

Enhanced Auditory Menu Cues Improve Dual Task Performance and are Preferred with In-vehicle Technologies

Myounghoon Jeon, Benjamin K. Davison, Michael A. Nees, Jeff Wilson, & Bruce N. Walker

Sonification Lab, School of Psychology

Georgia Institute of Technology

654 Cherry Street

Atlanta, GA 30332 USA

+1 404 894 8265

{mh.jeon, davison, mnees}@gatech.edu, jeff.wilson@bitc.gatech.edu, bruce.walker@psych.gatech.edu

ABSTRACT

Auditory display research for driving has mainly focused on collision warning signals, and recent studies on auditory in-vehicle information presentation have examined only a limited range of tasks (e.g., cell phone operation tasks or verbal tasks such as reading digit strings). The present study used a dual task paradigm to evaluate a plausible scenario in which users navigated a song list. We applied enhanced auditory menu navigation cues, including *spearcons* (i.e., compressed speech) and a *spindex* (i.e., a speech index that used brief audio cues to communicate the user's position in a long menu list). Twenty-four undergraduates navigated through an alphabetized song list of 150 song titles—rendered as an auditory menu—while they concurrently played a simple, perceptual-motor, ball-catching game. The menu was presented with text-to-speech (TTS) alone, TTS plus one of three types of enhanced auditory cues, or no sound at all. Both performance of the primary task (success rate of the game) and the secondary task (menu search time) were better with the auditory menus than with no sound. Subjective workload scores (NASA TLX) and user preferences favored the enhanced auditory cue types. Results are discussed in terms of multiple resources theory and practical IVT design applications.

Categories and Subject Descriptors

H.5.2 [Information Interfaces And Presentation (e.g., HCI)]: User Interfaces – Auditory (non-speech) feedback, graphical user interfaces (GUI), interaction styles (e.g., commands, menus, forms, direct manipulation), user-centered design, voice I/O

H.5.1 [Information Interfaces And Presentation (e.g., HCI)]: Multimedia Information Systems – audio input/output

D.2.2 [Software Engineering]: Design Tools and Techniques – user interfaces

General Terms

Design, Experimentation, Human Factors, Performance

Keywords

auditory display, dual task, infotainment, IVTs (In-Vehicle

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

AutomotiveUI'09, September 21-22, 2009, Essen, Germany.
Copyright 2009 ACM 1-58113-000-0/00/0009...\$5.00.

Technologies), spearcon, spindex, TTS (Text-to-Speech), auditory menus, multiple resources

1. INTRODUCTION

Emerging wireless and digital technologies have allowed for an abundance of information to be delivered in mobile devices. This information portability has extended to the automobile cockpit in the form of so-called in-vehicle technologies [IVTs, see 1, 2]. IVTs can include such diverse digital media as pictures, video, and audio, and IVTs have been developed to deliver driving-relevant information (e.g., navigation instructions; weather and traffic updates), in-vehicle entertainment (e.g., digital music and video or television), and productivity applications (e.g., cellular phone and wireless internet) for the driver and passengers (see, for example, www.centrafuse.com).

Complex in-vehicle technologies may increasingly distract drivers, and research has suggested that problems of driver inattention have become worse [3-5]. A critical concern that has been validated in research involves the extent to which visually-demanding tasks like driving are prone to interference from secondary tasks such as those encouraged by IVTs. Secondary tasks have been shown to negatively affect driving performance, and subjective workload increased while driving and performing a secondary task [6, 7].

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 good practice guidelines have even been proposed [8]. 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 [9]. Given that IVTs and other secondary 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, and auditory information presentation represents an obvious alternative to visual information presentation for IVTs.

1.1 Auditory and Multimodal Presentation for IVTs

Information presentation via audio has been shown to facilitate performance with interfaces where visual overload may otherwise occur [10, 11]. Research has further suggested that auditory and multimodal IVTs may overcome some of the problems associated with visually-taxing IVTs. For example, Liu found that both driving and secondary task performance

were better using auditory, and particularly multi-modal, in-vehicle information displays [12].

Multiple resources theory [see, e.g., 13] often has 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) to the extent that they also do not require the same processing code (cognitive representational) resources, stages of processing, or response modalities (manual versus oral). Other studies [1, 14], 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, perhaps owing to the auditory modality's superior ability to attract attention [15, 16]. As such, an auditory cost has been found in a number of studies that examined the modality of in-vehicle information displays [1, 17-19]. The results from these studies suggested that the potential modality benefits of auditory (rather than visual) presentation of secondary task information might be mitigated 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 [20]. In other studies, both an auditory cost and an auditory benefit for in-vehicle information displays has been shown [21-23], while much research has shown the intuitively predicted auditory benefit for both tasks [12, 24-28].

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 that are currently being considered for implementation in an existing IVT on performance of a perceptual-motor visual primary task.

1.2 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 29] must be distinguished from the use of sound as a means of interacting with the IVT systems (i.e., "infotainment" systems). The content in 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 particular context, even though audio might improve overall performance and safety (as well as user workload, stress, and satisfaction ratings) as compared to visual menu structures.

Typically, sound is used in such menus simply by speaking 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 report 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.

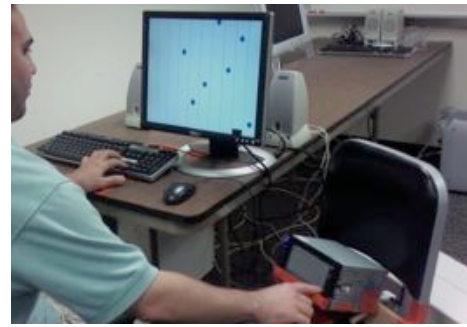


Figure 1. View of conducting dual tasks. Participants navigated a song list while playing a ball-catching game.

1.2.1 Spearcon: Compressed Speech Sounds

Spearcons (speech earcons) are brief sounds that are produced by speeding up spoken phrases, even to the point where the resulting sound is no longer comprehensible as a particular word [30]. These unique sounds are analogous to fingerprints because of their acoustic relationship with the original speech phrases. 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 the system to cope with dynamic menus. Typically, spearcons are prepended to (or may even entirely replace) the spoken menu item, which allows faster learning and navigation of the auditory menu.

Spearcons have shown better performance and learning rates than other well known nonspeech auditory cues such as auditory icons [31], earcons [32], and TTS alone. For example, Walker et al. [30] showed that spearcons resulted in faster and more accurate performance than other auditory cues for a search task. Spearcons also improved navigation efficiency over auditory menus using only TTS or no sound when combined with visual cues [33-35]. Other studies [36, 37] have demonstrated that spearcons are as learnable as speech, but auditory icons and earcons were more difficult to learn.

1.2.2 Spindex: Speech Index

A spindex [speech index, see 38] is created by associating an auditory cue with each menu item, and the cue is based on the pronunciation of the first letter of each menu item. For instance, the spindex cue for "All the above" would be a sound based on the spoken sound "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 right section of a thick reference book such as a dictionary or a telephone book, and analogous visual indices have been used, for example, in newer Apple iPods. The benefit of an auditory index (spindex) can be explained by the fact that users employ a combination of rough and fine navigation strategies in the search processes [39]. 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 reach the target zone, they need more precise and detailed information to select the target. 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 final search stage.

Spindex cues are natural sounds (based on speech) and part of the original word, thus they do not require training to learn the mapping from the sound to its intended meaning. In previous

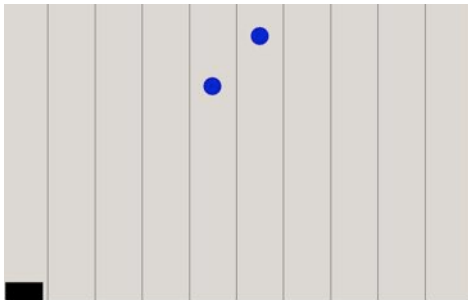


Figure 2. Screen capture of the primary task (game). Balls fall from the top of the screen, and the task is to “catch” them by moving the “bucket” (bottom left).

research, participants showed better performance in a TTS + spindex condition than in a TTS-only condition. Moreover, spindex-enhanced menus were learned more quickly, with peak performance reached in about half the number of trials, as compared to TTS-only menus. A subsequent study showed that alternative designs (*decreased* and *attenuated* types, discussed below) further improved user acceptance and performance [for more details of the spindex cue types, see 40].

1.3 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, while 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 (spearcons and spindex) may improve IVTs has yet to be established. To investigate these issues, the current study devised a plausible secondary task in which participants navigated a song list on an in-vehicle head unit. For this scenario, a divided attention paradigm [41] was used to examine the effectiveness of five types of auditory cues on performance for both a primary visual attention task (a simple ball-catching game that required perceptual vigilance and nearly constant manual control) and a concurrent secondary menu search task.

We predicted that the displays with auditory cues would shorten the navigation time in the secondary task, and also that the primary task (a visual task with perceptual and manual control components) should be less affected by the secondary task when auditory cues are 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 auditory cues, we predicted that enhanced auditory cues (i.e., those using spearcons and spindex cues) would outperform traditional TTS cues.

2. METHOD

2.1 Participants

Twenty-four undergraduate students (10 female; mean age = 20.2, $SD = 1.2$) participated in this study for credit in psychology courses. Participants reported normal or corrected-to-normal vision and hearing and gave informed consent.

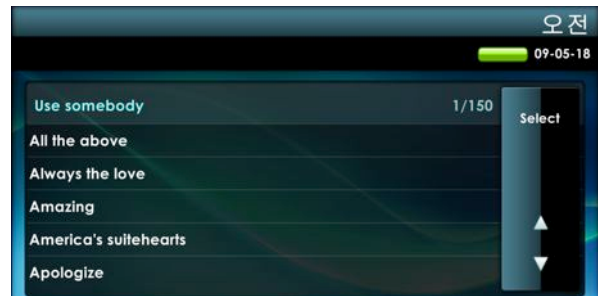


Figure 3. Screen capture of the secondary task (song list navigation) from the IVT. The task is to navigate to, then select, the target menu item (“Use somebody” in this case).

2.2 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 and 512 MB of RAM. A 17” monitor was placed on a table 50 cm in front of the seated participant. For the secondary task, stimuli were presented using an AZENTEK in-vehicle head unit, running Windows VISTA on a Pentium 4, 1.83 GHz processor and 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 behind the primary task monitor. The head unit included a 6.5” resistive touch screen panel. The head unit was located on an in-vehicle head unit position (approximately 34 cm below and 37 cm to the right from the center of the primary task monitor) [1] (See Figure 3).

2.3 Stimuli

2.3.1 Primary Task

The primary task was a visual perceptual-motor vigilance task and was piloted to be of sufficient difficulty to observe dual task decrements when the secondary menu task was introduced [for a discussion of the importance of task difficulty in dual task scenarios, see 42]. The simple computer game (see Figure 2) 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, with a black box that participants moved along the bottom of the screen. The purpose of the game was to catch all of the balls with the box before they reached the bottom of the screen. When a ball was successfully captured, the box 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% ($SD = 5.31$) accuracy for catching the balls over a 1 minute trial.

2.3.2 Secondary Task

The secondary IVT menu navigation task 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 3) was created in C# using the CentraFuse SDK programming tools for use as a plugin for the Centrafuse 2.1 head unit 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 where the next item was on a different page. The participants' objective was to reach the given target name in the list menu as fast as possible. Participants logged their selection as the current active item by pressing a "select" button (top right of figure 3). 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, the button-pressing handler triggered the auditory sound playback action. The sounds were prerecorded as a single 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 the no sound condition. The auditory cues included speech (TTS) and non-speech enhanced auditory cues as described below (also see Table 1).

2.3.2.1 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.

2.3.2.2 Spearcon Cues

Spearcons were created from the TTS files of each name by running them through the GT Sonification Lab's spearcon generation algorithm, in the form of a MATLAB script that compresses each TTS cue logarithmically while maintaining original sound frequency. Logarithmic compression is currently considered 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 prepended to each TTS menu item, with a 250 ms silent interval between the spearcon and the beginning of the TTS phrase

2.3.2.3 Spindex Cues

Since the *attenuated* spindex design has been shown to be the most preferred and simplest to implement [40], we used that version in this experiment. The attenuated version contained cues that were attenuated by -20 dB from the first menu item in a letter category. Spindex cues were created by generating TTS files for each letter (e.g., "A"). Each spindex cue consisted of only one syllable, pronouncing each of 26 letters which represented the initial letter of the 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 [33, 38]. If a participant touched an arrow button very fast, the spindex cues were generated preemptively without a lag between items.

2.3.2.4 Mixed Cues

We also created mixed cues with combined TTS, spearcons, and spindex cues. For this, we employed the *minimal* spindex type because event this showed the same level of performance on auditory menu searches as the other spindex types [40]. 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.

Condition	Auditory Cue Order (250 ms delay between)
No sound	(empty sound played)
TTS-only	TTS
Spearcon + TTS	Spearcon, TTS
Spindex + TTS	Spindex, TTS
Spindex + Spearcon + TTS	(Spindex,) Spearcon, TTS

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

2.4 Design and Procedure

Before the start of the dual tasks, participants performed the primary task alone for one minute to obtain a baseline for the single task condition. Participants then began the dual task portion of the study. In order to more accurately analyze the timing of both tasks, we synchronized the system clocks of the computers using a network time server. The primary task was initiated, and the target name for the secondary task was presented through the speakers after a delay randomly selected from 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 secondary task IVT head unit (e.g., "Use somebody" in Figure 3). After hearing the target menu item, participants navigated the list of songs on the touch screen while simultaneously playing maintaining performance of the visual primary task. They were instructed to always allocate 80% of their effort/attention to the primary task (game) and 20% to the secondary task (navigation) [see, e.g., 43]. 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. 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. One block included five trials of different targets. To evenly spread out the target menu positions across conditions, one target in each block was randomly selected from menu items 1-15, one from 16-30, and so on. Each condition was composed of two successive blocks, and the order of presentation of the cue conditions was counterbalanced across participants. At the end of the block (i.e., after all menu targets had been presented), participants saw a pop-up window and pressed the 'Q' key on their keyboard to quit the primary task. After each condition, participants completed the electronic version of NASA TLX [e.g., 44] to obtain measurements of perceived workload for the overall task combinations. Finally, after completing all conditions, participants filled out a short questionnaire for demographic information, indicated their preferred auditory cue condition, and provided comments on the study.

3. RESULTS

3.1 Primary Task Performance

Figure 4 shows overall mean percentages of success in the primary task for the single task and each auditory cue type. Results were analyzed with a 5 (Auditory cue type) x 2 (Block) repeated measures analysis of variance (ANOVA), which

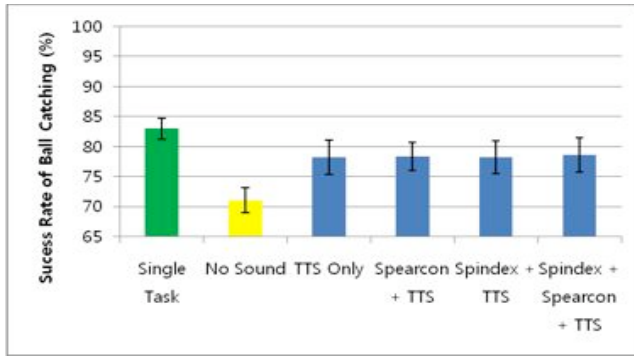


Figure 4. Primary task performance across auditory cue types.

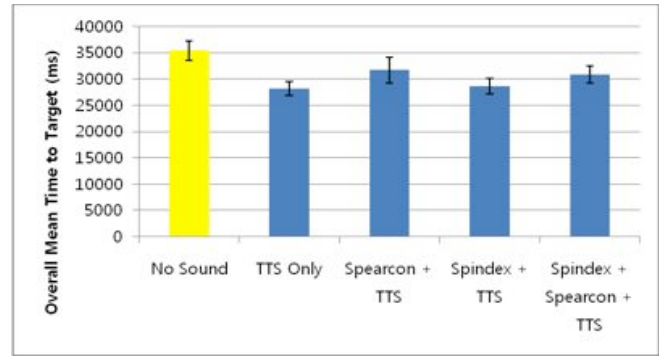


Figure 5. Secondary task performance across auditory cue types.

revealed a statistically significant difference between auditory cue types in mean success rate, $F(4, 92) = 8.372, p < .001, \eta_p^2 = .267$. Also, Block 2 ($M = 78.025, MSe = 2.248$) led to significantly higher score than Block 1 ($M = 75.711, MSe = 2.454$), $F(1, 23) = 15.737, p = .001, \eta_p^2 = .406$. The interaction of cue type with block was not significant, $F(4, 92) = 0.263, p = .901$. For the multiple comparisons among single task and the auditory cue types, we conducted paired-samples *t*-tests. Participants caught significantly more balls in the single task and all of the auditory-enhanced conditions than in the no sound condition. Success rate in the single task condition ($M = 82.96, SD = 8.86$) was higher than that in the no sound condition ($M = 71.01, SD = 10.12$), $t(23) = 7.325, p < .001^1$. Also, TTS-only ($M = 78.16, SD = 13.54$), $t(23) = -3.753, p = .001$, the spearcon + TTS cue ($M = 78.37, SD = 11.39$), $t(23) = -5.365, p < .001$, the spindex + TTS cue ($M = 78.21, SD = 13.10$), $t(23) = -5.509, p < .001$, and the spindex + spearcon + TTS cue ($M = 78.59, SD = 13.80$), $t(23) = -4.054, p < .001$ were also higher than the no sound cue. Primary task performance decreased in the no sound condition relative to baseline, but statistically performance recovered to the single task level in all sound conditions.

3.2 Secondary Task Performance

Errors (selection of non-target) in the secondary task were minimal, so the primary focus of the analyses for the secondary task was on the reaction time. For the sake of completeness, however, a one-way (Auditory cue type) repeated measures analysis of variance (ANOVA) was conducted and revealed a statistically significant differences between auditory cue types in navigation errors, $F(2.939, 92) = 3.613, p < .05, \eta_p^2 = .136$. For the multiple comparisons among the auditory cue types, we conducted paired-samples *t*-tests. The TTS-only cues ($M = .29, SD = .86$), $t(23) = 3.149, p = .004$ and the spindex + spearcon + TTS cues ($M = .33, SD = .56$), $t(23) = 3.204, p = .004$ showed significantly lower 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 ($M = .54, SD = .88$), $t(23) = 1.969, p = .061$ showed only marginally significant improvements in errors over the no sound condition for the

secondary task.

We included only correct responses in reaction time analyses. Figure 5 shows overall mean time to target (i.e., “search time”, in ms) in the secondary task for each of the auditory cue types. These results were also analyzed with a 5 (Auditory cue type) \times 2 (Block) repeated measures analysis of variance (ANOVA), which revealed a statistically significant difference between auditory cue types in mean search time, $F(4, 92) = 3.530, p < .05, \eta_p^2 = .133$. Also, Block 2 ($M = 29881.537, MSe = 1224.721$) led to significantly shorter search times than Block 1 ($M = 32036.963, MSe = 1213.727$), $F(1, 23) = 7.912, p < .05, \eta_p^2 = .256$. For the multiple comparisons among the auditory cue types, we conducted paired-samples *t*-tests. Participants searched significantly faster in TTS-only ($M = 28195.23, SD = 6791.48$), $t(23) = 3.888, p = .001$ and the spindex + TTS ($M = 28607.71, SD = 7324.10$), $t(23) = 3.330, p = .003$ conditions than in the no sound condition ($M = 35412.12, SD = 8996.31$). The spindex + spearcon + TTS ($M = 30871.16, SD = 7942.12$) also showed a tendency to be faster than the no sound, although this result was only marginally significant, $t(23) = 1.923, p = .067$. The spearcons condition ($M = 31710.03, SD = 2459.80$) was not significantly different from the no sound condition, $t(23) = 1.499, p = .147$. The interaction of block with cue type was not significant, $F(2.76, 92) = 1.167, p = .328$ with a Greenhouse-Geiser correction for sphericity violations employed.

3.3 Overall Workload and Preference

Figure 6 shows the overall 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 analysis of variance (ANOVA), which revealed a statistically significant difference between auditory cue types in workload score, $F(4, 92) = 14.348, p < .001, \eta_p^2 = .384$. For the multiple comparisons among the auditory cue types, we conducted paired-samples *t*-tests. The TTS-only cues ($M = 64.12, SD = 14.09$) showed lower workload than no sound cues ($M = 75.64, SD = 10.90$), $t(23) = 4.332, p < .001$. Also, the spearcon + TTS type ($M = 66.35, SD = 17.40$), $t(23) = 3.661, p = .001$, the spindex + TTS type ($M = 59.65, SD = 13.18$), $t(23) = 6.650, p < .001$, and the spindex + spearcon + TTS type ($M = 60.68, SD = 13.57$), $t(23) = 5.485, p < .001$ showed lower perceived workload than the no sound type. Further, the spindex + TTS, $t(23) = 2.294, p = .031$ and the spindex + spearcon + TTS, $t(23) = 1.933, p = .066$ showed marginally lower perceived workload than TTS-only. For the

¹ All pairwise comparisons in this study applied a Bonferroni adjustment to control for Type-I error, which meant that we used more conservative alpha levels (for the primary task, critical *alpha* level = .003; for the secondary task and workload scores, critical *alpha* level = .005).

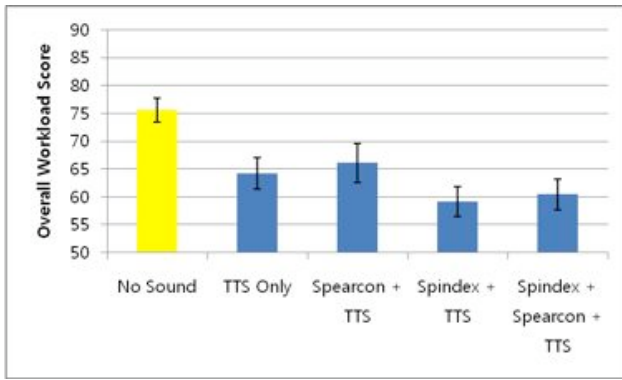


Figure 6. Overall workload score across auditory cue types.

best choice of the auditory cue types, participants clearly preferred the spindex + TTS ($N = 10$) and the spindex + spearcon + TTS ($N = 10$) to others (the no sound, $N = 1$; TTS-only, $N = 2$; the spearcon + TTS, $N = 1$) (See Figure 7).

4. DISCUSSION

We evaluated performance, 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 the auditory cues for in-vehicle head units could improve both primary and secondary task performance and ameliorate the 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 with more practice using the IVT interface during a visual primary task, although more longitudinal research will be required to examine these practice effects.

In terms of the primary task, all of the auditory conditions outperformed the no audio 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 driving scenarios. 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 recovered the primary task performance to the baseline single task level.

With respect to secondary task performance, all of the 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 (TTS-only and the spindex + TTS) showed significantly faster performance than the condition with no sound cue. While the spearcon + TTS and mixed cue conditions only showed marginally faster performance than the no sound condition, the mean difference of nearly 5 seconds may represent a practically relevant finding that would reach statistical significance with a larger sample size.

In addition to our findings with respect to performance, we found promising results that showed an overall reduction in perceived workload and also a subjective preference for enhanced auditory presentations. Participants perceived workload to be lower with auditory cues as compared to no

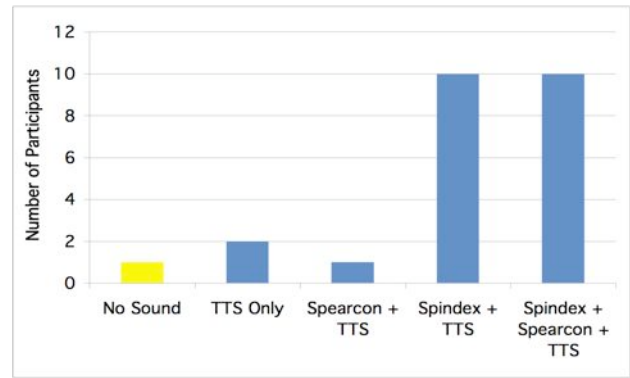


Figure 7. Overall preference across auditory cue types.

sound, and enhanced auditory cues (particularly the spindex and the spindex + spearcon conditions) resulted in lower workload than TTS-only. It can be inferred that the 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 TTS-only. The intersection of performance and aesthetics preferences remains a challenge for auditory display design [45, 46], and the user may reject non-preferred or undesirable auditory displays even when performance measures are improved by the use of such displays. We believe 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. Any audio design could be more successfully deployed to the extent that it meets user preferences and improves performance [40].

Our results do not offer evidence for a cognitive mechanism of preemption [1, 14] with respect to the effects of the discrete auditory secondary task on the continuous visual primary task. In other words, our data suggested that participants in this study did not preempt or interrupt performance of the visual primary task in order to accomplish the secondary task. Primary task performance was better in the redundant presentation condition than that in the visual only condition, so no auditory cost was observed. Moreover, the use of auditory cues seemed to contribute more to improve the primary task performance than the secondary task performance.

The findings of the present study are perhaps most readily explained by the time sharing predictions of multiple resources theory. For the no audio 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 the 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 have exceeded 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 as compared to a task configuration in which information is divided across modalities [e.g., 41].

5. CONCLUSION & FUTURE WORK

Our results that the auditory modality, and enhanced auditory cues in particular, may allow a user to more safely operate the menus of IVTs in presence of a visually-demanding primary task. 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. For a more representative primary task, enhanced auditory cues should be evaluated in a high fidelity driving simulation using a wheel remote controller for the navigation 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 auditory displays, and particularly enhanced auditory cues such as spearcons and spindex, may improve dual-task performance and also be preferred for interacting with IVTs.

6. ACKNOWLEDGMENTS

We thank CentraFuse for lending the head unit and software and Yarden Moskovitch for gathering data.

7. REFERENCES

- [1] W. J. Horrey and C. D. Wickens, "Driving and side task performance: The effects of display clutter, separation, and modality," *Human Factors*, vol. 46, p. 611-624, 2004.
- [2] W. J. Horrey, C. D. Wickens, and K. P. Consalus, "Modeling drivers' visual attention allocation while interacting with in-vehicle technologies," *Journal of Experimental Psychology: Applied*, vol. 12, pp. 67-78, 2006.
- [3] S. Ashley, "Driving the info highway," *Scientific American*, vol. 285, pp. 44-50, 2001.
- [4] T. Dukic, L. Hanson, and T. Falkmer, "Effect of drivers' age and push button locations on visual time off road, steering wheel deviation and safety perception," *Ergonomics*, vol. 49, pp. 78-92, 2006.
- [5] C. Patten, A. Kircher, J. Ostlund, and L. Nilsson, "Using mobile telephones: Cognitive workload and attention resource allocation," *Accident Analysis and Prevention*, vol. 36, pp. 341-350, 2004.
- [6] T. C. Lansdown, N. Brook-Carter, and T. Kersloot, "Distraction from multiple in-vehicle secondary tasks: Vehicle performance and mental workload implications," *Ergonomics*, vol. 47, pp. 91-104, 2004.
- [7] O. Tsimhoni and P. A. Green, "Visual demand of driving and the execution of display-intensive in-vehicle tasks," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting (HFES 2001)*. vol. 2, 2001, pp. 1586-1590.
- [8] P. C. Burns and T. C. Lansdown, "E-Distracted: The challenges for safe and usable internet services in vehicles.," in *Internet Forum on the Safety Impact of Driver Distraction When Using In-Vehicle Technologies* www.driverdistraction.org, 2000, July-August.
- [9] F. Schieber, A. Holtz, B. Schlorholtz, and R. McCall, "Analysis of visual demands of in-vehicle text displays reveals an age-related increase in time needed to reallocate attention to the road," in *Human Factors and Ergonomics Society (HFES2008)*. vol. 52, 2008, pp. 149-153.
- [10] S. Brewster, "Using non-speech sound to overcome information overload," *Displays*, vol. 17, pp. 179-189, 1997.
- [11] M. L. Brown, S. L. Newsome, and E. P. Glinert, "An experiment into the use of auditory cues to reduce visual workload," in *ACM CHI 89 Human Factors in Computing Systems Conference (CHI 89)*, 1989, pp. 339-346.
- [12] Y. C. Liu, "Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveller information systems," *Ergonomics*, vol. 44, pp. 425-442, 2001.
- [13] C. D. Wickens, "Multiple resources and performance prediction," *Theoretical Issues in Ergonomics Science*, vol. 3, pp. 159-177, 2002.
- [14] C. D. Wickens and Y. Liu, "Codes and modalities in multiple resources: A success and a qualification," *Human Factors*, vol. 30, pp. 599-616, 1988.
- [15] C. Spence and J. Driver, "Audiovisual links in attention: Implications for interface design," in *Engineering Psychology and Cognitive Ergonomics Vol. 2: Job Design and Product Design*, D. Harris, Ed. Hampshire: Ashgate Publishing, 1997, p. 185-192.
- [16] N. A. Stanton, R. T. Booth, and R. B. Stammers, "Alarms in human supervisory control: A human factors perspective," *International Journal of Computer Integrated Manufacturing*, vol. 5, pp. 81-93, 1992.
- [17] G. Matthews, T. J. Sparkes, and H. M. Bygrave, "Attentional overload, stress, and simulate driving performance," *Human Performance*, vol. 9, pp. 77-101, 1996.
- [18] J. D. Lee, B. F. Gore, and J. L. Campbell, "Display alternatives for in-vehicle warning and sign information: Message style, location, and modality," *Transportation Human Factors*, vol. 1, pp. 347-375, 1999.
- [19] T. A. Dingus, D. V. McGehee, N. Manakkal, S. K. Jahns, C. Carney, and J. M. Hankey, "Human factors field evaluation of automotive headway maintenance/collision warning devices," *Human Factors*, vol. 39, pp. 216-229, 1997.
- [20] D. L. Strayer and F. A. Drews, "Cell-phone-induced driver distraction," *Current Directions in Psychological Science*, vol. 16, pp. 128-131, 2007.
- [21] M. A. Mollenhauer, J. D. Lee, K. Cho, M. C. Hulse, and T. A. Dingus, "The effects of sensory modality and information priority on in-vehicle signing and information systems," in *Human Factors and Ergonomics Society (HFES 1994)*. vol. 38: Human Factors and Ergonomics Society, 1994, pp. 1072-1076.
- [22] A. Parkes and G. Burnett, "An evaluation of medium range "advance information" in route-guidance displays for use in vehicles," in *IEEE Vehicle Navigation & Information Systems Conference Ottawa, Canada*, 1993, pp. 238-241.
- [23] T. A. Ranney, J. L. Harbluk, and Y. I. Noy, "Effects of voice technology on test track driving performance: Implications for driver distraction," *Human Factors*, vol. 47, p. 439-454, 2005.
- [24] G. E. Burnett and S. M. Joyner, "An assessment of moving map and symbol-based route guidance systems," *Ergonomics and Safety of Intelligent Driver Interfaces*, pp. 115-136, 1997.

- [25] T. A. Dingus, M. C. Hulse, D. V. McGehee, and R. Manakkal, "Driver performance results from the TravTek IVHS camera car evaluation study," in *Human Factors and Ergonomics Society (HFES 1994)*, vol. 38: Human Factors and Ergonomics Society, 1994, pp. 1118-1122.
- [26] K. W. Gish, L. Staplin, J. Stewart, and M. Perel, "Sensory and cognitive factors affecting automotive head-up display effectiveness," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1694, pp. 10-19, 1999.
- [27] L. Streeter, D. Vitello, and S. Wonsiewicz, "How to tell people where to go: Comparing navigational aids," *International Journal of Man-Machine Studies*, vol. 22, pp. 549-562, 1985.
- [28] J. Walker, E. Alicandri, C. Sedney, and K. Roberts, "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, 1991, p. 107.
- [29] C. Ho and C. Spence, "Assessing the effectiveness of various auditory cues in capturing a driver's visual attention," *Journal of Experimental Psychology: Applied*, vol. 11, pp. 157-174, 2005.
- [30] B. N. Walker, A. Nance, and J. Lindsay, "Spearcons: Speech-based earcons improve navigation performance in auditory menus," in *International Conference on Auditory Display (ICAD2006)* London, UK, 2006, pp. 95-98.
- [31] W. W. Gaver, "Using and creating auditory icons," in *Auditory display: Sonification, audification, and auditory interfaces*, G. Kramer, Ed. MA: Addison-Wesley, 1994, pp. 417-446.
- [32] M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg, "Earcons and icons: Their structure and common design principles," *Human-Computer Interaction*, vol. 4, pp. 11-44, 1989.
- [33] D. Palladino and B. N. Walker, "Navigation efficiency of two dimensional auditory menus using spearcon enhancements," in *Annual Meeting of the Human Factors and Ergonomics Society (HFES 2008)* New York, NY (22-26 September), 2008, pp. 1262-1266.
- [34] D. Palladino and B. N. Walker, "Efficiency of spearcon-enhanced navigation of one dimensional electronic menus," in *International Conference on Auditory Display (ICAD08)* Paris, France, 2008.
- [35] B. N. Walker and A. Kogan, "Spearcons enhance performance and preference for auditory menus on a mobile phone," in *5th International Conference on Universal Access in Human-Computer Interaction (UAHCI) at HCI International 2009* San Diego, CA, USA, 2009.
- [36] T. Dingler, J. Lindsay, and B. N. Walker, "Learnability of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech," in *15th International Conference on Auditory Display (ICAD 08)* Paris, France, 2008.
- [37] D. Palladino and B. N. Walker, "Learning rates for auditory menus enhanced with spearcons versus earcons," in *International Conference on Auditory Display (ICAD2007)* Montreal, Canada, 2007, pp. 274-279.
- [38] M. Jeon and B. N. Walker, "'Spindex': Accelerated initial speech sounds improve navigation performance in auditory menus," in *Human Factors and Ergonomics Society (HFES 2009)* San Antonio, TX, 2009, p. TBD.
- [39] P. Klante, "Auditory interaction objects for mobile applications," in *24th annual ACM International Conference on Design of Communication* Myrtle Beach, SC, USA, 2004, pp. 157-163.
- [40] M. Jeon and B. N. Walker, "Optimized spindex improves both acceptance and performance with auditory menus," Georgia Institute of Technology Sonification Lab Technical Report, 2009.
- [41] A. Treisman and A. Davies, "Divided attention to ear and eye," *Attention and Performance IV*, pp. 101-117, 1973.
- [42] D. Gopher, M. Brickner, and D. Navon, "Different difficulty manipulations interact differently with task emphasis: Evidence for multiple resources," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 8, pp. 146-157, 1982.
- [43] A.-M. Bonnel and E. R. Hafter, "Divided attention between simultaneous auditory and visual signals," *Perception & Psychophysics*, vol. 60, pp. 179-190, 1998.
- [44] S. G. Hart, "NASA-Task Load Index (NASA-TLX); 20 years later," in *Human Factors and Ergonomics Society 50th Annual Meeting* San Francisco, CA, 2006, pp. 904-908.
- [45] J. Edworthy, "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*, vol. 10, pp. 401-409, 1998.
- [46] G. Kramer, "An introduction to auditory display," in *Auditory display: Sonification, audification, and auditory interfaces*, G. Kramer, Ed. MA: Addison Wesley, 1994, pp. 1-78.