertical list, one item per line. All text was presented on a 19-in. LCD monitor in font Helvetica, Bold and size 16. A colored box was used to highlight the currently selected item. For the audio output, auditory information was generated using a real-time simulation library. When the user scrolled over an item in the menu, the items label was spoken in a female human voice. Each audio clip took ∼1 s to playback. The audio recording was interruptible, such that if the user quickly scrolled to the next item, the playback of the first would terminate and the next item would be outputted. Audio output was presented to the participant through standard stereo headphones. For audio–visual feedback, both of these output streams were presented simultaneously to the user.

4.1.3. Design

A 2 × 3 (layout style × output modality) within-subjects design was used to systematically explore a design space for the menu selection task. Menu items were presented in either a radial or linear layout style. The output was given as audio-only, visual-only or audio–visual. The order that each output modality was used was counterbalanced between participants, while the ordering of menu layout style was randomized within each modality. The main dependent variables of interest were the time taken to select a target item from the menu and the number of errors that were made. Subjective feedback on participants' preferences for each point in the design space was also gathered during a post-experiment interview.

4.1.4. Procedure

Participants were informed that they would be required to perform a menu selection task using different design alternatives. They were told that the whole experiment would take about an hour to complete and that they were free to withdraw at any time if they wished without loss of credit. After receiving these instructions, participants were instructed to put on the headphones and to hold the touchpad with their right hands leaving the thumb off the touchpad.

Participants completed a series of menu selections with each interface type (i.e. for each combination of different layout style and output modality). Before each condition, participants received eight practice trials (one block) with a particular interface type so as to familiarize themselves with it. This was to ensure that any prior experience that users had with a certain menu type would not affect the findings of the experiment.

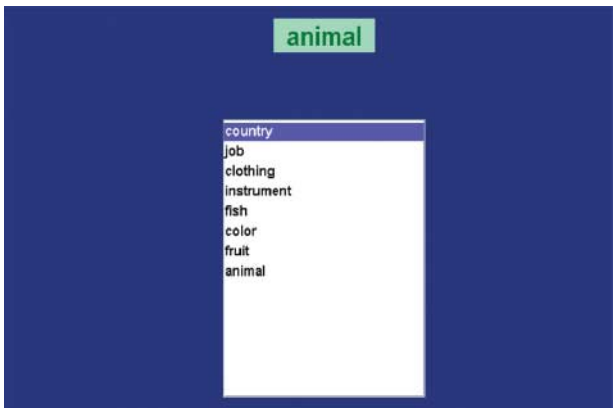

Figure 6. Visual stimulus at top of the screen.

Once familiar with the interface, each participant completed 96 trials, consisting of 12 blocks of trials for each of the 8 possible target positions. That is, in total, each participant completed 576 experimental trials (2 layout style × 3 output modality × 8 target items × 12 blocks) along with 48 practice trials.

For each trial, the to-be-selected item was presented in the center of the monitor (e.g. 'Animal'). Once they had encoded the to-be-selected item, the participant could start the trial menu by pressing the touchpad (i.e. making a selection gesture). The participant then searched the menu for the target. Dependent on condition, participants either received audio-only, visual-only or audio–visual output as they searched. In the audio conditions, participants heard the spoken names of each traversed menu item through their headphones. In the visual condition, menu items were displayed on the screen (Fig. 6). Participants selected the currently highlighted item by lifting their thumb off the touchpad. If an incorrect selection was made, then participants were notified by a visual prompt that informed them of their error. Participants were required to make another selection from the menu and did not progress to the next trial until the target was selected. The trial ended when the participant selected the target and participants were instructed to locate the target as quickly and as accurately as possible. After each trial, a visual message in the center of the screen instructed participants to press the spacebar to proceed to the next trial. Participants were allowed to take breaks between trials and breaks were enforced between each interface condition. After completing all of the trials, participants were asked to answer a set of questions about their preferences for each interface design.

4.2. Results

For each trial, we consider data from when the menu first appeared to when the participant selected an item. For statistical analysis, a 2 × 3 × 12 repeated-measures ANOVA was used, and effects were judged significant if they reached a 0.05 significance level.

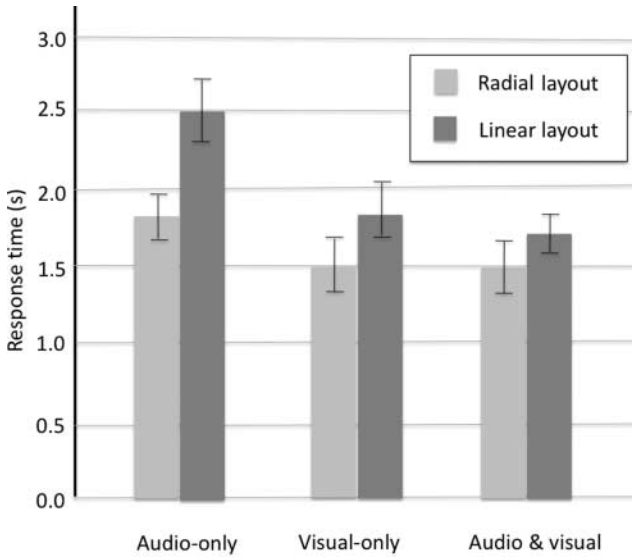

Figure 7. Response time for all six interfaces, sorted by modalities.

4.2.1. Errors

Participants made very few selection errors, occurring on only 5% of trials (SD = 0.005). There was no reliable difference in error-rate regardless of the menu layout style (radial 5.3% versus linear 4.7%) or of the output modality used (audio 4.8%, visual 5.4% and audio–visual 4.9%). Neither was there any evidence to suggest a change in error-rate over consecutive blocks of trials. Indeed, statistical analysis showed all effects to be non-significant (all F's < 1.03). We next consider the response time data.

4.2.2. Response time

For response time data, trials in which an incorrect item was selected on the first selection were removed—thus, we consider only response time for correct selections. Figure 7 shows the mean response time for each of the experimental conditions. Participants were significantly slower at selecting target items when the menu used a linear layout ($M = 2.33$ s) rather than a radial layout ($M = 1.58$ s) ($F_{1,11} = 3.06$, $P < 0.001$). Participants were also significantly slower when they received only audio feedback ($M = 2.12$ s) compared with when they received visual feedback, either in the visual-only condition ($M = 1.7$ s) or the audio–visual condition ($M = 1.83$ s) ($F_{2,22} = 8.69$, $P < 0.001$). As can be seen in Fig. 7, the difference in response time between the radial and linear layouts increased when audio feedback was used. Indeed, statistical analysis shows that there was a significant two-way interaction between output modality and layout style ($F_{2,22} = 8.69$, $P < 0.01$). This suggests that using a radial layout only carries performance benefits when the participant has to rely only on audio feedback.

4.2.3. *Observations & subjective preference*

Feedback from the post-experimental interviews indicated that the visual radial and audio–visual radial interfaces were rated most favorably by participants, while the audio linear interface had the lowest user satisfaction score. Almost all of the participants (10 of 12) reported that they preferred visual feedback to audio feedback. There were however two participants who said that they preferred receiving audio feedback to visual feedback. It is interesting to note that in terms of performance metrics, these participants were nonetheless faster at completing the selection task when they received visual feedback. This suggests that these two participants preference for audio feedback might stem from the novelty of using this interaction technique.

4.3. Discussion

The aim of the Experiment 1 was to evaluate different points in design space for devices that support menu selection. Results show that visual feedback modality affords faster selections presumably because audio takes time to listen, whereas for visual, we can quickly search visually for the to-be-selected item. In terms of layout, radial confers some benefit to a tradition linear; but these benefits are for audio—presumably because the participant has learnt. These results are consistent with Yin and Zhai (2006) and Callahan *et al.*, (1988) and suggest that visual feedback is optimal for supporting menu selection in single-task desktop environment. We next consider how different design alternatives might be better for alternative contexts of use. In particular, we consider how each of the above design alternatives fair when the user is engaged in some ongoing safety-critical primary task.

5. EXPERIMENT2—DUAL-TASK DRIVING SETTING ENVIRONMENT

Today's mobile devices are often used while a person is performing another task. In particular, the driver of a car typically drives the vehicle while performing other tasks or dealing with various distractions and this dual-tasking environment has attracted considerable research attention.

According to a survey of American drivers [GMAC2006], menus on mobile devices (specifically the iPod) are commonly used while driving, especially by young drivers ages 18–24. Although widely prohibited in many countries (legislation has been introduced in many countries including Australia, France, Germany, Japan, Russia, Singapore and the UK), people continue to use nomadic technology devices, such as cell phones and digital music players, while driving. For instance, compliance with the UK ban has slipped from 90% from its introduction in 2003 to around 75% in 2007.

Previous work on driver distraction resulting from cell phone use shows that it competes for limited visual attention resources, thus harming performance (e.g. Alm and Nilsson, 1994; McKnight and McKnight, 1993; Brumby *et al.*, 2009). Other research suggests that cognitive load alone, separate from perceptual/motor load, is sufficient to produce distraction effects. For instance, Strayer and Johnson (2001) and Strayer *et al.* (2003) in a series of studies indicated that the cognitive act of generating a word is sufficient to cause noticeable distraction effects. It is unclear then whether designing a mobile device so that it does not place additional demands on visual attention resources would mitigate the harmful effects of distraction. The increased cognitive load of interacting, even with an eyes-free device such as the *earPod*, may be sufficient to result in adverse effects for driving performance.

Given that it is difficult to make people stop engaging in secondary tasks while driving, there may be substantial value in directing efforts to better designing mobile devices to make their use by the driver of a car less egregious. A user-centered design approach that is sensitive to the environmental constraints imposed by using a mobile device in the context of an on-going dynamic task.

Experiment 1 provided empirical results for various menu interaction techniques under the desktop setting. The results of this first study suggest that visual output modality works well for supporting menu selection in static single-task settings. However, it is an open empirical question whether visual interfaces also offer performance gains when the user is concurrently engaged in an ongoing dynamic task, such as driving a car. In particular, because visual interfaces demand visual attention, we might assume that this might lead to greater driver distraction than using audio. In contrast then, we might expect that audio output may confer benefits in this dual-task setting. In the next section of this paper, we describe an experiment that is designed to address this question.

5.1. Method

5.1.1. *Participants*

Another 12 participants (1 female) ranging in age from 20 to 35 years (mean 27), recruited within the university community, volunteered for the experiment.

5.1.2. *Materials*

The driving experiment was conducted using a desktop driving simulator. The simulation environment, coded in Java with OpenGL graphics, incorporates a three-lane highway with the driver's vehicle in the center lane, as shown in Fig. 8. The highway includes alternating straight segments and curved segments with varying curvatures, all of which can be driven at normal highway speeds. A second automated vehicle, visible in the rear-view mirror, follows behind the driver's car at a distance of roughly 50 feet (15 m) so the driver to keep at an adequate speed and distance between the lead car and the car behind. Construction cones are placed on each side of the driver's lane to motivate as accurate lane keeping as possible. Previous versions

Figure 8. Driving simulation environment.

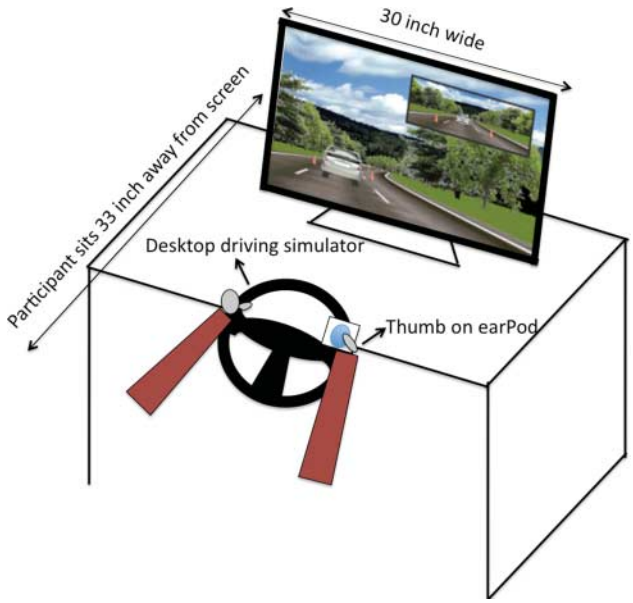

Figure 9. Set-up of experiment (not drawn to scale).

of a very similar environment have been used successfully to study various aspects of driver behavior (e.g. Salvucci, 2001, 2005; Salvucci *et al.*, 2007].

The hardware setup comprised a desktop computer controlled by a Logitech MOMO® steering wheel with force feedback. The simulation was run on an Apple desktop computer with an Intel Xeon CPU running at 2.00 GHz, 2 GB RAM and an NVIDIA GeForce 7300 GT graphics card. The environment was displayed on a 30 in. (69 cm) monitor at a distance of roughly 33 in. (85 cm) from the driver. The earPod was held in the dominant hand of the participant. For added realism, a soundtrack of real driving noise was run on continuous loop during the driving portions of the study. A rough set up of the experiment is shown in Fig. 9.

The experiment was divided into two parts. The first part (desktop condition) replicated Experiment 1 except it was much shorter. This setting allowed users to get familiar with the menu and the interaction techniques before they moved to the second and arguably more difficult part: driving and menu selection at the same time (driving condition). The experiment was designed to simulate a realistic usage scenario.

For the desktop condition, the setup was exactly the same as for Experiment 1 except the following difference.

In Experiment 2, both audio and visual stimuli were provided simultaneously and this allowed users to pick their preferred stimuli for different interfaces. Because driving is very different from desktop interaction, visual stimuli could be a possible source of distraction, thus complicating the interpretation of results. However, only using audio stimuli would not permit analysis of modality effects. To address these issues, we decided to provide both types of stimuli and allow users to pick the one that they would attend to during the experiment. This allowed us to find out which stimuli they actually used during the experiment. To allow better comparison between the two settings, we also used dual stimuli for the desktop condition.

The dual-task simulated driving condition is the focus of this experiment, but the desktop condition is also essential because it provides the necessary training for users to get familiar with techniques. This closely simulates real-world scenarios where users typically already have some experience with their devices before using them inside vehicles.

5.1.3. Design

A within-participants design was used. The exact design is summarized below. Desktop condition: 12 participants × 6 techniques × 8 items of 1 menu configurations: (Condition 8) × 5 blocks (4 blocks + 1 practice block for desktop conditions) + driving condition: 12 participants × 6 techniques × 8 consecutive selection of 1 menu item: (Condition 8) × 2 blocks (1 block + 1 practice block for driving conditions) = 4032 menu selections in total (2880 + 1152).

5.1.4. Instructions

During the desktop condition, the instructions remain the same as Experiment 1, where the participants were asked to complete the menu selection as quickly and accurately as possible. During the driving condition, for each trial, the participants were asked to complete the menu selection task as quickly and as accurately as possible while following an automated lead car that runs at a constant speed of 65 miles/h (∼105 km/h) and to maintain a reasonable, realistic following distance.

5.1.5. Procedure

After completing the desktop trials, participants were asked to answer a set of questions regarding their experience for the desktop conditions. They then moved to the driving simulator

and completed the menu selection tasks while driving. At least 10 s elapsed between the end of one menu-selection trial and the start of the next trial; this time allowed participants to perform any necessary corrective steering after each trial and re-center the vehicle to a normal driving state(note that this constraint reduced the number of trials possible in the driving context, but was absolutely necessary to maintain the integrity of the driver performance data.). The participants were allowed to take breaks between trials. Breaks were enforced after a maximum of 15 min of driving to avoid fatigue. Before each of the desktop and driving conditions, participants received eight practice trials (one block) for that particular interface. Each participant performed the entire experiment in one sitting which took ∼90 min (the desktop condition typically finished within 20 min and the driving condition typically lasted 40 min, with the extra time being used for questionnaires and breaks). After completion of the driving session, the same set of questions with the desktop condition was asked again regarding user experience during the trials.

5.2. Results

Both accuracy and response time results for the desktop setting were consistent with the ones from Experiment 1. Actual numbers varied, but no change was found concerning significance of effects.

5.2.1. *Observations & subjective preference*

Although this experiment differed little from Experiment 1, a set of additional questions in the post experimental questionnaire allowed us to gain more insights into users' experience. Since we use both visual and audio stimuli in this experiment, users were asked 'Which stimuli did you attend to during the experiment?' The answers were consistently visual (11 of 12 subjects) and only one subject said both. For the question 'Which feedback modality did you use under the dual-modality conditions?', the answers were again consistently 'visual' or 'primary visual'. For the question 'If you only used one type of feedback or primarily used only one type of feedback, did you found the other kind of feedback (audio or visual) distracting?', 3/12 users answered 'Yes, I found the audio feedback a bit distracting', while most subjects (9/12) said 'No'. Based on this feedback, it is clear that the visual modality is preferred under the single task desktop environment.

5.2.2. *Accuracy*

There were no significant differences in accuracy for either modality (audio 88.9%, visual 88.3% and dual 88.8%) or menu style (radial 88.3% and linear 89.0%). This is consistent with the findings from the desktop settings.

5.2.3. *Selection time*

Tests in the driving setting showed some unexpected findings. There was no significant main effect of modality on response time while driving. This is somewhat surprising since response time for audio was significantly slower than for visual in the desktop conditions. However, there was still a significant main effect of menu style, ($F_{1,11} = 32.86$, $P < 0.001$). Radial (3.34 s) was significantly faster than linear (4.12 s), which is also consistent with the findings from the desktop conditions. The average selection time for the six interfaces was: audio radial (3.27 s), audio linear (4.09 s), audio visual radial (3.53 s), audio visual linear (4.15 s), visual radial (3.27 s) and visual linear (4.10 s).

5.2.4. *Lateral velocity*

In testing interaction in the driving context, arguably the most important aspect of this interaction is the effect on driver performance. One common way to measure performance involves analysis of the vehicle's lateral (side-to-side) velocity as an indicator of vehicle stability. We computed the average lateral velocity over a time window that included both the interaction with the device and a period of 5 s after the completion of the interaction; this latter period accounts for vehicle 'correction' that typically takes place after distraction—during which the driver corrects the lateral position of the vehicle—which is best attributed to the immediately preceding interaction trial.

For our experiment, we found a significant effect of modality ($F_{2,22} = 6.99$, $P < 0.01$) but no significant effect of menu style ($F_{1,11} = 1.01$, $P =$ n.s.) and no significant interaction between modality and menu style ($F_{2,22} = 0.03$, $P =$ n.s.). The effect of modality is shown in Fig. 10. Pairwise comparisons showed no significant differences between the audio and the dual modalities, but both of these modalities differed significantly from the visual modality ($P < 0.05$). The lower lateral velocity (i.e. higher stability) for the audio versus the visual condition indicates, not surprisingly, that the visual attention needed for the visual condition causes additional distraction and reduced performance. Interestingly, the dual condition produces essentially the same reduced distraction as the audio condition, suggesting that drivers relied on the audio portion of the dual interaction while driving (which is supported by the drivers' post-experiment reports as discussed below).

5.2.5. *Following distance*

Lateral velocity is a measure of the results of distraction on driver performance. Another measure of distraction is the following distance to the lead car: in essence, as drivers feel themselves being distracted, they tend to back away from the car in front of them for safety reasons. We computed the average following distance using the same time window around a particular trial as used for the analysis of lateral velocity.

Overall, as was found in the case of lateral velocity, there was a significant effect of modality on following distance ($F_{2,22} = 6.66$, $P < 0.01$) but there was neither significant main effect of menu style ($F_{1,11} = 0.07$, $P =$ n.s.) nor a significant interaction between modality and menu style ($F_{2,22} = 1.44$,

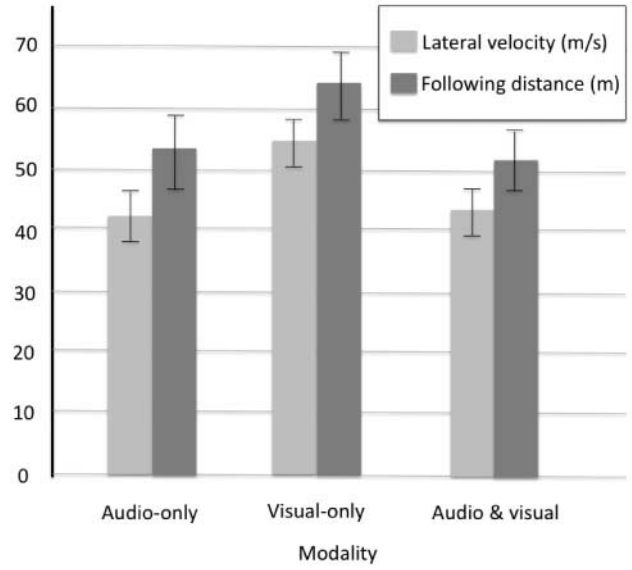

Figure 10. Lateral velocity and following distance by modality.

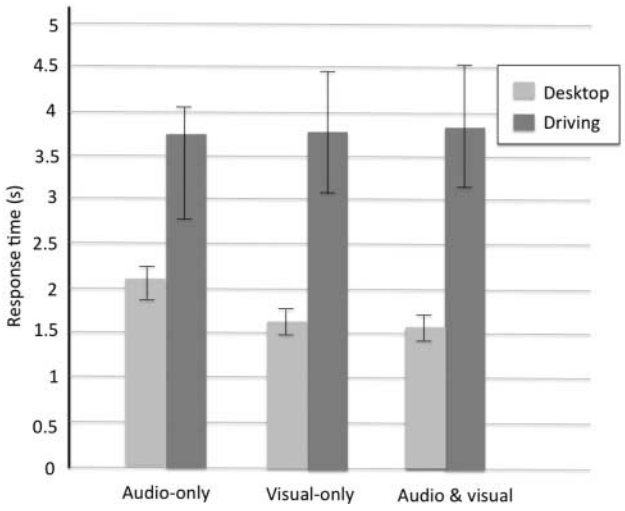

Figure 11. Experiment type × modality interaction.

P = n.s.). The average following distances by modality are shown in Fig. 10. Comparing this graph with the graph for lateral velocity, their similarity strongly suggests that drivers have a sense of the distraction potential for the three modalities: increased distraction as measured by larger lateral velocities led to increased following distances. Thus, drivers responded to the increased distraction by backing off from the lead car and giving themselves, in essence, more room for error.

5.2.6. *Observations & subjective preference*

For the same set of questions asked after the single task conditions, the preferences changed completely for driving. For the question, 'Which stimuli did you attend to during the experiment?' The answers were consistently audio (10 of 12 subjects) and only two subjects said both. For the question, 'Which feedback modality did you use under the dual-modality conditions?' The answers were again consistently 'audio'. All participants felt that audio was much safer to use than visual while driving. For the question, 'If you only used one type of feedback or primarily used only one type of feedback, did you find the other kind of feedback (audio or visual) distracting?', most users (9 of 12) reported that they totally ignored the visual feedback thus turning the dual-modality interface into an audio only interface. However, users who occasionally glanced at the visual interface found the visual feedback not just distracting, but dangerous and this point will be revisited in Section 6.

5.3. Desktop versus driving

Further interesting observations come from comparing the desktop conditions with the driving conditions. A new variable called experiment type was introduced into our analysis. The experiment type had two possible values: desktop conditions and driving conditions.

5.3.1. *Accuracy*

There was a significant main effect of experiment type, ($F_{1,11} = 27.26$, $P < 0.001$). The mean accuracy for desktop conditions (94.4%) was significantly higher than for the driving conditions (88.6%). This is not surprising since the user had to perform a more difficult task (two selections in a row) and also had to deal with a secondary driving task.

5.3.2. *Response Time*

There was a significant main effect of experiment type on response time ($F_{1,11} = 133.74$, $P < 0.001$). The mean selection time in the driving conditions (3.73 s) was significantly slower than the corresponding performance in the desktop conditions (2.63 s). This delay is likely due to the secondary driving task.

There was a significant experiment type × modality interaction, ($F_{2,22} = 12.13$, $P < 0.001$). While response time was significantly slower for the audio conditions in the desktop settings, it was no slower than the other conditions while driving (Fig. 11). This finding along with the empirical data obtained on lateral velocity and following distance all strongly suggest that the audio modality may be useful in driving, since it may increase safety without harming performance when interacting with a device.

There was also a significant experiment type × menu style interaction, ($F_{1,11} = 32.98$, $p < 0.001$). A closer examination indicates that the radial menu style has a larger advantage in terms of response time than the linear menu style for the desktop setting (Fig. 12).

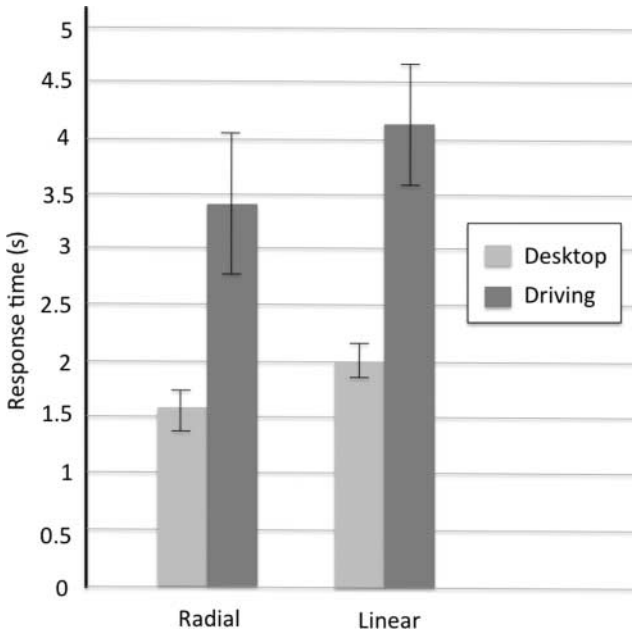

Figure 12. Experiment type × menu style interaction.

6. DISCUSSION AND DESIGN RECOMMENDATION

6.1. Audio versus visual versus dual

The most dramatic differences were found for the modality of feedback. In fact, both the results and subjective feedback differed markedly between the desktop and driving settings.

6.1.1. Desktop setting

Visual feedback was generally preferred by users, although 2 of the 12 participants told us that they preferred audio even in the desktop settings. However, even for them, performance on the visual and dual interface was much faster than the audio interfaces. Thus, visual feedback is advantageous in this setting. Users' reported experience for the dual-modality interface was interesting. These interfaces received the highest overall ranking and were ranked either as the favorite ones or right next to the favorite interfaces. However, for users who strictly preferred one kind of feedback (perhaps because they are either audio learners or visual learners), user reaction toward the other modality differed. For users who strictly prefer visual interfaces, they often found audio slightly annoying. However, people who prefer audio feedback are not affected by the presentation of the visual interfaces and tended to rank them as equally preferred to the audio only interfaces. This is perhaps due to people having the ability to close their eyes or to not pay attention to the screen if they are tired of looking (Gaver, 1997).

Overall, visual and dual interfaces were the favorites for desktop settings. Perhaps, the best strategy for the desktop setting is the dual interface but having the ability to turn off the audio or visual feedback when needed.

6.1.2. Driving setting

The change in user reaction between desktop and driving settings was dramatic. While the preference of modality still varied slightly for the desktop conditions, audio was consistently judged to be much better than visual for driving. This was true even for users who strongly preferred visual feedback in the desktop settings. One such user said after completing the driving conditions 'Although I prefer visual feedback for desktop, I found it completely useless while driving, where audio is much better'. For the dual interfaces in the desktop setting, users tend to use both modalities of feedback while performing trials. While driving, most users (9 of 12) completely ignored the visual feedback. Even for users, who occasionally glanced at the screen for extra information, they felt negative about it. As one subject put it, 'having the option to look at the visual information while driving is potential a safety hazard. I found myself tending to look at it while I was having difficulty finding the desirable item through audio, but it felt very dangerous, and I prefer an audio-only interface since it doesn't allow me to look at all'.

6.2. Linear versus radial

Compared with modality, menu style had a less dramatic effect, but still generated some interesting findings. Under both desktop and driving conditions, radial menu style yielded better performance than linear and was more preferred by users. However, compared with the desktop settings, the radial menu style did relatively better in terms of speed in the driving conditions, as described earlier by the experiment type × menu style interaction effect, (Fig. 12). This indicates that in the more difficult or complex environment, there is actually more incentive to switch to the radial menu style if possible.

6.3. Design for multiple scenarios

As devices become more powerful in terms of the number of features and more portable in terms of size and form factor, they are more likely to be used under a variety of scenarios. This imposes serious issues for interface designers since different scenarios often have very different requirements and constraints, as demonstrated by our study. The best solutions for one scenario may not be the best solution for another scenario. How to design a good design solution that works across a variety of conditions therefore presents a significant challenge for HCI Human-Computer Interaction research and interface design. Our exploration here represents one step toward further understanding user interaction with multiple modalities across multitasking environments.

6.4. Implication for design for individual scenarios

Based on the results of the experiments reported in this paper, we offer the following recommendation for the design of

menu selection interface for use in single-task (i.e. desktop-like) conditions or while the user is engaged with an on-going dynamic task (such as while driving a car).

For desktop settings, while radial and linear menu style each have their advantages—radial style is quicker to access, while linear style is more flexible and easier to design the structure and content of the menu—if the designer has a reasonably sized static menu, using visual or dual radial layout would likely yield better performance. Otherwise, visual or dual linear are also quite usable and are perhaps more suitable for menus that are longer or that have dynamic content.

For driving conditions, we recommend audio radial for reasonably sized static menu. If the menu size is longer or contains dynamic content, audio linear is probably more suitable and we do not recommend visual interfaces at all. Even the dual interfaces should be excluded if possible to avoid potential danger. In addition, although audio interfaces are safer under the driving conditions, they still impose a cognitive load which could affect a user's driving performance.

6.5. Implication for design for integrated scenario and shared input multimodal mobile interfaces

While multi-tasking in mobile scenarios has been heavily investigated (Pascoe *et al.*, 2000), a significant amount of efforts have been taken into the design of eyes-free interface/interaction techniques for mobile devices; however, as Pascoe *et al.* (2000) have pointed out, while mobile devices can be accessed on the move, they are frequently used in stationary scenarios where users have the majority of their visual attention available for mobile HCI tasks. In such cases, the visual interface will often have an advantage. While earlier research often tends to focus on either the general use or eyes-free use of mobile devices (Fig. 13, usage Scenario 2), we believe that both stationary and eyes-free usage scenarios are equally important for mobile interface design and need to considered as an integrated whole rather than two separate scenarios.

In this integrated scenario, users will use their mobile devices either stationary or on-the-move and often need to switch between these scenarios according to context. To design for this integrated scenario, it requires designers to consider interface solutions that are suitable for both scenarios and support the easy switch between their usages to seek a balanced design.

However, the design of such type of integrated interface can be difficult since the design requirement for the stationary use is quite different from on-the-move use. As demonstrated in our experiment, eyes-free auditory interface is more suitable in the driving scenario while the visual interface is optimal for the desktop usage scenario.

A possible and seemingly promising solution for this type of integrated mobile scenario is the shared input multimodal mobile interface (Fig. 14, right). Since shared input multimodal interfaces share the same input mechanism, they require less additional effort to learn. Furthermore, since the input mechanism is shared between two interfaces, the motor skill required to operate both interfaces is the same. Using either

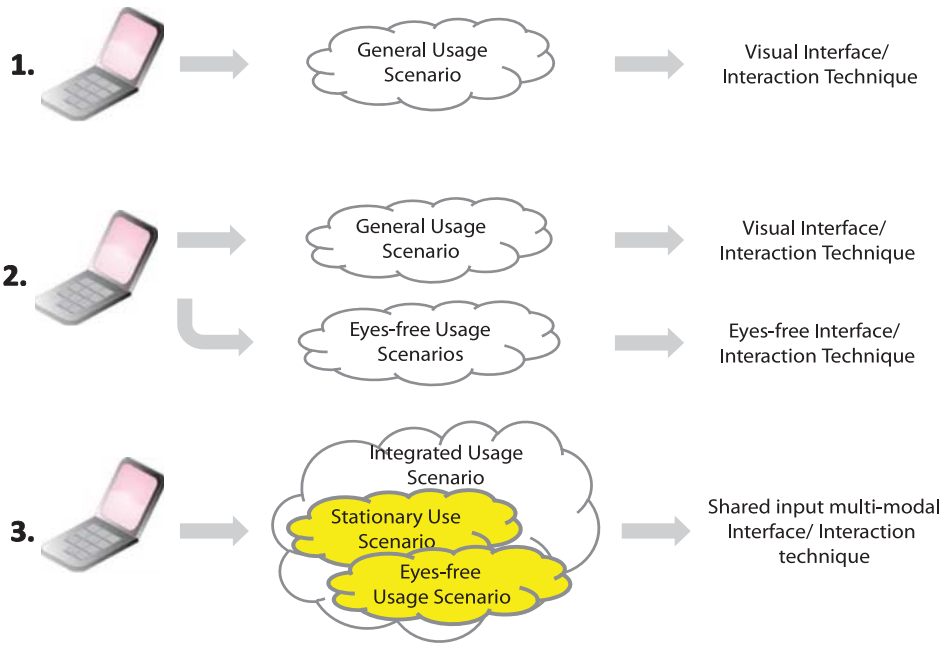

Figure 13. Three approaches for designing mobile interfaces and interaction techniques. The first two represent more traditional approaches while the third one is our proposed approach.

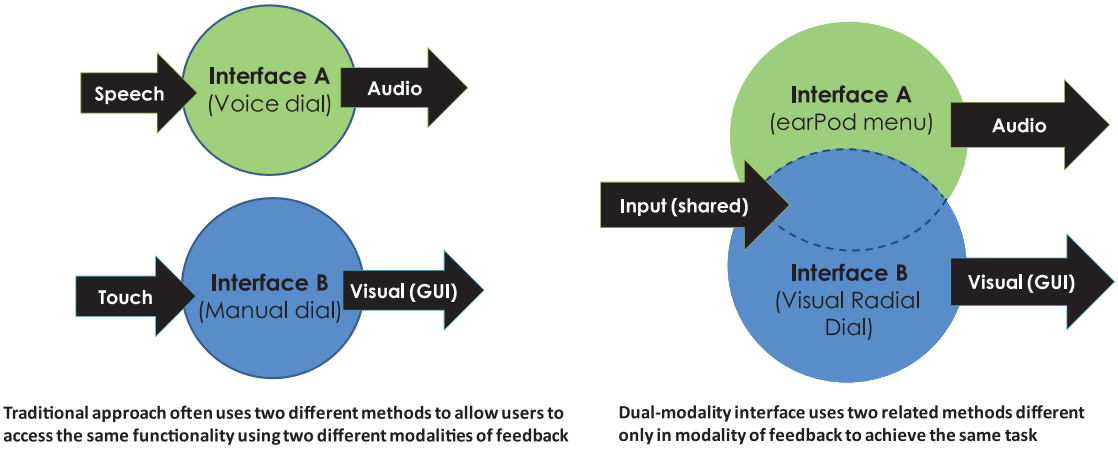

Figure 14. Two interfaces with different input-output modalities (left) versus the shared input multimodal interface.

interface also trains the use of the other interface, which could potentially reduce the learning time for users to achieve expert performance in both interfaces. The concept of using common input and multiple output modalities is not new as seen in Brewster and Crease's (1999) audio-enhanced widgets, who also used multiple output modalities for the same input mechanism. However, in their work, the multiple output modalities complement each other for the same task in the same scenario. As stated by Brewster and Crease (1999), their 'aim was to enhance standard graphical menus with more salient feedback to see if menu errors could be solved and also to see if sound was effective as the feedback'. Their approach is not designed for the diverse usage scenarios (which often include both stationary and on-the-move use (Pascoe *et al.*, 2000)) often encountered by users on their mobile devices today. In our approach, the multiple modalities work independently and the audio modality have an equivalent role as the visual modality. Our approach is designed for both stationary and on-the-move usage of mobile devices.

The type of multimodal systems, we are proposing, is a specific type of multimodal system where the audio and visual feedbacks are independent and used non-concurrently. This is considered as an exclusive system according to Nigay and Coutaz's (1993) taxonomy and an equivalent output multimodal system according to Coutaz *et al.* (1995). We envision that this type of interface is particularly useful for mobile devices due to their diverse usage scenarios. Therefore, we call our proposed approach the 'shared input multimodal mobile interfaces'.

This type of interface can be accessed independently by either modality alone (not excluding the possibility of using them at the same time); this is different from a multimodal interface where an interface generates feedback using a number of modalities that often complement each other, but not independently. Our results showed that performance and user preference can change dramatically from one scenario to the other; therefore, coercing users to adopt one type of modality is not desirable, in our opinion. Instead it is far better to let users decide which output modality they wish to attend to in a given situation. Of course this leaves open the important question of how users will go about deciding which output modality to use in a given situation (but see, Brumby *et al.*, 2011, for an approach for tackling this question).

Critically, we advocate that the method for completing a task on a device should be invariant to the interaction technique used. For instance, in the case of *earPod*, both visual and auditory methods of feedback operate in a consistent manner, meaning that the user need learn only a single method for selecting an item from the menu. For this to be effective it is important that there is a common interaction style that operates across the various output modalities that can be used. At the time of this writing, there are very few interfaces that support such shared input multimodal styles of interaction, and further exploration and evaluation for such interfaces in other application domains might provide a rich and fruitful avenue of research. We believe that the shared input multimodal systems deserve more attention from both mobile researchers and designers.

7. STUDY LIMITATIONS

Due to constraints to keep the experimental procedure within a reasonable time limit, we did not have the opportunity to consider possible learning behavior over a prolonged period of usage; the requirements of the driving task did not allow time for the inclusion of this interesting factor. It is therefore difficult to rule out the possibility that some of the users in our experiment could come to learn to drive more 'safely', even in the visual condition. Indeed, it is well known that performance improves

following a power law of practice (Newell and Rosenbloom, 1981). Further research is required to investigate asymptotic performance. However, since driving is considered a high-risk task, the potential cost associate with in-car training is extremely high. Even if an interface can be learned to be safer in a car, any mistakes during practice could potentially be catastrophic. Thus, the current experimental setting may have practical value in guiding safe vehicle interface design.

8. CONCLUSION

In conclusion, we investigated the effect of alternative feedback modalities, i.e. audio, visual and audio–visual feedback and menu layout on user performance and preference for menu selection tasks in a single-task desktop setting and a dual-task driving setting. Experimental results indicated that different operational environments can have strong effects on the performance of menu selection using different types of modality of feedback (audio, visual and audio–visual) and different styles of menu layout (linear or radial). Visual feedback produced better user performance and is preferred under single-task conditions; in dual-task conditions it presented a significant source of driver distraction. In contrast, auditory feedback mitigated some of the risk associated with menu selection while driving.

Although driving is an important mobile scenario, there are other common usage scenarios that we have not evaluated here. While not formally tested, the results of experiments are likely to apply to walking and running scenarios since in both cases, users need to pay attention to the road and the environment. We expect the audio menus can benefit the users more in the running scenarios than that of the walking scenario due to increased inconvenience of visually checking status of the mobile device. It will be interesting to verify this hypothesis in future studies.

ACKNOWLEDGMENTS

We thank members of the NUS-HCI Lab for their support. This research is supported by National University of Singapore Academic Research Fund WBS R-252-000-414-101. We thank all of the reviewers who have provided constructive comments on this work and have helped strengthen this paper.

REFERENCES

2006 GMAC Insurance National Drivers Test http://www.gmacinsurance.com/SafeDriving/2006/ (accessed October 14, 2010).

Akamatsu, M., Mackenzie, I.S. and Hasbroucq, T. (1995) A comparison of tactile, auditory, and visual feedback in a pointing task using a mouse-type device. Ergonomics, 38, 816–827.

Alm, H. and Nilsson, L. (1994) Changes in driver behaviour as a function of hands-free mobile phones—a simulator study. Accident Anal. Prev., 26, 441–451.

Arons, B. (1997) SpeechSkimmer: a system for interactively skimming recorded speech. ACM Trans. Comput.-Hum. Interact., 4, 3–38.

Brewster, S.A. (1998) Using nonspeech sounds to provide navigation cues. ACM Trans. Comput.-Hum. Interact. (TOCHI), 5, 224–259.

Brewster, S.A. and Crease, M.G. (1999) Correcting menu usability problems with sound. Behav. Inf. Technol., 18, 165–177.

Brewster, S., Lumsden, J., Bell, M., Hall, M. and Tasker, S. (2003) Multimodal 'Eyes-Free' Interaction Techniques for Wearable Devices. In ACM CHI Conf. Human Factors in Computing Systems, Ft. Lauderdale, FL, pp. 473–480.

Brumby, D.P., Salvucci, D.D. and Howes, A. (2009) Focus on Driving: How Cognitive Constraints Shape the Adaptation of Strategy when Dialing while Driving. In Proc. CHI 2009, pp. 1629–1638. ACM Press.

Brumby, D.P., Davies, S.C.E., Janssen, C.P. and Grace, J.J. (2011) Fast or safe? How performance objectives determine modality output choices while interacting on the move. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI'11 (pp. 473–482). ACM Press, New York.

Careless Talk http://news.bbc.co.uk/2/hi/uk_news/magazine/6382077.stm (accessed October 14, 2010).

Callahan, J., Hopkins, D., Weiser, M. and Shneiderman, B. (1988) An empirical comparison of pie vs. linear menus. In Proc. CHI 1988, pp. 95–100. ACM Press.

Coutaz, J, Nigay, L., Salber, D., Blandford, A., May, J. and Young, R.M. (1995) Four Easy Pieces for Assessing the Usability of Multimodal Interaction: The CARE Properties. In Interact 1995, Lillehammer, Norway, pp. 115–120.

Countries that Ban Cell Phones while Driving (2010) http://www.cellular-news.com/car_bans/ (accessed October 14, 2010).

Dingler, T., Lindsay, J. and Walker, B.N. (2008) Learnability of Sound Cues for Environmental Features: Auditory Icons, Earcons, Spearcons, and Speech. In Proc. Int. Conf. Auditory Display (ICAD 2008), Paris, June 24–27.

Gaver, W. (1989) The sonic finder: an interface that uses auditory icons. Hum. Comput. Interact., 4, 67–94.

Gaver, W.W. and Smith, R.B. (1991) Auditory icons in large-scale collaborative environments. SIGCHI Bull. 1991, 23, 96.

Gaver, W. (1997) Auditory Interfaces. Handbook of Human–Computer Interaction, pp. 1003–1041. Elsevier Science, The Netherlands.

Graf, S., Spiessl, W., Schmidt, A., Winter, A. and Rigoll, G. (2008) In-Car Interaction Using Search-Based User Interfaces. In Proc.CHI 2008, Florence, Italy, pp. 1685–1688. ACM Press.

Ho, C., Reed, R. and Spence, C. (2007) Multisensory in-car warning signals for collision avoidance. Hum. Factors J. Hum. Factors Ergonom. Soc., 49, 1107–1114

Jacko, J., Scott, I., Sainfort, F., Barnard, L., Edwards, P., Emery, V., Kongnakorn, T., Moloney, K. and Zorich, B. (2003) Older adults and visual impairment: what do exposure times and accuracy tell

us about performance gains associated with multimodal feedback? In Proc. CHI 2003, Fort Lauderdale, FL, pp. 33–40.

KidsClick! (2010) Web search for kids by librarians. http://sunsite.berkeley.edu/KidsClick!/ (accessed October 14, 2010).

Kurtenbach, G. (1993) The design and evaluation of marking menus. Ph.D. Thesis, University of Toronto, Toronto.

Kurtenbach, G. and Buxton, W.S. (1993) The Limits of Expert Performance Using Hierarchic Marking Menus. In Proc. CHI 1993, pp. 482–487, ACM Press.

Kurtenbach, G. and Buxton, W.S. (1994) User Learning and Performance with Marking Menus. In Proc.CHI 1994, pp. 258–264. ACM Press.

Luk, J., Pasquero, J., Little, S., Maclean, K., Levesque, V. and Hayward, V. (2006) A Role for Haptics in Mobile Interaction: Initial Design Using a Handheld Tactile Display Prototype. In Proc. CHI2006, pp. 171–180. ACM Press.

Mcknight, A.J. and Mcknight, A.S. (1993) The effect of cellular phone use upon driver attention. Accident Anal. Prev., 25, 259–265.

Minoru, K. and Schmandt, C. (1997) Dynamic Soundscape: Mapping Time to Space for Audio Browsing. In Proc. CHI1997, pp. 194–201. ACM Press.

Mynatt, E.D. (1995) Transforming Graphical Interfaces Into Auditory Interfaces. In Proc. CHI1995, pp. 67–68. ACM Press.

Newell, A. and Rosenbloom, P.S. (1981) Mechanisms of Skill Acquisition and the Law of Practice. In Anderson, J. R. (ed.), Cognitive skills and Their Acquisition, pp. 1–51. Lawrence Erlbaum Associates, Hillsdale, NJ.

Nigay, L. and Coutaz, J. (1993) A Design Space for Multimodal SysTEMS: Concurrent Processing and Data Fusion. In Proc. INTERCHI'93, Amsterdam, pp. 172–178. ACM Press.

Pascoe, J., Ryan, N. and Morse, D. (2000) Using while moving: HCI issues in fieldwork environments. ACM Trans. Comput. Hum. Interact., 7, 417–437.

Pfleging, B., Kienast, M., Schmidt, A., Döring, T. and et al. (2011) SpeeT: A Multimodal Interaction Style Combining Speech and Touch Interaction in Automotive Environments. In Adjunct Proc. 3rd Int. Conf. Automotive User Interfaces and Vehicular Applications, Salzburg, Austria.

Pirhonen, A., Brewster, S. and Holguin, C. (2002) Gestural and Audio Metaphors as a Means of Control for Mobile Devices. In Proc. CHI 2002, pp. 291–298. ACM Press.

Rinnot, M. (2005) Sonic Texting. ACM CHI Extended Abstracts on Human Factors in Computing Systems, pp. 1144–1145. ACM Press.

Rockbox (2010) http://www.rockbox.org/ (accessed October 14, 2010).

Roy, D.K. and Schmandt, C. (1996) NewsComm: A Hand-Held Interface for Interactive Access to Structured Audio. In Proc. CHI 1996, pp. 173–180. ACM Press.

Salmen, A., Grobmann, P., Hitzenberger, L. and Creutzburg, U. (1999) Dialog Systems in Traffic Environment. In Proc. ESCA: Tutorial and Research Workshop on Interactive Dialogue in Multi-Modal Systems, Kloster Irsee, Germany.

Salvucci, D.D. (2001) Predicting the effects of in-car interface use on driver performance: an integrated model approach. Int. J. Hum.-Comp. Stud., 55, 85–107.

Salvucci, D.D. (2005) A multitasking general executive for compound continuous tasks. Cogn. Sci., 29, 457–492.

Salvucci, D.D., Markley, D., Zuber, M. and Brumby, D.P. (2007) iPod distraction: effects of Portable Music-Player Use on Driver Performance. In Proc. CHI 2007, pp. 243–250. ACM Press.

Sawhney, N. and Schmandt, C. (2000) Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments. ACM Trans. Comput.-Hum. Interact., 7, 353–383.

Schmandt, C. (1998) Audio Hallway: A Virtual Acoustic Environment for Browsing. In Proc. UIST1998, pp. 163–170. ACM Press.

Schmandt, C., Lee, K., Kim, J. and Ackerman, M. (2004) Impromptu: Managing Networked Audio Applications for Mobile Users. In Proc. Int. Conf. Mobile Systems, Applications, and Services, pp. 59–69. ACM Press.

Sears, A. and Shneiderman, B. (1994) Split menus: effectively using selection frequency to organize menus. ACM Trans. Comput.-Hum. Interact., 1, 27–51.

Stifelman, L., Arons, B. and Schmandt, C. (2001) The Audio Notebook: Paper and Pen Interaction with Structured Speech. In Proc. CHI 2001, pp. 182–189. ACM Press.

Stifelman, L.J., Arons, B., Schmandt, C. and Hulteen, E.A. (1993) VoiceNotes: A Speech Interface for a Hand-Held Voice Notetaker. In Proc. INTERCHI1993, pp. 179–186. ACM Press.

Strayer, D.L. and Johnston, W.A. (2001) Driven to distraction: dual-task studies of simulated driving and conversing on a cellular phone. Psychol. Sci., 12, 462–466.

Strayer, D.L., Drews, F.A. and Johnston, W.A. (2003) Cell phone-induced failures of visual attention during simulated driving. J. Exp. Psychol. Appl., 9, 23–32.

Sodnik, J., Dicke, C., Tomazic, S. and Billinghurst, M. (2008) A user study of auditory versus visual interfaces for use while driving. Int. J. Hum. Comput. Stud., 66, 318–332.

Vitense, H.S., Jacko, J.A. and Emery, V.K. (2002) Foundation for improved interaction by individuals with visual impairments through multimodal feedback. Univers. Access Inf. Soc., 2, 76–87.

Wagner, C.R., Lederman, S.J. and Howe, R.D. (2004) Design and performance of a tactile shape display using RC servomotors. Haptics-e Electron. J. Haptics Res., 3, 4. www.haptics-e.org.

Walker, B.N., Nance, A. and Lindsay, J. (2006) Spearcons: Speech-based Earcons Improve Navigation Performance in Auditory Menus. In Proc. Int. Conf. Auditory Display (ICAD 2006), London, June 20–24, pp. 63–68

Wickens, C.D. (2002) Multiple resources and performance prediction. Theor. Issues Ergon. Sci., 3, 159–177.

Wikipedia Contents: Categories (2010) http://en.wikipedia.org/wiki/Portal:Contents/Categories (accessed October 14, 2010).

Yi, Bo, Xiang, C., Morten, F. and Zhao, S. (2012) Exploring User Motivations for Eyes-free Interaction on Mobile Devices. In ACM CHI 2012, pp. 2789–2792.

Yin, M. and Zhai, S. (2006) The Benefits of Augmenting Telephone Voice Menu Navigation with Visual Browsing and Search. In Proc. CHI 2006, pp. 319–328. ACM Press.

Zhao, S., Agrawala, M. and Hinckley, K. (2006) Zone and Polygon Menus: Using Relative Position to Increase the Breadth of Multi-Stroke Marking Menus. In Proc. CHI 2006, pp. 1077–1086. ACM Press.

Zhao, S. and Balakrishnan, R. (2004) Simple vs. Compound Mark Hierarchical Marking Menus. In Proc. UIST 2004, pp. 33–42. ACM Press.

Zhao, S., Dragicevic, P., Chignell, M., Balakrishnan, R. and Baudisch, P. (2007) earPod: Eyes-Free Menu Selection with Touch Input and Reactive Audio Feedback. In Proc. CHI 2007, pp. 1395–1404. ACM Press.