Menu Navigation With In-Vehicle Technologies: Auditory Menu Cues Improve Dual Task Performance, Preference, and Workload

Myounghoon Jeon[1], Thomas M. Gable[2], Benjamin K. Davison[2], Michael A. Nees[3], Jeff Wilson[2], and Bruce N. Walker[2]

[1]Mind Music Machine Lab

Department of Cognitive and Learning Sciences

Michigan Technological University

Houghton, MI, USA

[2]Sonification Lab

School of Psychology

Georgia Institute of Technology

Atlanta, GA, USA

[3]Human Factors, Perception, & Cognition Lab

Department of Psychology

Lafayette College

Easton, PA, USA

Corresponding author:

Myounghoon Jeon

Department of Cognitive and Learning Sciences

Michigan Technological University

1400 Townsend Drive, Houghton, MI 49931-1295

Phone: +1 906. 487. 3273

Fax: +1 906. 487. 2468

mjeon@mtu.edu

ABSTRACT

Auditory display research for driving has mainly examined a limited range of tasks (e.g., collision warnings, cell phone tasks). In contrast, the goal of this project was to evaluate the effectiveness of enhanced auditory menu cues in a simulated driving context. The advanced auditory cues of ‘spearcons’ (compressed speech cues) and ‘spindex’ (a speech-based index cue) were predicted to improve both menu navigation and driving. Two experiments used a dual task paradigm in which users selected songs on the vehicle’s infotainment system. In Experiment 1, 24 undergraduates played a simple, perceptual-motor ball-catching game (the primary task; a surrogate for driving), and navigated through an alphabetized list of 150 song titles—rendered as an auditory menu—as a secondary task. The menu was presented either in the typical visual-only manner, or enhanced with text-to-speech (TTS), or TTS plus one of three types of additional auditory cues. In Experiment 2, 34 undergraduates conducted the same secondary task while driving in a simulator. In both experiments, performance on both the primary task (success rate of the game or driving performance) and the secondary task (menu search time) was better with the auditory menus than with no sound. Perceived workload scores, as well as user preferences favored the enhanced auditory cue types. These results show that adding audio, and enhanced auditory cues in particular, can allow a driver to operate the menus of in-vehicle technologies more efficiently while driving more safely. Results are discussed with multiple resources theory.

Keywords: auditory display; dual task; spearcon; spindex; TTS (text-to-speech); multiple resources theory

1. INTRODUCTION

Emerging wireless and digital technologies have allowed an abundance of information to be delivered via mobile devices. This information portability has extended to the driver's seat in the form of 'in-vehicle technologies' (IVTs, see Horrey and Wickens, 2004; Horrey, Wickens, and Consalus, 2006). IVTs can deliver such diverse digital media as driving-relevant information (e.g., navigation instructions or weather and traffic updates), in-vehicle entertainment (e.g., music, video, or television), and productivity applications (e.g., cellular phone or wireless Internet) for the driver and passengers.

As IVTs become more complex, problems of driver inattention have become worse (Ashley, 2001; Dukic, Hanson, and Falkmer, 2006; Patten, Kircher, Ostlund, and Nilsson, 2004). A critical concern involves the extent to which secondary tasks encouraged by IVTs interfere with the visually demanding driving task because those secondary tasks have been shown to negatively affect driving performance and increase perceived workload (Lansdown, Brook-Carter, and Kersloot, 2004; Tsimhoni and Green, 2001).

Despite the potential pitfalls of IVTs with respect to driver distraction, it has been argued that such technologies can be safely integrated into automobiles, and some practice guidelines have even been proposed (e.g., Burns and Lansdown, 2000; Harvey, Stanton, Pickering, McDonald, and Zheng, 2011). Research has found that younger adults accomplished a task that required reading text messages aloud from an IVT system with surprisingly little impact on simulated driving performance, although this promising finding did not hold for older adults (Schieber, Holtz, Schlorholtz, and McCall, 2008). Aftermarket controllers have also been applied to vehicles for safety purposes. However, Lee, Roberts, Hoffman, and Angell (2012) recently showed that some aftermarket controllers could degrade rather than enhance performance, and thus should be further validated. Given that IVTs and accompanying distractions appear to be a common component of the modern automobile, the appropriate design of safe IVTs remains a challenge that must be addressed by further research. Auditory information presentation represents an obvious alternative to visual information presentation for IVTs.

1.1. Auditory and Multimodal Presentation for IVTs

Much research has demonstrated the detrimental effects of using mobile phones in a car (Lansdown et al., 2004; Strayer and Drews, 2007; Strayer, Drews, and Johnston, 2003; Strayer and Johnston, 2001; Young and Regan, 2007). In those studies, impairment of the primary task (a measure of driving behavior) occurred, even when participants used a hands-free phone (Strayer and Drews, 2007). However, the potential distraction caused by phones and IVTs has not deterred people from using them for diverse purposes. One issue with research on use of phones while driving is that even though it can clearly show negative effects on driving, it cannot force phone providers to enhance their phones specifically for in-vehicle use. In contrast, IVT research can suggest practical guidelines to carmakers or IVT designers (e.g., how to develop IVTs to be safer, as well as how to integrate third-party mobile devices to be compatible with their IVTs).

Studies have demonstrated that negative effects of a secondary task can be mitigated by reducing the task's complexity and physical, cognitive, and attentional demands (Dukic, Hanson, and Falkmer, 2006; Patten et al., 2004; Ranney, Harbluk, and Noy, 2005). For example, offloading drivers' information processing onto alternative perceptual channels via auditory and multimodal presentation can facilitate performance with interfaces where visual overload tends to occur (Brewster, 1997; Brown, Newsome, and Glinert, 1989). Research has further suggested that auditory and multimodal IVTs may overcome some of the problems associated with visually taxing IVTs (for a review, see Nees and Walker, 2011). To illustrate, Liu (2001) found that both driving and secondary task performance can be better when using auditory—and particularly multimodal—in-vehicle information displays.

1.2. Multiple Resources vs. Auditory Preemption

Multiple resources theory (see, e.g., Wickens, 2002; Wickens and McCarley, 2008) has often been invoked to explain the apparent benefit of dividing information display across modalities during multitasking. Multiple resources theory would predict that concurrent auditory and visual tasks draw upon separate pools of modality resources and thus, should be 'time-shared' efficiently (i.e., without interrupting each other) as long as they do not require the same processing code, stage of processing, or response modality. Other studies (Horrey and Wickens, 2004; Wickens and Liu, 1988), however, have suggested that a discrete auditory task 'preempts,' or causes a brief lapse in, the performance of a continuous visual task while the auditory stimulus is attended to, owing to the auditory modality's superior ability to attract attention (Spence and Driver, 1997; Stanton, Booth, and Stammers, 1992). As such, an auditory cost has been found in a number of studies that examined the modality of in-vehicle information displays (Dingus et al., 1997; Horrey and Wickens, 2004; Lee, Gore, and Campbell, 1999; Matthews, Sparkes, and Bygrave, 1996). The results from these studies suggest that the potential modality benefits of auditory (rather than visual) presentation of secondary task information might be suppressed by processing mechanisms (such as preemption*,* described above) and display characteristics. Related research has shown that even hands-free auditory cell phone conversations impair driving (Strayer and Drews, 2007). In other studies, both auditory costs and auditory benefits for in-vehicle information displays have been shown (Mollenhauer, Lee, Cho, Hulse, and Dingus, 1994; Parkes and Burnett, 1993; Ranney et al., 2005), whereas much research has shown the intuitively predicted auditory benefit for both tasks (Burnett and Joyner, 1997; Dingus, Hulse, McGehee, and Manakkal, 1994; Gish, Staplin, Stewart, and Perel, 1999; Liu, 2001; Streeter, Vitello, and Wonsiewicz, 1985; Walker, Alicandri, Sedney, and Roberts, 1991).

Taken together, these findings suggest qualified successes for the implementation of auditory displays in IVTs, but the precise circumstances in which auditory cues help or harm performance of a visual primary task and the exact locus of interference remain to be determined. The current study examined the impact of a number of recently developed enhanced auditory cues on performance of a perceptual-motor visual primary task (Experiment 1) and a simulated driving task (Experiment 2). See 1.4. The Current Study and Hypotheses section for more details about the relationship of the two tasks.

1.3. Enhanced Auditory Cues in Menu Navigation

The use of sound to communicate information about the driving task itself (e.g., warnings relating to the vehicle status or the presence of an approaching vehicle; see Ho and Spence, 2005) must be distinguished from the use of sound as a means of interacting with the IVT systems (i.e., ‘infotainment’ systems; see Nees and Walker, 2011). The content of infotainment IVTs is often organized into a menu structure through which the driver (or passenger) must navigate in order to select the desired option (e.g., to play a particular song or to retrieve directions to a particular restaurant). Relatively little research has examined the use of sound in this specific context, even though audio might improve overall performance and safety (as well as user workload, stress, and satisfaction ratings) compared to visual-only menu structures (Walker and Nees, 2012).

Typically, sound is used in such menus simply by playing aloud the menu items via text-to-speech (TTS) synthesis, but more can be done to enhance auditory menus. Non-speech cues, for example, can supplement spoken menu items. The present study focuses on the use of non-speech cues to enhance a spoken auditory menu. Our recent research in this area has specifically examined spearcons and spindex cues, described below.

1.3.1. Spearcons: Compressed Speech Sounds

Spearcons (short for ‘speech earcons’) are brief sounds that are produced by speeding up spoken phrases, to the point where the resulting sound is no longer comprehensible as a particular word (Walker, Nance, and Lindsay, 2006). Despite the compression, spearcons have the same amount of information value (no truncation) as speech and thus, they should be better in terms of performance and learning than other arbitrary mappings of non-speech sounds. Because of this acoustic relationship with the original speech phrases, these unique sounds are analogous to fingerprints. Spearcons are easily created by converting the text of a menu item to speech via

TTS and speeding it up using a pitch-constant compression algorithm, a process that allows a user interface to support dynamic menus. Typically, spearcons are played just before (or may even entirely replace, e.g., Suh, Jeon, and Walker, 2012) the spoken menu item, which leads to better performance and learning rates, compared to menus composed of TTS alone, or to menus that use TTS along with other well-known non-speech auditory cues, such as auditory icons (representative sounds of events or objects, Gaver, 1986) and earcons (brief musical motives, Blattner, Sumikawa, and Greenberg, 1989). For example, Walker et al. (2006) showed that spearcons resulted in faster and more accurate performance than other auditory cues for a search task. Spearcons also improved navigation efficiency over TTS only or no sound when combined with visual cues (Palladino and Walker, 2008a, 2008b; Walker and Kogan, 2009). Other studies (Dingler, Lindsay, and Walker, 2008; Palladino and Walker, 2007) have demonstrated that spearcons are as learnable as speech, but auditory icons and earcons are more difficult to learn. See Walker, Lindsay, Nance, Nakano, Palladino, Dingler, and Jeon (2012) for a summary of spearcons research.

1.3.2. Spindex: Speech Index

A spindex (short for speech index, Jeon and Walker, 2009) is created by associating an auditory cue with each menu item, based on the pronunciation of the first letter (or phoneme) of each menu item. For instance, the spindex cue for “All the above” would be a sound based on the spoken sound “A” (i.e., /ei/ or even /a/). 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 correct section of a thick reference book such as a dictionary or a telephone book; analogous visual indexes have been used, for example, in newer Apple iPods. Given that the song list or address list in the infotainment system typically uses an alphabetical order as its default, spindex cues are also expected to apply easily to in-vehicle interfaces.

The benefit of adding a spindex to a menu can be explained by the fact that users employ a combination of rough and fine navigation strategies in the search process (Klante, 2004). In the rough navigation stage, users invoke top-down knowledge about the serial order of the alphabet to exclude non-targets until they approach the alphabetical area proximal to the target. After users perceive that they have reached the target zone, they then need more precise, detailed information to select the target menu item; this is the fine navigation stage. The spindex-enhanced auditory menu can contribute per-item speedups during the rough search stage, while still supporting detailed item information via the TTS phrase in the fine search stage.

Spindex cues are natural sounds (based on speech) and part of the original word, and thus, do not require training to learn the relationship between the sound and its intended meaning. In previous research, participants showed better performance in a TTS + spindex condition than in a TTS-only condition. A subsequent study showed that alternative designs (e.g., *attenuated* types, discussed below) further improved user acceptance and performance (for more details of the spindex cue types, see Jeon and Walker, 2011). Research demonstrated benefits of spindex for visually impaired users (Jeon and Walker, 2011) and while using various interaction styles on a touch screen device (Jeon, Walker, and Srivastava, 2012).

1.4. The Current Study and Hypotheses

With respect to the menu-oriented tasks often required to select content in IVTs, relatively little research has examined the potential for audio cues to reduce conflicts with a visual primary task. Conflicting results have suggested that auditory secondary tasks may sometimes preempt performance of a visual primary task, whereas other results have shown an advantage for auditory presentation of a secondary task in the presence of a visual primary task. Furthermore, the extent to which enhanced auditory cues (spearcon and spindex) may reduce distraction and improve safety of IVTs has yet to be established. Speaking out menu items using TTS per se could preempt performance of a visual primary task just as Strayer and Drews (2007) found. Further, adding *unexpected* enhanced auditory cues (spearcon or spindex) to TTS can make the overall auditory presentation unfamiliar and longer and thus, auditory preemption might occur and the benefits of using auditory menus might decrease even more. On the other hand, the use of auditory displays could improve performance in the dual task situation, as the multiple resources theory would predict.

To investigate these issues, the current study devised a plausible secondary task in which participants navigated a song list on an in-vehicle infotainment system. For this scenario, a dual task (or divided attention) paradigm (Treisman and Davies, 1973) was used to examine the effectiveness of five types of auditory cues on performance for a primary driving-like task and and a concurrent secondary menu search task.

Conducting research on in-vehicle interfaces might not need a full driving simulator to study prototype interfaces, just as using a driving simulator provides meaningful data as actual driving (Bedard, Parkkari, Weaver, Riendeau, and Dahlquist, 2010). Indeed, researchers often use an abstracted task for driving research, such as peripheral detection tasks or the lane change task (Mattes and Hallen, 2009). Note that in any case, the driving-like task has to be visual-manual, include tracking, and be instrumented so that researchers can measure performance on the task. Abstraction might be good in terms of generalization. If we conduct research with actual driving, the results could be applied to *only* driving. However, if we conduct research on a psychologically abstracted level, we might postulate the similar hypotheses and principles to various dynamic situations, such as aviation, driving, or other domains that have the same stimulus-response relationship. To this end, in Experiment 1 we used a simple ball-catching game as an abstracted task and in Experiment 2 we used a simulated driving task as a more specific application. Of course, the repetition of the study can also allow us to achieve more reliable insight about the effects of auditory cues.

To evaluate overall effects of adding auditory displays, we measured multiple dependent variables (e.g., success rate, reaction time, workload, preference, etc.) (e.g., Chang, Hwang, and Ji, 2011). We predicted that the displays enhanced with auditory cues would shorten the navigation time in the secondary task, and also that the primary task would be less affected by the secondary task when auditory cues were used. The combined workload of the task configuration was predicted to be attenuated by the use of auditory cues. With respect to the relative effectiveness of and participant preference for auditory cues, we predicted that enhanced auditory cues (i.e., spearcons and spindex) would outperform traditional TTS cues. To test these hypotheses, we conducted two empirical experiments with five different auditory cue types (TTS-only; spearcon + TTS; spindex + TTS; spindex + spearcon + TTS; and no sound).

2. Experiment 1

Experiment 1 used a simple ball-catching computer game that required perceptual vigilance and nearly constant manual control, which is a perceptual-motor task. This type of abstracted task (e.g., Mattes, 2003; Mattes and Hallen, 2009) has frequently been used in driving research domain and is psychologically comparable to vehicle control during driving.

2.1. Method

2.1.1. Participants

Twenty-four undergraduate students (10 female and 14 male; mean age = 20.2, *SD* = 1.2) participated in this study for credit in a psychology course in Spring 2009 at the Georgia Institute of Technology. Participants reported normal or corrected-to-normal vision and hearing, and gave informed consent.

2.1.2. Apparatus

Figure 1 shows the experimental apparatus. The primary task stimuli were presented using a Dell Dimension XPS T600 computer, running Windows XP on a Pentium 3, 598 MHz processor with 512 MB of RAM. A 17-inch monitor was placed on a table 50 cm (20 in.) in front of the seated participant. For the secondary task, stimuli were presented using an Azentek in-vehicle infotainment system, running Windows Vista on a Pentium 4, 1.83 GHz processor with 1 GB of RAM. A Sigma Tel High Definition audio output device was used for sound rendering. Participants listened to auditory stimuli using harman/kardon HK195 speakers located 30 cm (12 in.) behind the primary task monitor. The infotainment system included a 6.5-inch resistive touch screen panel. The infotainment system was located at a usual in-vehicle position—approximately

34 cm (13.4 in.) below and 37 cm (14.6 in.) to the right from the center of the primary task monitor (Horrey and Wickens, 2004) (See Figure 1).

2.1.3. Stimuli

2.1.3.1. Primary Task.

The primary task was a visual perceptual-motor vigilance task and was pilot-tested to be sufficiently difficult for dual task decrements to be observed when the secondary menu task was introduced (for a discussion of the importance of task difficulty in dual task scenarios, see Gopher, Brickner, and Navon, 1982). The simple computer game (see Figure 1) was programmed in Visual Basic 6.0 and consisted of balls that dropped along 10 vertical columns from the top of the screen at a rate of approximately 1 ball per second. The purpose of the game was to catch all of the balls before they reached the bottom of the screen, by moving a “basket” across the bottom of the screen. When a ball was successfully captured, the basket flashed from black to green. To control the box, participants placed the index and middle finger of their left hand on the right and left arrow keys on the keyboard, respectively. Five pilot subjects allowed us to establish the baseline performance of the primary task at 92.11% accuracy (*SD* = 5.31) for catching the balls over a 1-minute trial period.

2.1.3.2. Secondary Task.

The secondary task, an IVT menu navigation, was designed as a song selection task. A song list 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/). A visual menu (see Figure 2) was created in C# using the Centrafuse SDK programming tools for use as a plug-in for the Centrafuse 2.1 infotainment user interface (www.centrafuse.com). The menu items were in alphabetical order, and the participant was able to scroll downward and upward in the menu by pressing arrow buttons on the touch screen. One arrow press moved the selected item down by one menu position, and the display advanced upon any arrow press in which the next item was on a different page. The participant's objective was to reach the given target name in the list menu as quickly as possible. Participants logged their selection as the current active item by pressing a "select" button (top right of Figure 2). If the participant reached the top or bottom of the menu, the list did not wrap around.

In addition to the visual display, each menu item could also have auditory cues (depending upon the experimental condition) that played when the menu item was highlighted. When the arrow button was pushed, this triggered the auditory sound playback for the new menu item. All auditory stimuli were interruptible so that when the next item was selected, the previous auditory cue was stopped. The sounds were prerecorded as a single audio file for each menu item (with negligible loading delay). In order to maintain a code-based performance similarity between the no-sound and sound conditions, a non-audible sound file of similar playback length was played for each menu item in the no-sound condition. The auditory cues included speech (TTS) and non-speech enhanced auditory cues as described below (also see Table 1).

2.1.3.3. Text-To-Speech Cues.

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 this condition simply consisted of an auditory TTS phrase that played for each menu item as the participant navigated the song list.

2.1.3.4. Spearcon Cues.

Spearcons were created from the TTS file of each name by running them through the GT Sonification Lab's spearcon generation algorithm (Walker et al., 2012) in the form of a MATLAB script that compresses each TTS cue logarithmically while maintaining original sound frequency. Logarithmic compression is currently considered as the preferred compression technique for creating spearcons, because it compresses longer phrases more than shorter phrases. Shorter words tend to sound more like "clicks" if they are compressed too much and lose their original acoustic properties. In this condition, spearcons were played before each TTS menu item, with a 250 ms silent interval between the spearcon and the beginning of the TTS phrase.

2.1.3.5. Spindex Cues.

Spindex cues were created by generating TTS files for each letter of the alphabet. Each spindex cue consisted of only one syllable, pronouncing each of 26 letters representing the initial letter of the song names. Spindex cues used in the list were presented before the TTS cues, with a 250 ms interval between the spindex cue and the TTS phrase (Jeon and Walker, 2009; Palladino and Walker, 2008b). If participants touched an arrow button quickly, they would hear only spindex cues, without a lag between items or TTS. Since the "attenuated" spindex design has been shown to be the most preferred and simplest to implement (Jeon and Walker, 2011), we used that version in this experiment. The attenuated spindex version contained cues that were attenuated by 20 dB after the first menu item in each letter category (e.g., Aaaaa…Bbbbb…Ccccc…, assuming lower case letters' volume represents 20 dB less than that of uppercase [first item] letters).

2.1.3.6. Mixed Cues.

We also created mixed cues with combined spindex, spearcons, and TTS cues. For this, we employed the "minimal" spindex type because it has shown the same level of performance on auditory menu searches as the other spindex types (Jeon and Walker, 2011). The minimal spindex cues were used only when the user crossed category boundaries in the search list (e.g., for the first menu item starting with A, then the first item starting with B, and so on). Therefore, the spindex cues were added to only the category boundaries of the spearcon version of the auditory menu.

2.1.4. Design and Procedure

There were five within-subjects conditions, based on auditory cue type: no-sound, TTS-only, spearcon + TTS, attenuated spindex + TTS, and minimal spindex + spearcon + TTS.

After the participant gave informed consent, the experimenter explained the detailed procedure and demonstrated how to interact with the ball-catching game and the menu system on the infotainment system. Before the start of the dual tasks, participants performed the primary task alone for one minute to provide a baseline for the single task condition. The overall goal of the primary task was to catch as many falling balls as possible by moving the basket with arrow keys on the computer keyboard (Figure 1).

Participants then began the dual task portion of the study. In order to more accurately analyze the timing of both tasks, the system clocks of the computers were synchronized using a network time server. They were instructed to always allocate 80% of their effort/attention to the primary task (game) and 20% to the secondary task (navigation) (see Bonnel and Hafter, 1998). The goal of the secondary task was to reach the target song title in the song list menu as quickly as possible, and select it by touching the “select” button. The primary task was initiated, and the target name for the secondary task was presented through the speakers after a randomly selected delay of either 5, 10, or 15 seconds from the start of the primary task. The target name was also displayed visually on the first line of the list on the infotainment system (e.g., “Use Somebody” in Figure 2). After hearing the target menu item, participants navigated the list of songs on the touch screen while simultaneously maintaining performance of the visual primary task. After the selection of the target, there was another randomly selected delay of 5, 10, or 15 seconds before the next target item was presented.

Menu navigation time was operationalized as the time between the first menu navigation button press, and the pressing of the select button. In the experiment, each block included five 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 1-15 (bin 1), one from 16-30 (bin 2), and so on to 136-150 (bin 10). The order of these bins was also randomized. At the end of the block (i.e., after all five menu targets had been presented), participants saw a pop-up window on the infortainment interface and pressed the ‘Q’ key on their computer keyboard to quit the primary task. Each condition was comprised of two successive blocks. Every participant completed 2 blocks (10 trials) in each of five auditory cue conditions (in total 50 trials), which were counterbalanced across participants.

After each auditory cue condition, participants completed the electronic version of NASA- TLX (e.g., Hart, 2006) to report their perceived workload for the overall task combinations. Finally, after completing all conditions, participants filled out a short questionnaire, providing demographic information, their preferred auditory cue condition, and comments on the study.

2.2. Results of Experiment 1

2.2.1. Primary Task Performance

Table 2 shows a summary of results in Experiments 1 and 2. Figure 3 shows overall mean percentages of success on the primary task for the single task condition, and each auditory cue type in the dual-task conditions. Results were analyzed with a 5 (auditory cue type) x 2 (block) repeated-measures analysis of variance (ANOVA), which revealed a statistically significant difference between auditory cue types in mean success rate, $F(4, 92) = 8.37$, $p < .001$, $\eta_p^2 = .27$. Also, Block 2 ($M = 78.03$, $MSe = 2.25$) elicited significantly higher scores than Block 1 ($M = 75.71$, $MSe = 2.45$), $F(1, 23) = 15.74$, $p = .001$, $\eta_p^2 = .41$. The interaction of cue type with block was not significant, $F(4, 92) = 0.26$, $p = .90$. For multiple comparisons among the single task and the dual-task auditory cue types, we conducted paired-samples t-tests. For all pairwise comparisons in Experiment 1, a Bonferroni adjustment was applied to control for Type I error, which meant that more conservative alpha levels were used (for the primary task, critical alpha level = .003; for the secondary task and workload scores, critical alpha level = .005). Participants caught significantly more balls in the single task condition, and in all of the auditory-enhanced dual-task conditions than in the no-sound dual-task condition. Success rate in the single task condition ($M = 82.96$, $SD = 8.86$) was higher than that in the no-sound condition ($M = 71.01$, $SD = 10.12$), $t(23) = 7.33$, $p < .001$. For the dual-task conditions, success rates in the TTS-only ($M = 78.16$, $SD = 13.54$), $t(23) = -3.75$, $p = .001$, the spearcon + TTS ($M = 78.37$, $SD = 11.39$), $t(23) = -5.37$, $p < .001$, the spindex + TTS ($M = 78.21$, $SD = 13.10$), $t(23) = -5.51$, $p < .001$, and the spindex + spearcon + TTS conditions (*M* = 78.59, *SD* = 13.80), $t(23) = -4.054$, $p < .001$ were higher than in the no-sound condition. Primary task performance decreased in the no-sound condition relative to baseline, but performance seemed to recover to about the single task level in all sound conditions.

2.2.2. Secondary Task Performance

Errors in the secondary task were minimal, so the primary focus of the analyses for the secondary task was on the search time. For the sake of completeness, however, a one-way (auditory cue type) repeated-measures ANOVA was conducted on navigation errors, and revealed a statistically significant difference between auditory cue types, $F(2.94, 92) = 3.61$, $p < .05$, $\eta_p^2 = .14$. For the multiple comparisons among the auditory cue types, we conducted paired-samples t-tests (with Bonferroni correction, as previously indicated). The TTS-only cues (*M* = .29, *SD* = .86), $t(23) = 3.15$, $p = .004$ and the spindex + spearcon + TTS cues (*M* = .33, *SD* = .56), $t(23) = 3.20$, $p = .004$ showed significantly fewer errors than the no-sound condition (*M* = 1.17, *SD* = 1.20). The spearcon + TTS cues (*M* = .54, *SD* = .98), $t(23) = 1.871$, $p = .074$ and the spindex + TTS cues (*M* = .54, *SD* = .88), $t(23) = 1.97$, $p = .061$ showed numerical, but not significant improvements in errors over the no-sound condition for the secondary task.

We included only correct responses in search time analyses. Figure 4 shows overall mean time to target (i.e., “search time”, in seconds) in the secondary task for each of the auditory cue types. These results were also analyzed with a 5 (auditory cue type) x 2 (block) repeated-measures ANOVA, which revealed a statistically significant difference between auditory cue types in mean search time, $F(4, 92) = 3.53$, $p < .05$, $\eta_p^2 = .13$. Also, Block 2 (*M* = 29.9, *MSe* = 1.2) led to significantly shorter search times than Block 1 (*M* = 32.0, *MSe* = 1.2), $F(1, 23) = 7.91$, $p < .05$, $\eta_p^2 = .26$. For the multiple comparisons among auditory cue types, we conducted paired-samples t-tests (with Bonferroni correction). Participants searched significantly faster in the TTS-only (*M* = 28.2, *SD* = 6.8), $t(23) = 3.89$, $p = .001$ and the spindex + TTS (*M* = 28.6, *SD* = 7.3), $t(23) = 3.33$, $p = .003$ conditions than in the no-sound condition (*M* = 35.4, *SD* = 9.0). The spindex + spearcon + TTS condition (*M* = 30.9, *SD* = 7.9) also revealed numerically faster search times than the no-sound condition, although this result was not statistically significant, $t(23) = 1.92$, $p = .067$. The spearcon condition ($M = 31.7$, $SD = 2.5$) was not significantly different from the no-sound condition, $t(23) = 1.50$, $p = .15$. The interaction of block with cue type was not significant, $F(2.76, 92) = 1.17$, $p = .33$ with a Greenhouse-Geisser correction for sphericity violations.

2.2.3. Overall Perceived Workload and Preference

Figure 5 shows the overall perceived workload scores for each of the auditory cue types. All of the auditory cue types decreased the perceived workload of both tasks. These results were supported by a one-way (auditory cue type) repeated-measures ANOVA, which revealed a statistically significant difference between auditory cue types in workload score, $F(4, 92) = 14.35$, $p < .001$, $\eta_p^2 = .38$. For the multiple comparisons among auditory cue types, we conducted paired-samples t-tests (with Bonferroni). The TTS-only cues ($M = 64.12$, $SD = 14.09$) led to lower perceived workload than in the no-sound condition ($M = 75.64$, $SD = 10.90$), $t(23) = 4.33$, $p < .001$. Also, the spearcon + TTS cues ($M = 66.35$, $SD = 17.40$), $t(23) = 3.66$, $p = .001$, the spindex + TTS cues ($M = 59.65$, $SD = 13.18$), $t(23) = 6.65$, $p < .001$, and the spindex + spearcon + TTS cues ($M = 60.68$, $SD = 13.57$), $t(23) = 5.49$, $p < .001$ showed lower perceived workload than the no-sound condition. Finally, the spindex + TTS, $t(23) = 2.29$, $p = .031$ and the spindex + spearcon + TTS, $t(23) = 1.93$, $p = .066$ showed numerically, but not significantly lower perceived workload than the TTS-only condition. For the best choice of the auditory cue types, participants clearly preferred spindex + TTS ($N = 10$) and spindex + spearcon + TTS ($N = 10$) to others (no-sound, $N = 1$; TTS-only, $N = 2$; spearcon + TTS, $N = 1$) (see Figure 6).

2.3. Discussion of Experiment 1

We evaluated performance, perceived workload, and preference measures for five types of auditory presentation cues for an IVT menu navigation task in the presence of a visual perceptual-motor vigilance primary task. The results showed that the application of auditory cues for in-vehicle infotainment interfaces could improve both primary and secondary task performance and decrease overall workload. The significant performance improvements over time (i.e., from Block 1 to Block 2) for both primary and secondary task measures suggest that participants may continue to acquire skill with the system and further improve performance on both tasks using the IVT interface during a visual primary task.

In terms of the primary task, all of the auditory conditions outperformed the no-sound condition. This suggested that redundant multimodal presentation was less disruptive to performance of the primary task than visual-only presentation. Given the visually intensive nature of the primary task employed here, we expect that these results may generalize to the vehicle control scenario. Specifically, auditory cues for IVTs might allow drivers to devote more attention to the roadway than visual-only menus in IVTs, as all of the auditory cue conditions allowed primary task performance to recover to the baseline, single-task level. However, because the single task was consistently conducted before the dual task experiment, this result could be due to an order effect. Extending these results to a driving task, and considering potential order effects, was a main purpose of Experiment 2.

With respect to secondary task performance, all conditions with auditory cues reduced the mean number of secondary task errors (at statistically significant or at least marginally significant levels) as compared to the condition with no sound cues. Additionally, some auditory cues (the TTS-only and the spindex + TTS) showed significantly faster performance than the condition with no sound cues. While the spearcon + TTS and mixed cue conditions only showed numerically faster performance than the no-sound condition, the mean difference of roughly 4-5 seconds may represent a practically relevant finding.

In addition to our findings regarding performance, we found positive results that showed an overall reduction in perceived workload and also a subjective preference for enhanced auditory presentations. Participants reported lower perceived workload with auditory cues as compared to no sound cues, and enhanced auditory cues (particularly the spindex and the spindex + spearcon conditions) resulted in numerically lower perceived workload than the TTS-only. It can be inferred that lower workload in complex multitasking situations might increase the capacity for driving or other visually demanding tasks to be performed while interacting with IVT menus.

Participants also favored the spindex + TTS and the spindex + spearcon + TTS cues, despite the fact that these conditions showed equivalent levels of performance with the TTS-only condition. The intersection of performance and aesthetic preferences remains a challenge for auditory display design (Edworthy, 1998; Kramer, 1994), and the user may reject non-preferred or undesirable auditory displays even when performance measures are improved by the use of such displays. We contend that the appropriate implementation of audio in IVTs will require the consideration not only of performance consequences, but also of user preferences and perceived desirability. An audio design will be most successfully deployed when it both meets user preferences and improves performance (Jeon and Walker, 2011).

3. Experiment 2

The methods of Experiments 1 and 2 were nearly identical. The main difference was that the primary task in Experiment 2 was driving in a medium-fidelity driving simulator, rather than playing the computerized ball-catching game as in Experiment 1. This extension into the driving simulator was intended to strengthen and generalize the findings from Experiment 1, and enable us to specify the effects of adding auditory cues in terms of driving performance measures (e.g., speed, brake pedal force, steering angle, etc.).

3.1. Method

3.1.1. Participants

Thirty-seven undergraduate students participated in the study for credit in psychology courses in Spring 2011 at the Georgia Institute of Technology. Three of these participants were not able to continue the study due to simulator sickness experienced during the screening phase (described below). Accordingly, the 34 participants (22 female and 12 male; mean age = 18.6, *SD* = 4.9) who completed the study will be referred to in the following analyses. All participants had previous driving experience (mean years of driving = 4.2, *SD* = 1.6) with a valid driver's license within the United States, reported normal or corrected-to-normal vision and hearing, and gave informed consent. None had participated in Experiment 1.

3.1.2. Apparatus

Figure 7 shows the experimental apparatus. The primary task stimuli were presented using a quarter cab version of National Advanced Driving Simulator's MiniSim version 1.8.3.3. The simulator software runs on a computer running Microsoft Windows 7 Pro on an Intel Core i7 processor, 3.07 GHz and 12 GB of RAM, and presents sound through a 2.1 audio system. Three Panasonic TH-42PH2014 42" plasma monitor displays are installed in an array, allowing for 130-degree field of view in front of the seated participant with a 1280x800 resolution. The center monitor is 28 inches from the center of the steering wheel and the left and right monitors are 37 inches from the center of the steering wheel. The MiniSim also includes a real steering wheel, adjustable car seat, gear-shift, and a gas and brake pedal, as well as a Toshiba Ltd. WXGA TFT LCD monitor with a 1280x800 resolution for the instrument display.

For the secondary task, stimuli were presented using a Lilliput 7" resistive touch screen VGA infotainment system monitor with a 800x480 resolution, running Windows Vista on a Dell OptiPlex 990 with an Intel core i5-2400 at 3.10 GHz and 4 GB of RAM. The infotainment system monitor was secured at 19.5 inches from the center of the steering wheel, with the center of the screen 7 inches below the top of the steering wheel (see Figure 7). Participants listened to the auditory stimuli for the secondary task through Dell desktop A215 speakers located directly behind the center monitor of the driving simulator.

3.1.3. Stimuli

3.1.3.1. Primary Task.

The primary task was a simple driving task and was piloted to be sufficiently difficult to produce dual task decrements when the secondary menu task was introduced (as in Experiment 1). The task (see Figure 7) was programmed using the NADS ISAT software and consisted of pink "beach balls" (similar in appearance to the balls that dropped in the primary task in Experiment 1) that were randomly placed in one of four highway lanes at intervals of 300 feet. These balls could appear in either the current lane or one of the adjacent lanes; a ball could appear in the same lane at most twice in a row. The purpose of the task was to drive through each of the balls with the car, while keeping the vehicle speed between 50-60 mph. To control the vehicle, participants drove as they would in a real vehicle. We intentionally design this driving task to include two characteristics: (1) It can be closer to actual driving than Experiment 1 (external validity). (2) However, in Experiment 2 we still want to have the perceptually equivalent stimulus-response relationship to Experiment 1 (ecological validity as the original meaning) so that it can be in the same line with Experiment 1. The core task of driving can be abstracted as a lane change task, which is used as one of the standard research tools in driving research (Mattes, 2003). Our primary task is equivalent to the lane change task.

3.1.3.2. Secondary Task.

The secondary IVT menu navigation task and all of the stimuli were identical to those in Experiment 1.

3.1.4. Design and Procedure

Whereas in Experiment 1 all participants conducted a 1-minute baseline task first, in Experiment 2, participants had a single driving task (i.e., primary task only) as an additional counterbalanced driving condition, to eliminate possible confounding learning effects. Thus, the addition of the single driving task made six within-subjects conditions, based on task and auditory cue types: the baseline primary task only (i.e., driving task alone); dual task with no-sound; dual task with TTS-only; dual task with spearcon + TTS; dual task with (attenuated) spindex + TTS; and dual task with (minimal) spindex + spearcon + TTS conditions.

Before participants began the experimental task, they went through the GT Simulator Sickness Screening Protocol (Gable & Walker, 2013) that consisted of rating current physical feeling on

17 categories, using an eleven-point Likert-type scale from zero to ten (ten being the strongest feeling). The participants were then asked to drive a 2-minute "city center" scenario in the simulator. Once the drive was completed, the participants again rated their physical feeling. If the participants felt sick at any time during the drive, the simulation was stopped and they were excused from testing. Participants whose scores showed signs of simulator sickness were also excused from testing (adapted from Gianaros, Muth, Mordkoff, Levine, and Stern, 2001). Participants who did not experience simulator sickness were introduced to the next phase of the experiment and began the actual experimental task.

Each block included five trials of different targets and each condition was comprised of two successive blocks, just as in Experiment 1. The order of presentation of the conditions was counterbalanced across participants. At the end of the block (i.e., after all menu targets had been presented), the drive was ended by the experimenter and the participants completed the electronic version of NASA-TLX to provide perceived workload measurements for the overall task combinations. After finishing all conditions, participants filled out a short questionnaire for demographic information, indicated their preferred auditory cue condition, and provided comments on the study.

3.2. Results of Experiment 2

In Experiment 1, which was an exploratory study, we conducted an overall ANOVA first and then did multiple comparisons with a Bonferroni adjustment of the alpha level. However, in Experiment 2 we directly conducted *planned pairwise comparisons* using paired-samples t-tests because we already had clear hypotheses and analysis directions (Keppel & Wickens, 2004). Therefore, we maintained a traditional alpha level (.05) without applying a Bonferroni adjustment.

3.2.1. Primary Task Performance

Figure 8 shows overall mean percentages of success in the primary (ball-hitting) task for the single task and each auditory cue type. Results were analyzed with paired-samples t-tests.

Participants in the single task ($M = 99.88$, $SD = 1.04$) hit a significantly higher percentage of balls than in any other condition: the no-sound ($M = 92.49$, $SD = 13.13$), $t(33) = -3.27$, $p = .003$; the TTS-only ($M = 92.16$, $SD = 10.13$), $t(33) = -4.50$, $p < .001$; the spearcon + TTS ($M = 92.12$, $SD = 8.91$), $t(33) = -5.08$, $p < .001$; the spindex + TTS ($M = 92.67$, $SD = 8.73$), $t(33) = -4.86$, $p < .001$; and the spindex + spearcon + TTS ($M = 93.84$, $SD = 7.99$), $t(33) = -4.51$, $p < .001$. There was no other significant difference in the ball-hitting task.

For other driving performance measures, as shown in Figure 9, participants in the single task showed significantly lower standard deviations of the steering wheel angle ($M = 4.75$, $SD = 0.73$) than in any other conditions: the no-sound ($M = 5.47$, $SD = 1.09$), $t(33) = 5.23$, $p < .001$; the TTS-only ($M = 5.03$, $SD = 0.98$), $t(33) = 2.39$, $p < .05$; the spearcon + TTS ($M = 5.18$, $SD = 1.36$), $t(33) = 2.22$, $p < .05$; the spindex + TTS ($M = 5.13$, $SD = 1.15$), $t(33) = 2.70$, $p < .05$; and the spindex + spearcon + TTS ($M = 5.18$, $SD = 1.08$), $t(33) = 2.96$, $p < .01$. More importantly, all of the auditory cue types led to significant decreases of the standard deviation of the steering wheel angle compared to the no-sound type: the TTS-only, $t(33) = 3.90$, $p < .001$; the spearcon + TTS, $t(33) = 2.33$, $p < .05$; the spindex + TTS, $t(33) = 3.10$, $p < .01$; and the spindex + spearcon + TTS, $t(33) = 2.29$, $p < .05$. No other driving variable showed significant differences among conditions, including speed and brake pedal force.

In summary, adding the secondary menu task led to significantly lower scores on the primary ball-hitting driving task. However, all of the auditory menu cue types led to significantly more stable steering behavior than the no-sound condition, even though those did not recover to the single task level.

3.2.2. Secondary Task Performance

Because errors in the secondary menu search task were minimal just as in Experiment 1, we focused more on the analyses of target item selection time. Again, we included only correct responses in these time-to-target analyses. Figure 10 shows overall mean time to target in the secondary task for each of the auditory cue types. For the planned comparisons among the auditory cue types, we conducted paired-samples t-tests. In all of the auditory menu conditions, time to target was numerically faster than in the no-sound condition. Only the spindex + TTS condition ($M = 28.2$, $SD = 9.9$) was statistically faster than the no-sound condition ($M = 32.9$, $SD = 12.5$), $t(33) = 2.22$, $p < .05$. This tendency appeared similarly in Block 1 and Block 2.

3.2.3. Overall Workload and Preference

Figure 11 shows the overall perceived workload scores for each condition. For the planned pairwise comparisons among the conditions, we conducted paired-samples t-tests again. First of all, as expected, the single (driving) task ($M = 29.23$, $SD = 18.31$) showed lower perceived workload than all of the dual (driving plus menu) task conditions ($ps < .001$). Perceived workload in the no-sound dual task condition was the highest of all conditions ($M = 68.72$, $SD = 18.66$). All of the auditory cue types except the spindex + spearcon + TTS significantly decreased the perceived workload of the dual tasks, compared to the no-sound dual task: the TTS-only ($M = 58.37$, $SD = 18.85$), $t(33) = 4.89$, $p < .001$; the spearcon + TTS ($M = 64.21$, $SD = 18.54$), $t(33) = 2.15$, $p < .05$; and the spindex + TTS ($M = 57.34$, $SD = 20.24$), $t(33) = 3.68$, $p = .001$. Moreover, the TTS-only showed lower workload than the spearcon + TTS, $t(33) = -2.58$, $p < .05$ and the spindex + spearcon + TTS, $t(33) = -3.72$, $p = .001$. Likewise, the spindex + TTS also showed lower workload than the spearcon + TTS, $t(33) = 4.14$, $p < .001$ and the spindex + spearcon + TTS, $t(33) = -4.55$, $p < .001$.

For the best choice of the auditory cue types (see Figure 12), participants clearly preferred spindex + TTS ($N = 21$) to others (no-sound, $N = 4$; TTS-only, $N = 3$; spearcon + TTS, $N = 0$; spindex + spearcon + TTS, $N = 6$) similar to the findings of Experiment 1. This result was statistically supported by a chi-square test. Actual frequencies of the best choice were significantly different from the null case in which all frequencies are equal, $\chi^2(3, 34) = 25.06$, $p < .001$.

3.3. Discussion of Experiment 2

In Experiment 2, we replicated and extended Experiment 1 using a medium-fidelity driving simulator and obtained similar results. The results also supported the hypothesis that the application of auditory cues for in-vehicle infotainment interfaces would improve both driving behavior and menu navigation performance while reducing overall perceived workload.

For the primary (driving) task, even though there was no statistically reliable difference in the ball-hitting success rate, we found a clear benefit of adding auditory cues in terms of stable driving behavior. That is, the standard deviation of the steering wheel angle was significantly lower in all of the auditory conditions compared to the no-sound condition. Adding to the results of Experiment 1, this provides evidence that redundant multimodal displays were less disruptive to driving performance than visual-only displays. There are potential reasons why auditory cue conditions did not match the single task level in Experiment 2, as was the case in Experiment 1. First, in Experiment 1, the single task was always conducted first. Therefore, participants might not have had the benefit of practice or learning, and performance in the single task was relatively worse than in Experiment 2. Moreover, the (falling) ball-catching task (*M* = 82.96% in the single task) in Experiment 1 might generally be more difficult than the (driving) ball-hitting task (*M* = 99.88%) in Experiment 2. Thus, a comparatively easy primary task in Experiment 2 might not have had enough room for performance compensation by adding auditory cues. However, the results still showed that adding auditory cues to the menus in the infotainment interface can allow drivers to reliably devote more attention to steering than they could otherwise.

Regarding secondary (menu) task performance, all of the conditions with auditory cues reduced mean time to target compared to the no-sound condition, even though only the spindex + TTS condition led to the conventional level of statistical significance. More importantly, this pattern (all auditory cue conditions at least numerically reduced mean time to target compared to the no-sound condition) was repeatedly shown in Experiments 1 and 2. Taken together with the primary and secondary task results in Experiments 1 and 2, we may be closer to the generalization of the benefits of adding auditory cues to IVTs in terms of both primary and secondary tasks.

We consistently attained promising results with respect to not only task performance, but also perceived workload and subjective preference. Most of the auditory cue types (TTS-only, spearcon + TTS, and spindex + TTS) significantly reduced perceived workload compared to the no-sound condition. However, given that the workload results are not consistent between advanced auditory cues and TTS-only (see Table 2), they should be further validated. The spindex + spearcon + TTS might be considered as "too much" information (Sheridan, 2005) or too complex for drivers, who were engaging in the visually demanding primary task. Again, the lower workload that came with adding auditory cues can be expected to increase the driver's ability to allocate resources to more critical tasks. Just as in Experiment 1, participants favored the spindex + TTS ($N = 21$). Some preferred using the spindex + spearcon + TTS cues ($N = 6$). However, as we mentioned just above, the spindex + spearcon + TTS cues might increase driver workload; therefore, care is needed for practical application. These results support the use of multifaceted approaches and measures to specify subtle differences among auditory cue designs. Here, we are not merely arguing which modality is generally better between audio versus vision. Rather, we are exploring which designs are best in specific contexts among many plausible alternatives (regardless of whether it is a visual display, auditory display, or redundant display). We believe that for the successful deployment of an optimal display, this type of detailed effort is necessary, when "applying" design alternatives to the real interface as well as when iteratively "creating" design alternatives (e.g., Jeon and Walker, 2011). In addition, designers should weigh the benefits of customizable or adaptive user interfaces, considering that at least some users, albeit very few ($N = 1$ and 4 in each experiment), still prefer the vision-only interface over auditory-enhanced interfaces, despite clear performance benefits of auditory user interfaces.

4. GENERAL DISCUSSION

Meta-analyses by Wickens and colleagues (Wickens, Prinet, Hutchins, Sarter, and Sebok, 2011) recently showed that the secondary task performance might benefit from auditory presentation, but the primary task (visual vehicle control task) might not. However, our data in Experiments 1 and 2 indicated that there was no auditory cost in terms of either performance (both primary and secondary) tasks or perceived workload. In addition, it is promising that the enhanced auditory interface using spindex cues obtained a high user preference. This indicates that participants' perceptions of the utility of the auditory interfaces matched reality, which is not always the case (Andre and Wickens, 1995).

Our data suggest that participants in both experiments seemed not to preempt or interrupt performance of the visual primary task in order to accomplish the secondary task with discrete auditory cues (Horrey and Wickens, 2004; Wickens and Liu, 1988). Indeed, primary task performance was better in the redundant presentation conditions than in the visual-only condition.

Adding enhanced auditory cues such as spearcons or spindex to the plain TTS menu makes auditory presentation longer and thus, it should lead to worse primary task performance based on auditory preemption hypothesis. Our results do not support this hypothesis.

This gap could be explained by the strategies that participants used with enhanced auditory cues. As mentioned in the Introduction, users' list navigation behavior can be divided into the two stages: rough navigation and fine navigation. In the rough navigation stage, users exclude non-targets until they approach the alphabetical area that includes the target. This is because they already know the framework of alphabetic ordering and letters. Therefore, they do not need to hear the full auditory presentation of non-targets until they get to the target zone. It is sufficient for them to obtain only the information needed to decide whether they are in the target zone or not. Only once users reach the target zone do they need detailed information about each menu item to compare it with the target. In other words, in spearcon and spindex conditions, participants might not hear the full TTS in each item. Instead, they could skip over items quickly by hearing only enhanced cues, as originally intended and empirically supported (e.g., Jeon and Walker, 2011; Walker et al., 2012). An enhanced auditory menu can significantly contribute to this per-item speedup during the rough navigation. Then, the TTS phrase still supports detailed item information in the fine navigation stage. Therefore, enhanced auditory cues could even diminish plausible preemption effects when using plain TTS. If the menu items are not in alphabetical order, the benefits of the use of spindex cues might be relatively reduced. However, spearcons have shown performance improvements in 2D menu navigation (e.g., Microsoft Word-like menu with non-alphabetical order, Walker et al., 2012), and thus, advanced auditory cues could still be effective in menus with non-alphabetical order.

The reason enhanced auditory menus did not outperform TTS-only in the secondary task might be that there was no practice and there were too few trials (only 10 for each condition) for participants to become fully familiar with those new auditory cues, compared to previous studies (45-50 trials per condition in Jeon et al., 2011; Walker et al., 2012). For a more detailed analysis of list type menu navigation behavior with auditory cues, see Jeon et al. (2012).

The findings of the present research are perhaps most readily explained by the time-sharing predictions of multiple resources theory. For the no-sound condition, the primary task and the secondary task conflicted with each other in terms of both processing stage (both required motor response processes) and modality (both required focal vision) resources. We explicitly piloted and calibrated our primary visual task to be particularly demanding of visual resources, and the addition of the secondary task (which was also demanding, with the overall average time-to-target at around 31 seconds) seemed to exceed participants' capacity to effectively time-share the tasks equally across all secondary task conditions. Our primary task performance findings, in particular, suggested that supplementing the visual display of the secondary task with audio may have alleviated some of the demands on focal vision, thereby allowing for better primary task performance (as a function of lowered demands on visual resources), even when motor demands remained constant across conditions. Indeed, dual task performance is worse in many circumstances when two visual tasks must be time-shared, compared to a task configuration in which information is divided across modalities (e.g., Treisman and Davies, 1973).

We used our initial instruction about attention allocation as 80:20 (primary: secondary) ratio for the dual tasks. Even though there might be some arguments whether participants could/did allocate exactly 80% of their attentional resource to the primary task and 20% to the secondary task, psychological research on divided attention has shown that different instructions could lead to different performance results (80:20, 50:50, 20:80, etc.). See Bonnel and Hafter (1998) for more discussions on divided attention between simultaneous auditory and visual stimuli in the dual task paradigm.

5. CONCLUSION

Given that it will likely be impossible to avoid widespread implementation of in-vehicle technologies, banning IVTs may not be the best solution (Regan, Lee, and Young, 2009). Researchers need to establish human factors guidelines and ergonomic approaches so that these technologies can be used with minimal distraction and maximum driving safety (see, e.g., Nees and Walker, 2011).

Based on this rationale, our results showed that the auditory modality and enhanced auditory cues may allow users to more efficiently operate the menus of IVTs while driving safely. IVTs may be more gracefully embedded into a driving task through the application of enhanced auditory cues that can improve the performance and reduce perceived workload. As mentioned, we are well aware of the downsides of applying audio in interfaces (e.g., auditory preemption). A speech-based system might slow drivers' responses to the braking of a lead vehicle (Lee, Caven, Haake, and Brown, 2001) or compete with some of the cognitive resources needed for driving. In a follow-up study, we plan to investigate the possibility that auditory cues preempt the visual primary task, but users compensate for it using other cognitive strategies. Modeling would be a good way to precisely quantify the relationship between the use of auditory displays and visual distraction (e.g., Bi, Gan, and Liu, 2014). Designers should also consider task type and difficulty, interaction style, and driver situation when adding auditory-based interaction.

For the sake of practicality, enhanced auditory menu cues are being further evaluated using different input interfaces, such as a steering wheel-mounted controller for the menu task. Other critical issues remain to be examined, including the effects of cabin noise on IVT auditory displays in a real driving situation. The present research, however, has suggested that even basic auditory displays such as plain TTS may improve dual-task performance and reduce workload, leading to safer driving. Further, more sophisticated auditory designs, such as spindex cues, may obtain even clearer effects and also be more preferred by users of IVTs.

Acknowledgments

The authors would like to thank Centrafuse for lending the infotainment system and software and Yarden Moskovitch and Ashley Henry for gathering data.

This paper represents a compilation of results collected over several years. As is typical in such a program of research, some of the results have been discussed in preliminary form at academic conference over the course of the project. The data from Experiment 1 were discussed in preliminary form at AutomotiveUI 2009 (Jeon, Davison, Nees, Wilson, and Walker, 2009). Experiment 2 consists of entirely novel data that have never been presented or published, and is a follow-up study to Experiment 1. Therefore, discussion for the combined Experiments 1 and 2 is also new.

References

Andre, A. D., & Wickens, C. D. (1995). When users want what's not best for them. *Ergonomics in Design 3*(4), 10-14.

Ashley, S. (2001). Driving the info highway. *Scientific American 285*(4), 44-50.

Bedard, M., Parkkari, M., Weaver, B., Riendeau, J., & Dahlquist, M. (2010). Assessment of driving performance using a simulator protocol: Validity and reproducibility. *American Journal of Occupational Therapy, 64*(2), 336-340.

Bi, L., Gan, G., & Liu, Y. (2014). Using queuing network and logistic regression to model driving with a visual distraction task. *International Journal of Human-Computer interaction, 30*(1), 32-39.

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.

Bonnel, A-M.,& Hafter, E. R. (1998). Divided attention between simultaneous auditory and visual signals. *Perception & Psychophysics, 60*(2), 179-190.

Brewster, S. (1997). Using non-speech sound to overcome information overload. *Displays 17*, 179-189.

Brown, M. L., Newsome, S. L., & Glinert, E. P. (1989). An experiment into the use of auditory cues to reduce visual workload. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.*

Burnett, G. E., Joyner, S. M. (1997). An assessment of moving map and symbol-based route guidance systems. In Y. I. Noy (Ed.), *Ergonomics and safety of intelligent driver interfaces*. Lawrence Erlbaum Associates, New Jersey, pp. 115-136.

Burns, P. C., & Lansdown, T. C. (2000). E-Distraction: The challenges for safe and usable internet services in vehicles. *Proceedings of the internet forum on the safety impact of driver distraction when using in-vehicle technologies*.

Chang, W., Hwang, W., & Ji, Y. G. (2011). Haptic seat interfaces for driver information and warning systems, *International Journal of Human-Computer Interaction, 27*(12), 1119-1132.

Dingler, T., Lindsay, J., & Walker, B. N. (2008). Learnability of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech. *Proceedings of the 15th international conference on auditory display*.

Dingus, T. A., Hulse, M. C., McGehee, D. V., Manakkal, R. (1994). Driver performance results from the TravTek IVHS camera car evaluation study. *Proceedings of the annual meeting of the human factors and ergonomics society*.

Dingus, T. A., McGehee, D. V., Manakkal, N., Jahns, S. K., Carney, C., & Hankey, J. M. (1997). Human factors field evaluation of automotive headway maintenance/collision warning devices. *Human Factors 39*(2), 216-229.

Dukic, T., Hanson, L., & Falkmer, T. (2006). Effect of drivers' age and push button locations on visual time off road, steering wheel deviation and safety perception. *Ergonomics 49*(1), 78-92.

Edworthy, J. (1998). Does sound help us to work better with machines? A commentary on Rautenberg's paper 'about the importance of auditory alarms during the operation of a plant simulator'. *Interacting with Computers, 10*, 401-409.

Gable, T. M., & Walker, B. N. (2013). Driving Simulator Sickness Screening: Efficient and Effective New Protocol. *Proceedings of the Fifth International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI 2013)*. Eindhoven, The Netherlands (27-30 October).

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

Gianaros, P.J., Muth, E.R., Mordkoff, J.T., Levine, M.E., Stern, R.M., 2001. A questionnaire for the assessment of multiple dimensions of motion sickness. Aviation, Space, and Environmental Medicine 72 (2), 115-119.

Gish, K. W., Staplin, L., Stewart, J., & Perel, M. (1999). Sensory and cognitive factors affecting automotive head-up display effectiveness. *Transportation Research Record: Journal of the Transportation Research Board, 1694*(1), 10-19.

Gopher, D., Brickner, M., & Navon, D. (1982). Different difficulty manipulations interact differently with task emphasis: Evidence for multiple resources. *Journal of Experimental Psychology: Human Perception and Performance, 8*(1), 146-157.

Hart, S. G. (2006). NASA-TLX Load Index (NASA-TLX): 20 years later. *Proceedings of the Annual Meeting of Human Factors and Ergonomic Society.*

Harvey, C., Stanton, N. A., Pickering, C. A., McDonald, M., & Zheng, P. (2011). In-vehicle information systems to meet the needs of drivers. *International Journal of Human-Computer Interaction, 27*(6), 505-522.

Ho, C., & Spence, C. (2005). Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. *Journal of Experimental Psychology: Applied, 11*, 157-174.

Horrey, W. J., & Wickens, C. D. (2004). Driving and side task performance: The effects of display clutter, separation, and modality. *Human Factors, 46*(4), 611-624.

Horrey, W. J., Wickens, C. D., & Consalus, K. P. (2006). Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. *Journal of Experimental Psychology: Applied, 12*(2), 67-78.

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

Jeon, M., & Walker, B. N. (2011). Spindex (Speech Index) improves auditory menu acceptance and navigation performance. *ACM Transactions on Accessible Computing, 3*(3), 10:1-26.

Jeon, M., Walker, B. N., & Srivastava, A. (2012). “Spindex” (speech index) enhances menu navigation on touch screen devices with tapping, wheeling, and flicking gestures. *ACM Transactions on Computer-Human Interaction, 19*(2), 14:1-27.

Keppel, G., & Wicken, T. D. (2004) Design and analysis: A researcher's handbook. London: Person Prentice Hall.

Klante, P. (2004). Auditory interaction objects for mobile applications. *Proceedings of the 24th Annual ACM International Conference on Design of Communication*.

Kramer, G. (1994). An introduction to auditory display, In Kramer, G. (Ed.), *Auditory display: Sonification, audification, and auditory interfaces*. Addison Wesley, Massachusetts, pp. 1-78.

Lansdown, T. C., Brook-Carter, N., & Kersloot, T. (2004). Distraction from multiple in-vehicle secondary tasks: Vehicle performance and mental workload implications. *Ergonomics, 47*(1), 91-104.

Lee, J., Caven, B., Haake, S., & Brown, T. (2001). Speech-based interaction with in-vehicle computers: The effect of speech-based e-mail on drivers' attention to the roadway. *Human Factors, 43*(4), 631-640.

Lee, J. D., Gore, B. F., & Campbell, J. L. (1999). Display alternatives for in-vehicle warning and sign information: Message style, location, and modality. *Transportation Human Factors, 1*(4), 347-375.

Lee, J. D., Roberts, S. C., Hoffman, J. D., & Angell, L. S. (2012). Scrolling and driving: How an MP3 player and its aftermarket controller affect driving performance and visual behavior. *Human Factors, 52*(2), 250-263.

Liu, Y. C. (2001). Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveller information systems. *Ergonomics, 44*(4), 425-442.

Mattes, S. (2003). The lane change task as a tool for driver distraction evaluation. In H. Strasser, H. Rausch, and H. Bubb (Eds.), *Quality of work and products in enterprises of the future.* Stuttgart: Ergonomia Verlag.

Mattes, S., & Hallen, A. (2009). Surrogate distraction measurement techniques: The land change test. In M. A. Regan, J. D. Lee, and K. L. Young (Eds.), *Driver distraction: Theory, effects, and mitigation*, (pp. 107-122), NY: CRC Press.

Matthews, G., Sparkes, T. J., & Bygrave, H. M. (1996). Attentional overload, stress, and simulate driving performance. *Human Performance 9*(1), 77-101.

Mollenhauer, M. A., Lee, J. D., Cho, K., Hulse, M. C., & Dingus, T. A. (1994). The effects of sensory modality and information priority on in-vehicle signing and information systems. *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society.*

Nees, M. A., & Walker, B. N. (2011). Auditory displays for in-vehicle technologies. In P. DeLucia (Ed.), *Reviews of Human Factors and Ergonomics, 7*, (pp. 58-99). CA: Sage Publications.

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

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

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.*

Parkes, A., & Burnett, G. (1993). An evaluation of medium range “advance information” in route-guidance displays for use in vehicles. *Proceedings of the IEEE Vehicle Navigation & Information Systems Conference*.

Patten, C., Kircher, A., Ostlund, J., & Nilsson, L. (2004). Using mobile telephones: Cognitive workload and attention resource allocation. *Accident Analysis and Prevention, 36*(3), 341-350.

Ranney, T. A., Harbluk, J. L., & Noy, Y. I. (2005). Effects of voice technology on test track driving performance: Implications for driver distraction. *Human Factors, 47*(2), 439-454.

Regan, M. A., Lee, J. D., & Young, K. L. (Eds.). (2009). Driver distraction: Theory, effects, and mitigation. NY: CRC Press.

Schieber, F., Holtz, A., Schlorholtz, B., & McCall, R. (2008). Analysis of visual demands of in-vehicle text displays reveals an age-related increase in time needed to reallocate attention to the road. *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society.*

Seppelt, B., & Wickens, C. D, (2003). In-vehicle tasks: Effects of modality, driving relevance, and redundancy: Tech. Report AHFD-03-16/GM-03-2). Savoy, IL: University of Illinois, Aviation Human Factors Division.

Sheridan, T. B. (2005). Allocating bits to displays for dynamic control: When more is more and when more is less. *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society.*

Spence, C., & Driver, J. (1997). Audiovisual links in attention: Implications for interface design. In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics 2: Job design and product design.* (pp. 185-192). Hampshire: Ashgate Publishing.

Stanton, N. A., Booth, R. T., & Stammers, R. B. (1992). Alarms in human supervisory control: A human factors perspective. *International Journal of Computer Integrated Manufacturing, 5*(2), 81-93.

Strayer, D. L., & Drews, F. A. (2007). Cell-phone-induced driver distraction. *Current Directions in Psychological Science, 16*(3), 128-131.

Strayer, D. L., Drews, F. A., & Johnston, W. A. (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied, 9*(1), 23-32.

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science, 12*(6), 462-466.

Streeter, L., Vitello, D., & Wonsiewicz, S. (1985). How to tell people where to go: Comparing navigational aids. *International Journal of Man-Machine Studies, 22*(5), 549-562.

Suh, H., Jeon, M., & Walker, B. N. (2012). Spearcons improve navigation performance and perceived speediness in Korean auditory menus. *Proceedings of the 56th Annual Meeting of the Human Factors and Ergonomics Society*, Boston, MA.

Treisman, A., & Davies, A. (1973). Divided attention to ear and eye. *Attention and Performance, IV*, 101-117.

Tsimhoni, O., & Green, P. A. (2001). Visual demand of driving and the execution of display-intensive in-vehicle tasks. *Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society*.

Walker, B. N., & Kogan, A. (2009). Spearcons enhance performance and preference for auditory menus on a mobile phone. *Proceedings of the 5th International Conference on Universal Access in Human-Computer Interaction at HCI International 2009*.

Walker, B. N., Lindsay, J., Nance, A., Nakano, Y., Palladino, D. K., Dingler, T., & Jeon, M. (2012). Spearcons (Speech-based earcons) improve navigation performance in advanced auditory menus, *Human Factors, 55*(1), 157-182.

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

Walker, B. N., & Nees, M. A. (2012). Theory of Sonification. In T. Hermann, A. Hunt, J. Neuhoff (Eds.), *The sonification handbook*, (pp. 9-39). Berlin, Germany: Logos Publishing House.

Walker, J., Alicandri, E., Sedney, C., & Roberts, K. (1991). In-vehicle navigation devices: Effects on the safety of driver performance. McLean, VA: Performer: Federal Highway Administration, Office of Safety and Traffic Operations Research and Development.

Wickens, C. D., Prinet, J., Hutchins, S., Sarter, N., & Sebok, A. (2011). Auditory-visual redundancy in vehicle control interruptions: Two meta-analyses. *Proceedings of the 55th Annual Meeting of the Human Factors and Ergonomics Society*, Las Vegas, NV.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*(2), 159-177.

Wickens, C. D., & Liu, Y. (1988). Codes and modalities in multiple resources: A success and a qualification. *Human Factors, 30*(5), 599-616.

Wickens, C. D., & McCarley, J. S. (2008). Applied attention theory. Boca Raton: CRC Press Taylor & Francis Group.

Young, K., & Regan, M. (2007). Driver distraction: A review of the literature. In I. J. Faulks, M. Regan, M. Stevenson, J. Brown, A. Porter, and J. D. Irwin (Eds.), *Distracted driving*, (pp. 379-405). Australasian College of Road Safety, Sydney.

AUTHOR BIOGRAPHIES

Myounghoon "Philart" Jeon is an Assistant Professor in Cognitive & Learning Sciences and Computer Science at Michigan Tech. He received his Ph.D. from Georgia Tech in 2012. His Mind Music Machine Lab focuses on auditory displays, affective computing, accessible computing, aesthetic computing, and automotive user interface design research.

Thomas M. Gable is a PhD student specializing in engineering psychology in the School of Psychology at Georgia Institute of Technology. He earned a bachelor's degree in psychology from The College of Wooster. His research centers on attention and driving, including the use of multimodal interfaces to decrease driver distraction.

Benjamin Davison is a researcher at Google. He received his PhD in Interactive Computing in 2012 from Georgia Tech.

Michael A. Nees is an Assistant Professor in the Department of Psychology at Lafayette College. He was awarded a Ph.D. from the Georgia Institute of Technology in 2009. His research examines applied and theoretical aspects of auditory perception, auditory and multimodal displays, assistive technologies, and auditory displays for in-vehicle technologies.

Jeff Wilson is a research scientist in BITC and IMTC, and a PhD student in the School of Interactive Computing in the CS/HCI program at Georgia Tech.

Bruce N. Walker is an Associate Professor in Psychology and Interactive Computing at Georgia Tech. His Sonification Lab studies multimodal interfaces, including sonification and auditory displays. He teaches HCI, Sensation & Perception, and Assistive Technology, and has consulted for NASA, state and federal governments, the military, and private companies.

Table 1. Auditory cue orders for each experimental condition of the secondary task.

Condition	Auditory Cue Order (250 ms interval between)	Example
No sound	(empty sound played)	"…"
TTS-only	TTS	"All the above"
spearcon + TTS	Spearcon, TTS	"All the above" "…"(250ms) "All the above"
spindex + TTS	Spindex, TTS	"ei" "…"(250ms) "All the above"
spindex + spearcon + TTS	(Spindex), Spearcon, TTS	"ei" "…"(250ms) " All the above " "…" (250ms) "All the above"

Table 2. Summary of results in Experiments 1 and 2.

Experiment	**Measures**	**Results**
Experiment 1	Primary task performance	Single task and all auditory cuesǂ > no-sound
	Secondary task performance	- Error: TTS-onlyǂ, spindex + spearcon + TTSǂ < no-sound
		- Search time: TTS-onlyǂ, spindex + TTSǂ < no-sound
	Perceived workload	All auditory cuesǂ < no-sound
		Spindex + TTS* < TTS-only
	Preference	Spindex + TTS (10), spindex + spearcon + TTS (10)
Experiment 2	Primary task performance	- Success rate of game: Single task* > dual tasks
		- Std of steering wheel angle: Single task* > dual tasks
		- Std of steering wheel angle: All auditory cues* > no-sound
	Secondary task performance	- Search time: Spindex + TTS* < no-sound
	Perceived workload	Single task* < dual tasks
		TTS-only*, spindex + TTS*, spearcon + TTS* < no-sound
		TTS-only*, spindex + TTS* < spearcon + TTS, spindex + spearcon + TTS
	Preference	Spindex + TTS (21)*

* traditional α level (< 0.05), ǂ α level with a Bonferroni adjustment

Figure 1. View of conducting dual tasks in Experiment 1. Participants navigated a song list while playing a ball-catching game.

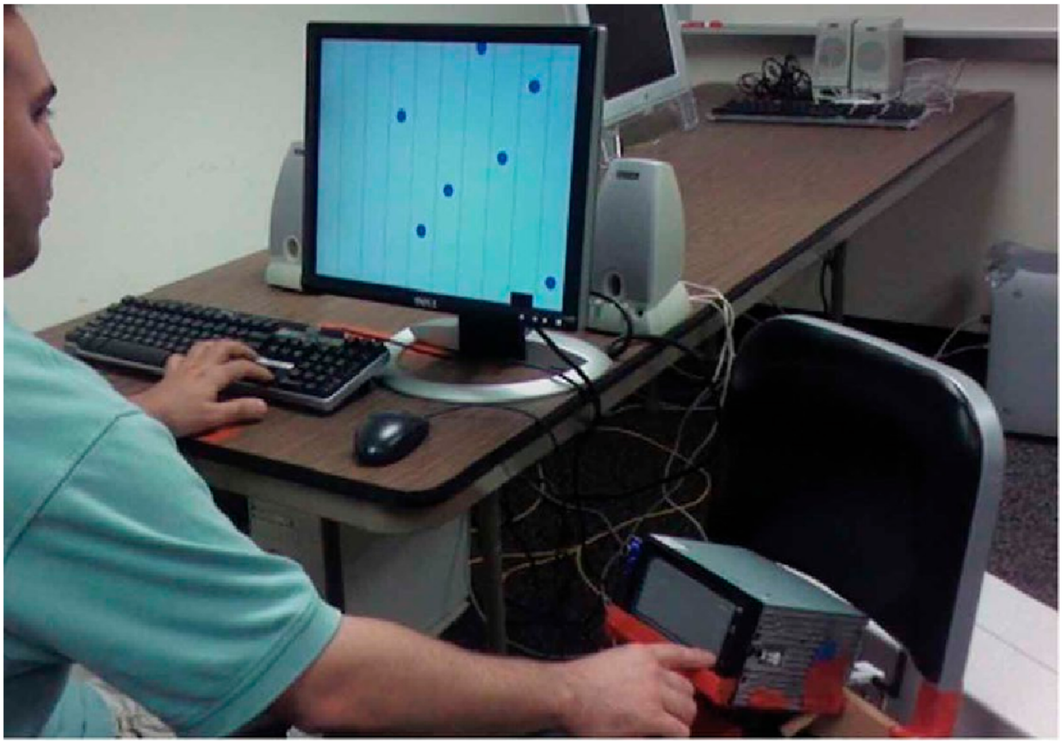

Figure 2. Screen capture of the secondary task (song list navigation) from the IVT in Experiments 1 and 2. The task was to navigate to, then select, the target menu item ("Use somebody" in this case).

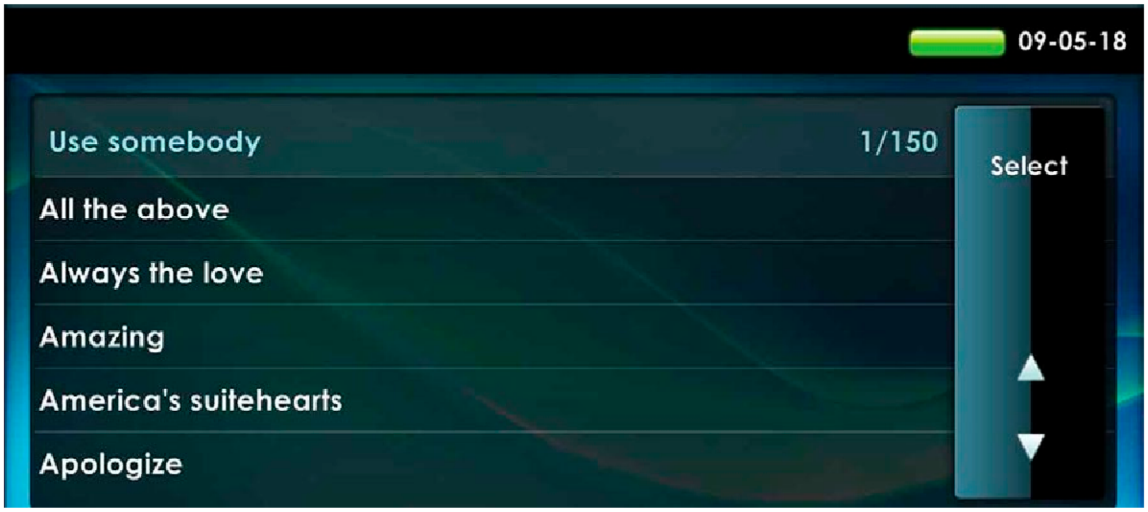

Figure 3. Primary task performance across auditory cue types in Experiment 1. Error bars indicate standard errors of the mean.

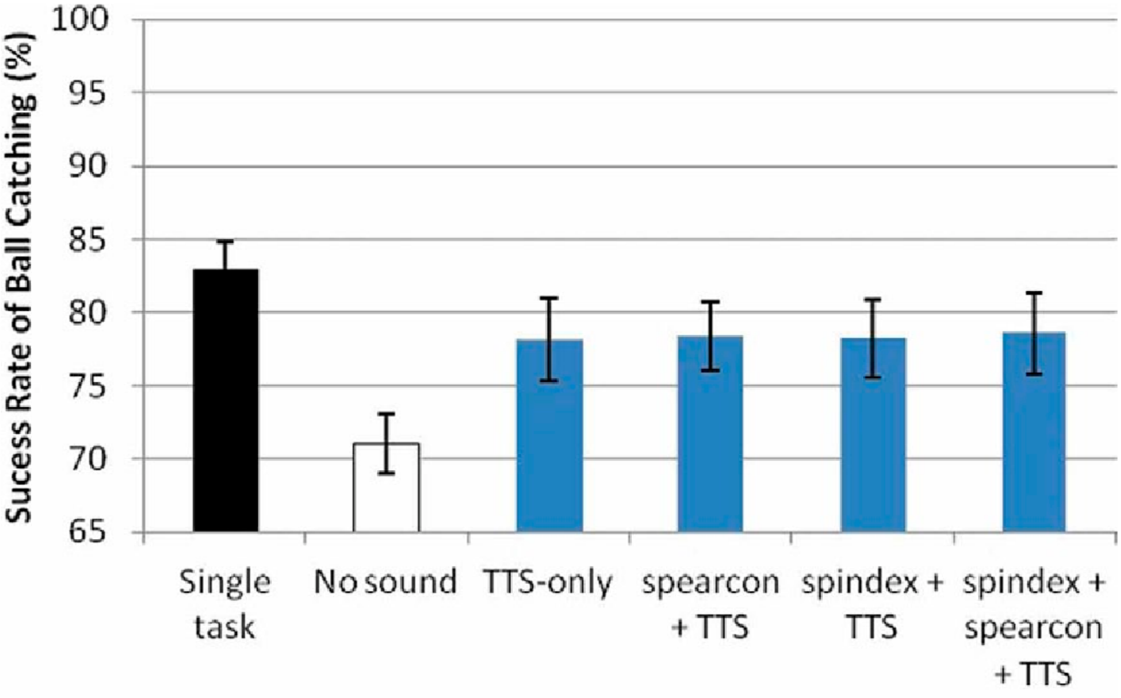

Figure 4. Secondary task performance across auditory cue types in Experiment 1. Error bars indicate standard errors of the mean.

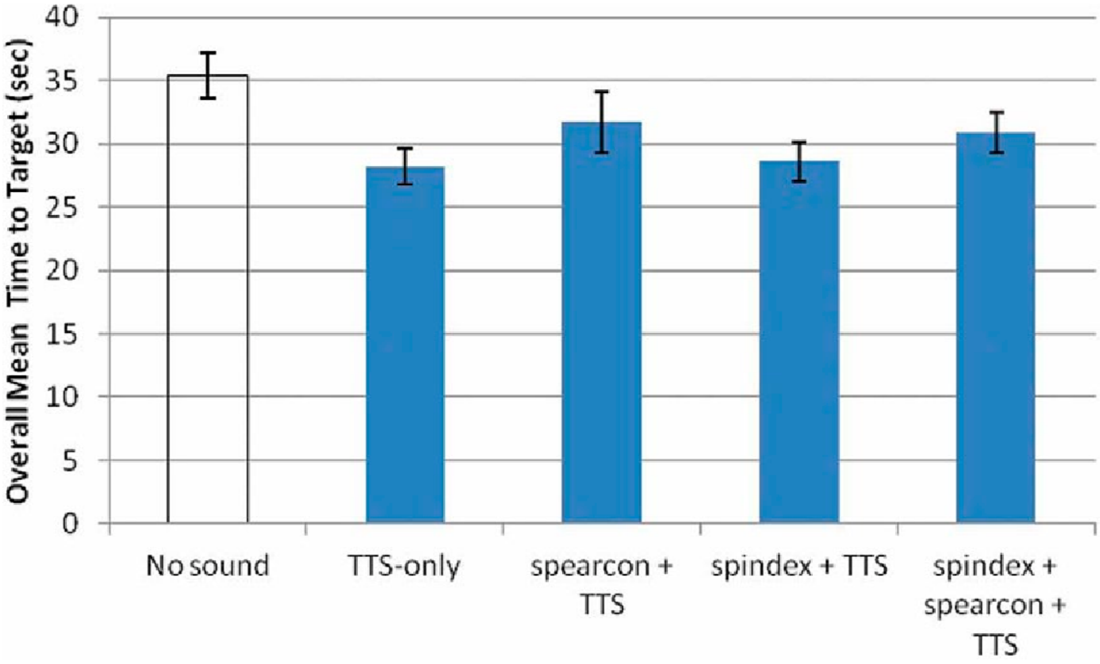

Figure 5. Overall workload score across auditory cue types in Experiment 1. Error bars indicate standard errors of the mean.

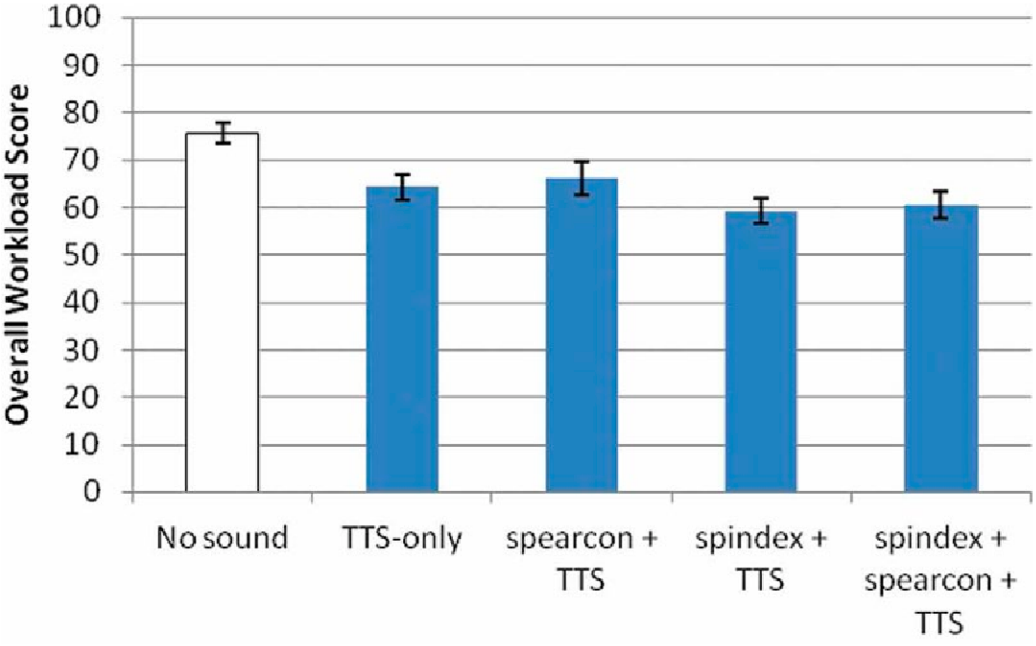

Figure 6. Overall preference across auditory cue types.

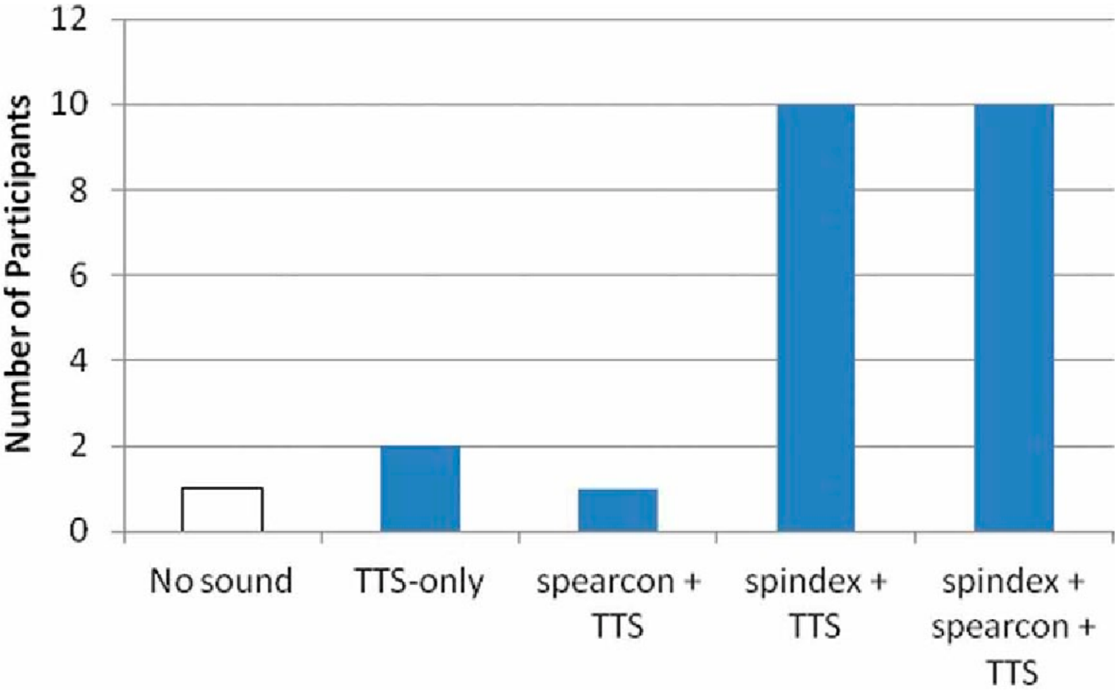

Figure 7. View of conducting dual tasks in Experiment 2. Participants navigated the same song list while driving in a simulator.

Figure 8. Primary task performance across auditory cue types in Experiment 2. Error bars indicate standard errors of the mean.

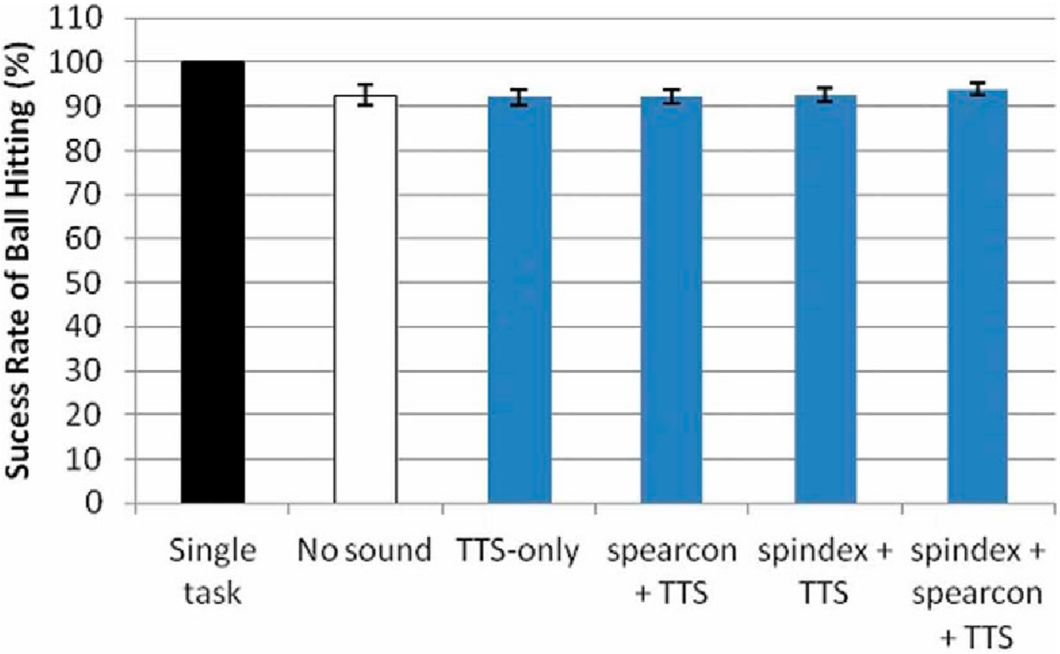

Figure 9. Mean of standard deviation of the steering wheel angle across auditory cue types in a driving task in Experiment 2. Error bars indicate standard errors of the mean.

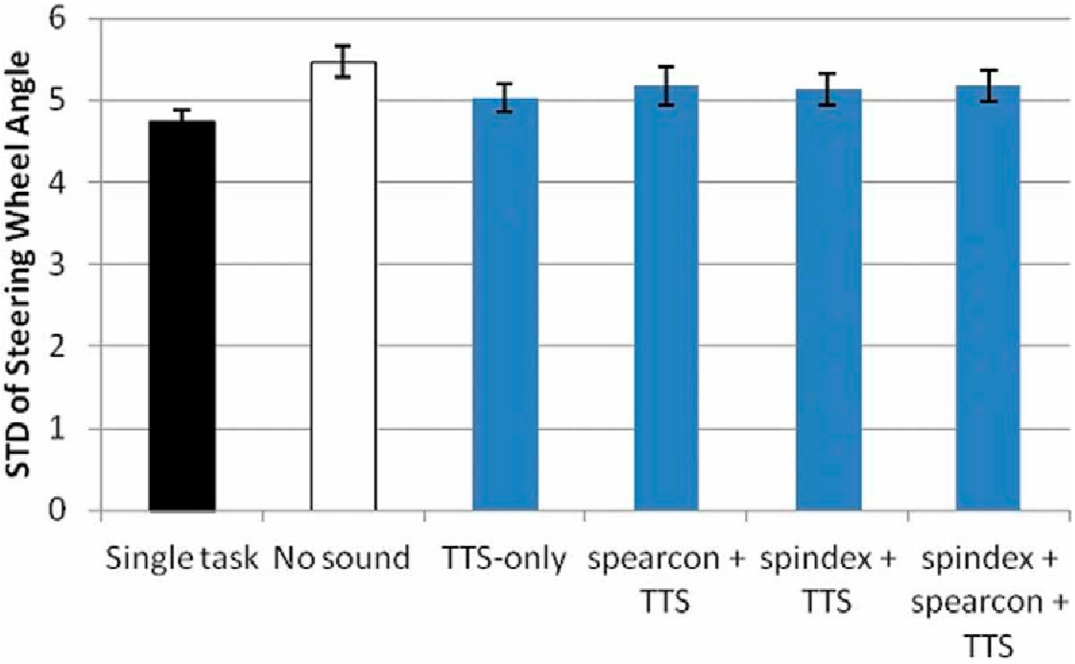

Figure 10. Secondary task performance across auditory cue types in Experiment 2. Error bars indicate standard errors of the mean.

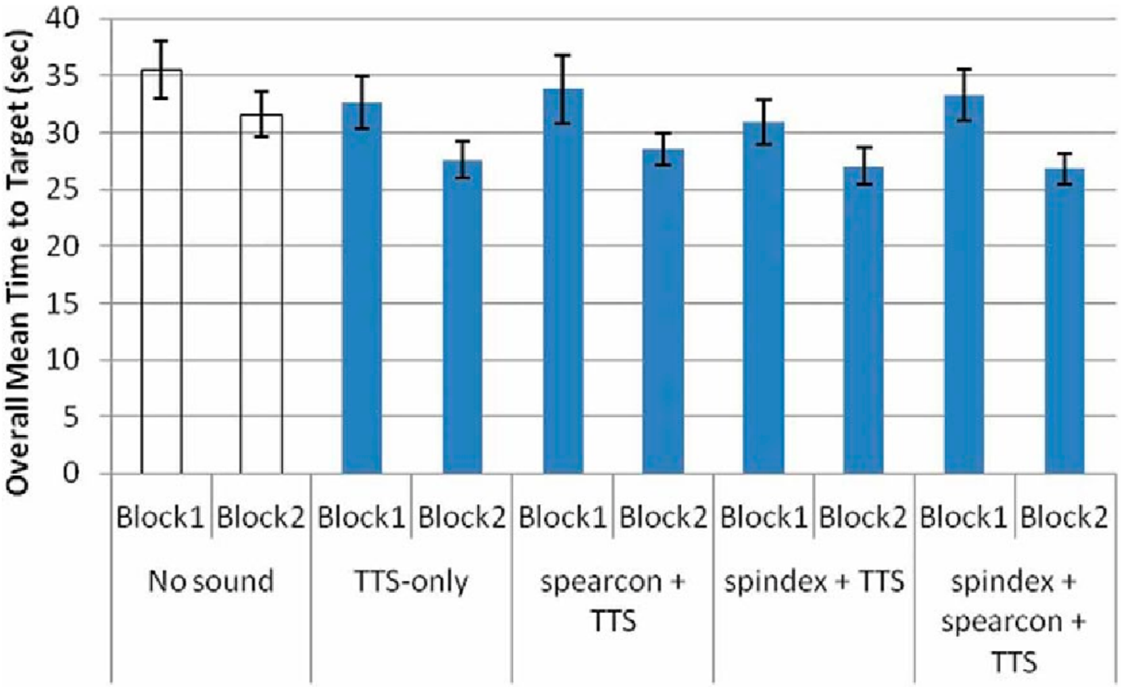

Figure 11. Overall workload score across auditory cue types in Experiment 2. Error bars indicate standard errors of the mean.

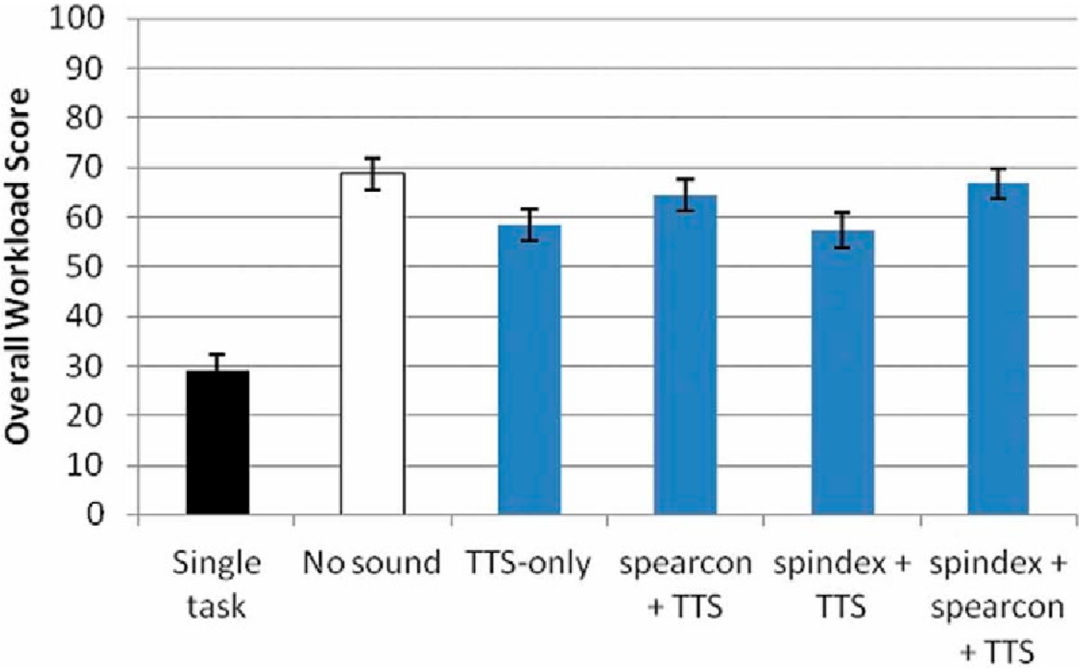

Figure 12. Overall preference across auditory cue types.