Evaluation of Different Speech and Touch Interfaces to In-Vehicle Music Retrieval Systems

L. Garay-Vega[a,1], A. K. Pradhan[b], G. Weinberg[c], B. Schmidt-Nielsen[d], B. Harsham[e], Y. Shen[f], G. Divekar[g], M. Romoser[h], M. Knodler[i], D. L. Fisher[j]

[a]Department of Civil and Environmental Engineering, University of Massachusetts, 216 Marston Hall, Amherst MA 01003, USA. lisandra.umass@gmail.com

[b]Department of Mechanical and Industrial Engineering, University of Massachusetts, 210D Marston Hall, Amherst MA 01003, USA. apradhan@ecs.umass.edu

[c]Mitsubishi Electric Research Labs, 201 Broadway, Cambridge, MA 02139, USA. weinberg@merl.com

[d]Mitsubishi Electric Research Labs, 201 Broadway, Cambridge, MA 02139, USA. bent@merl.com

[e]Mitsubishi Electric Research Labs, 201 Broadway, Cambridge, MA 02139, USA. harsham@merl.com

[f]Department of Mechanical and Industrial Engineering, University of Massachusetts, Amherst MA 01003, USA. yans@engin.umass.edu

[g]Department of Mechanical and Industrial Engineering, University of Massachusetts, 210D Marston Hall, Amherst MA 01003, USA. gdivekar@engin.umass.edu

[h]Department of Mechanical and Industrial Engineering, University of Massachusetts, 210D Marston Hall, Amherst MA 01003, USA. mromoser@ecs.umass.edu

[i]Department of Civil and Environmental Engineering, University of Massachusetts, 216 Marston Hall, Amherst MA 01003, USA. mknodler@ecs.umass.edu

[j]Department of Mechanical and Industrial Engineering, University of Massachusetts, 210D Marston Hall, Amherst MA 01003, USA. fisher@ecs.umass.edu

Correspondence should be addressed to:
Dr. Lisandra Garay-Vega
55 Broadway Cambridge, MA 02142, USA
Voice: 617-494-2808 Fax: 617-494-3622
lisandra.garay-vega@dot.gov

---

[1] Now at Volpe National Transportation Systems Center, 55 Broadway Cambridge, MA 02142.

Abstract

In-vehicle music retrieval systems are becoming more and more popular. Previous studies have shown that they pose a real hazard to drivers when the interface is a tactile one which requires multiple entries and a combination of manual control and visual feedback. Voice interfaces exist as an alternative. Such interfaces can require either multiple or single conversational turns. In this study, each of 17 participants between the ages of 18 and 30 years old was asked to use three different music-retrieval systems (one with a multiple entry touch interface, the iPod™, one with a multiple turn voice interface, interface B, and one with a single turn voice interface, interface C) while driving through a virtual world. Measures of secondary task performance, eye behavior, vehicle control, and workload were recorded. When compared with the touch interface, the voice interfaces reduced the total time drivers spent with their eyes off the forward roadway, especially in prolonged glances, as well as both the total number of glances away from the forward roadway and the perceived workload. Furthermore, when compared with driving without a secondary task, both voice interfaces did not significantly impact hazard anticipation, the frequency of long glances away from the forward roadway, or vehicle control. The multiple turn voice interface (B) significantly increased both the time it took drivers to complete the task and the workload. The implications for interface design and their relative safety merits are discussed.

Keywords:
iPod™, driving simulator, distraction, eye movement, attention, music retrieval systems

1. Introduction

Distraction has long been recognized as a major contributor to automobile crashes among all drivers (Wang, Knipling, and Goodman, 1996). The magnitude of the problem is likely to increase because of the growing popularity of in-vehicle tasks that require the driver to glance away from the forward roadway – most notably music retrieval operations (Chisholm, Caird and Lockhart, 2008; Salvucci, Markley, Zuber, Brumby, 2007) and text messaging with cell phones (Lerner and Boyd, 2004; Strayer, Drews and Crouch, 2003).

Two recent studies on driving simulators point directly to the impact on driving performance of interacting with an iPod. In the first study (Salvucci et al., 2007), 17 drivers (no information on age was available) navigated the virtual roadway while selecting various media (music, podcast or video) on an iPod that was placed in a hands-free device holder. Each new request for an iPod task was made 30 s after the driver had completed the previous task, allowing for 30 s of control driving between secondary tasks. Our interest here is primarily in the song selection results. Selecting a song took an average of almost 32 s to complete. Furthermore, when drivers were selecting a song, the lateral deviation around lane center was larger than baseline. Involvement in a secondary task for such a long period of time is a clear threat to safe driving (Green, 1999), as are increases in the lateral deviation around lane center.

In the second study (Chisholm et al., 2008), drivers between the ages of 18 and 22 (mean 19.1) were asked to navigate through a virtual world in which, among other things, a lead car braked suddenly, a pedestrian entered the roadway unexpectedly, and a vehicle

pulled out into the roadway without warning. Throughout six different sessions on the driving simulator the participants were asked to interact with an iPod, performing both easy (2-3 steps, e.g., turning off the iPod) and difficult (5-7 steps) retrieval tasks. Eye movements were monitored throughout. Task completion times for the difficult iPod tasks did not differ from one another in the last three sessions and hovered around 28 s. Task completion times for the easy iPod tasks were much faster and settled around 4 s. Perhaps not surprisingly, drivers engaged in a difficult iPod task performed less safely than did drivers engaged in no secondary task on all other measures: perception response times to the various events were longer; more collisions were recorded; larger variation in the steering wheel angle was measured; more glances inside the vehicle were required; and average glance durations inside the vehicle were longer.

There are a number of measures that have been proposed as indices of distraction, most notably those promulgated in the ISO standards for measuring visual behavior (ISO, 2002) and visual demand (ISO, 2007). Perhaps the most common one is the task completion time where tasks that take longer than 15 s to perform are considered unsafe (Green, 1999). More recently, a number of researchers have argued that distraction can be indexed by its impact on hazard anticipation and response (e.g., Chisholm et al., 2008). Arguably, the single most important predictor of crashes due to in-vehicle distractions such as the above is the existence of long glances away from the forward roadway during the performance of the in-vehicle task (Dingus et al., 1989; Green, 2007; Klauer et al., 2006). For example, in a naturalist study of drivers, Klauer et al. estimated that glances away from the forward roadway for more than 2.0 seconds were the cause of more than 23% of the crashes and near crashes (shorter glances were not associated with

a significant increase in crash risk). Other support for this argument is provided in the Chisholm et al. (2008) simulator study where the number of collisions increased with the duration of the glances inside the vehicle. (A *glance* is a single fixation or a sequence of fixations which begins when the driver first looks away from the forward roadway at the distracting task and ends when the driver looks back at the forward roadway.) Additionally, this position is consistent with a simulator study reported by Horrey and Wickens (2007) in which glances 1.6 s or longer inside the vehicle, while constituting only a relatively small fraction of the total glances (22%), are responsible for the great majority of crashes (86%).

Unfortunately, the sorts of in-vehicle tasks such as interacting with an iPod or texting while driving which are dangerous in general are just the sorts of tasks which are most popular with younger drivers who, for the most part, are much more easily distracted than older drivers. The evidence that distraction poses a significant problem for novice drivers comes from many different sources, including police crash reports (McKnight and McKnight, 2003; Wang, Knipling and Goodman, 1996), naturalistic studies (Klauer et al., 2006), field experiments (Wikman, Nieminen and Summala, 1998; Lee, Olsen and Simons-Morton, 2006), simulator studies (Chan, Pradhan, Pollatsek, Knodler and Fisher, 2008; Greenberg et al., 2003; Horrey and Wickens, 2007), and surveys of novice drivers (Olsen, Lerner, Perel and Simons-Morton, 2005). Moreover, younger drivers are much more likely while performing an in-vehicle task to engage in the very behaviors which are most likely to lead to crashes, i.e., prolonged glances inside the vehicle (Wikman et al., 1998; Chan et al., 2008).

In summary, several studies indicate that interactions with an iPod lead drivers to engage in risky behaviors such as glancing inside the vehicle for prolonged periods of time, spending too long on the task, and losing awareness of the surrounding environment by becoming over engaged in the music retrieval task. This is particularly troublesome given that the very individuals most likely to glance for extended periods of time inside the vehicle – younger adults -- are also the ones most likely to be interacting with an iPod or texting while driving. Assuming that music retrieval systems continue to be accessible inside the cabin of an automobile, one would like to design an interface to these systems which reduced both the relative frequency of prolonged glances inside the vehicle and the task completion time and which had a minimal effect on hazard anticipation.

To do such, one needs to consider the demands that the iPod makes on the visual, voice, motor and memory systems of the users and the effect that these demands have on eye glance behaviors, task completion time, and hazard anticipation. To begin, consider the demands. Specifically, the iPod (Interface A) requires multiple entries, makes demands on both the motor and visual systems (in order to see on which menu option the cursor is positioned), and requires some attentional resources (in order to navigate the menu hierarchy). There are several obvious ways in which these demands could influence these measures of distractions. First, multiple entries will lead to long task completion times. Second, the demands on the motor and visual systems will increase the frequency and duration of glances away from the forward roadway. And third, the demands on attentional resources may influence hazard anticipation.

Perhaps the most obvious way to modify the touch interface is to use voice commands to enter the request. There are two types of voice interfaces one might

consider evaluating. One is a multiple conversational turn voice interface (Interface B) that prompts the user at each step to enter a command appropriate to the level in the menu hierarchy (this style of voice interface is most typical of currently available commercial offerings because of the serious limitations of automatic speech recognition systems). The second, much less common type is a single turn voice interface (Interface C) that allows the user to give the entire request in one command. One would predict that when compared with an interface which made touch and visual demands (e.g., the iPod), the two voice interfaces would reduce the number of glances away from the forward roadway and the frequency of especially long glances because visual feedback was not necessary. Additionally, one would predict that the single turn interface (Interface C) had an average task completion time that was shorter than both multiple entry/turn interfaces (touch and voice). Finally, one would predict that all interfaces interfered with hazard anticipation since they make demands on the attentional resources of the drivers which could otherwise be used to scan the roadway and predict potential threats. Below, we compare the effect on various measures of safe driving performance (glance durations, task completion time, hazard anticipation, among others) of two different voice interfaces (single and multiple turn) and the standard iPod touch interface (multiple entry).

2. Method

Each participant had to navigate eight different four minute drives in a driving simulator, two control drives with no secondary task and six drives for the different music retrieval systems. The drives with the music retrieval system consisted of two with an iPod (with touch interface), two with voice interface B, and two with voice interface C. Music retrieval tasks were selected randomly from the music library. Task initiation

and completion times were recorded, with the number of retrieval attempts, and the task outcome (i.e., success or failure). Eye behaviors, vehicle behaviors, task performance and workload were monitored throughout the experiment.

2.1. Participants

Seventeen native English speakers (12 men and 5 women) participated in the experiment (the speech recognition systems used in this experiment are optimized for native speakers of English). Participant ages ranged from 18 to 30 years old, with an average age of 21.53 years (SD=2.93). In order to be included in the study, participants had to have owned an iPod (iPod Shuffle excluded) for at least a month and used it two or more times per week in the last month of ownership. Participants also had to have at least one full year of driving experience. Individuals were recruited from the student and staff population at the University of Massachusetts at Amherst and the surrounding community. Participants received $25 for their participation.

2.2. Music Retrieval Systems and Music Retrieval Task

Each of the music-retrieval systems contained a standard set of music, a collection of 101 albums in a variety of genres. This standard set of music was used across all participants to reduce confounds that could mask the effects of each interface on driving performance (e.g. variability in the size or content of an individual participant's personal music collection). This approach also eliminated complex logistics that would have been involved in preparing participants' personal iPods for use in the experiment. (However, it did mean that all participants were equally unfamiliar with the iPod contents, a situation which is somewhat different from the usual case). Music retrieval tasks consisted of finding a specified item from the music collection. There were three types of

retrieval tasks, corresponding to three very common use cases: Song tasks, to find a specified song; Album tasks, to find any song from a specified album; and Artist tasks, to find any song by a specified artist. Each of the three systems is described in detail below (the critical elements are also listed in Table 1). These correspond to the one touch and two voice systems described above.

Multiple Entry Touch Interface A. The iPod itself serves as the baseline system. The participant selects music by navigating the iPod's hierarchical menu structure (by means of the touch-sensitive "click wheel") and then pressing the "play" hardware button. A schematic of portions of the interface are displayed in Figure 1 below. The iPod was mounted near the dashboard in a dedicated holder; however users were allowed to hold it during retrieval tasks if they desired.

Multiple Turn Voice Interface B. A commercially available aftermarket in-dash navigation and entertainment unit (Pioneer AVIC-Z2) constitutes multiple turn voice interface B. The unit offers a "Music Library" mode that combines touch and speech input modalities for the retrieval and playback of individual albums, artists and songs. The unit's speech interface is *stateful* (i.e., only a subset of speech commands are valid in particular system/screen states) and *system paced* (i.e., the user pressed the push-to-talk button once to initiate a dialog; further user turns within the same dialog are initiated by the system). These interface aspects are typical of most current commercial offerings, which motivated our selection of this particular unit. The unit's Music Library screen presents a hierarchical album, artist, playlist, and song selection interface relatively similar to the iPod's own interface, with the notable exception that albums are shown in the order in which they were copied to the Music Library from CD, rather than in

alphabetical order. One initiates a speech dialog and navigates playlists using a steering wheel-mounted touch input device. Playlist navigation is necessary only when the speech recognition system does not provide a correct match.

Single Turn Voice Interface C. The Mitsubishi Electric Research Laboratories (MERL) "SpeakPod" prototype serves as the third system (Weinberg and Kondili, 2008). Its voice interface, in contrast to voice interface B, does not require the subject to enter into a context-sensitive dialog or more generally to remember particular commands. Instead, the participant requests music by using descriptive words in any order, much like using an Internet search engine. The best match for the requested music begins playing back right away, and a list of matching music is displayed on the screen. Alternate matches can be accessed using a steering wheel-mounted touch input device which incorporates a clickable joystick and auxiliary buttons. The browsing of such matches takes place using a hierarchical menu that closely resembles the iPod's own menu.

A single dashboard-mounted LCD was used as the display for voice interfaces B and C. Its viewable area measured approximately 7 inches diagonally.

2.3. Driving Simulator and Scenarios

2.3.1 Driving Simulator

The fixed-based driving simulator in the Human Performance Laboratory at the University of Massachusetts at Amherst was used in this experiment. The simulator, manufactured by Illusion Technologies, Inc., consists of a full size Saturn sedan. The scene is displayed on three screens which subtend 150 degrees of vision in the horizontal direction and 30 degrees in the vertical direction. The images can be displayed with a

resolution as high as 1024 X 768 pixels in each screen with a refresh rate of 60 Hz. The simulator also employs a surround sound audio system.

2.3.2 Virtual Drives

Eight virtual drives were constructed, four drives through suburban roads and four through city streets (see Figure 2). Four scenarios were embedded in each of the various suburban drives (for a total of 16 different scenarios) which were used to identify the effects of the different in-vehicle systems on drivers' tactical hazard anticipation skills (Fisher, Laurie, Glaser, Connerney, Pollatsek, Duffy, and Brock, 2002; Pollatsek, Narayanaan, Pradhan, and Fisher, 2006).

2.4. Eye Tracker

The Mobile Eye, a lightweight tetherless eye tracker system from Applied Science Laboratories (ASL), was used to monitor eye movements of the driver. The system uses pupil-corneal reflection as the measurement principle. The sampling and output rate is 30 Hz and the system allows the driver's head a full range of motion. The visual range is 50 degrees in the horizontal direction and 40 degrees in the vertical direction. The system's accuracy is 0.5 degrees of visual angle. The system converts eye position to external point of gaze by superimposing crosshairs on a video of the scene that is being viewed by the participant.

2.5. Experimental Design

As noted above, each participant navigated eight drives. The eight drives were partitioned into four different blocks, each block consisting of two different drives (one suburban and one city drive). In three of the four blocks, the participant engaged in a secondary music retrieval task using in each block a different one of the three different

music retrieval systems. In the remaining block, no secondary task was introduced. A different ordering of the blocks was used for each participant (i.e., each participant was exposed to the different music retrieval systems and control condition in a different sequence).

2.6. Procedure

Participants completed a calibration procedure for the eye tracker and a practice drive designed to make them comfortable with the simulated driving environment and the vehicle itself. Each participant then completed the four blocks of two drives each, changing interfaces between blocks (or simply doing a control block of drives). In each drive participants were asked to drive normally while following a lead vehicle (a black SUV) to an unknown destination. Prior to each block, a random list of retrieval tasks was automatically generated, with an equal likelihood of song, album, and artist retrieval tasks. Participants were prompted to perform retrieval tasks (sequentially from the list) throughout the entirety of each block except for the control block. The experimenter used a standardized format for specifying the music retrieval task for all systems, as shown by the following examples: "Find the artist Sarah McLachlan." "Find the album 'Fumbling Towards Ecstasy' by Sarah McLachlan." "Find the song 'Ice Cream' by Sarah McLachlan."

At the beginning of each block (exclusive of the control block), the experimenter demonstrated the procedure by which one retrieves music using either of the voice interfaces (B or C) or reviewed the iPod's hierarchical touch interface. The target items for this demonstration varied from system to system, but were the same for each participant. This in-vehicle training was completed with the driving simulation not yet

running. Each participant had at least one practice task for each item type (i.e., song, artist, album) with each system. In the case of voice interface C, the experimenter explained and demonstrated how multiple phrasings can be used to retrieve the same music (e.g., "Revolver The Beatles" and "The Beatles Revolver"). During the experiment, once a song was retrieved, the participant was given 10 seconds "reward" listening, after which the next task was initiated.

For each task in each drive of each block, the experimenter used a device to mark the initiation and completion of the task as well as to flag a task as a success or failure (the same rules applied to all interfaces). Artist tasks were considered successfully completed when any song by the given artist started playing. Album tasks, similarly, were considered successful when any song from the album in question started playing. Song tasks were considered successfully completed only when the actual song requested was played by the system. A task was considered unsuccessfully completed when the participant verbally indicated he/she had finished trying to find the music in question, or when 120 seconds elapsed, whichever came first. Multiple attempts were acceptable within the 120-second time limit. Drivers were informed when they reached the two-minute time limit, at which point a new task was given.

At the end of each block, participants completed a workload questionnaire for each condition (using a laptop computer). Instructions for the workload questionnaire were provided once at the start of the first session and then available throughout the experiment.

The experimental procedure lasted approximately 1.5 hours. Simulator fatigue was mitigated by including a brief break at the end of the second experimental block. The

break offered the participant (and experimenter) a chance to get out of the car, walk or go to the restroom if needed and or rest his or her eyes.

2.7. Dependent Variables

The dependent measures were categorized into five main groups: (1) task completion time; (2) eye behavior; (3) hazard anticipation; (4) lane deviation; and (5) a subjective measure of drivers' workload (NASA TLX questionnaire). Measures of drivers' eye behavior included the glance duration on the interface itself (not the speedometer, say), as well as a measure of tactical scanning (whether or not the driver looked at a hazard). Voice interface B has functionality beyond the music retrieval domain. Prior to performing music retrieval, the user had to activate the "music search" state. In order to record comparable data with all three systems, the activation of the music search state was not included in the measure of task time, i.e., the task timer for voice interface B did not start until the user had entered the "music search" state.

3. Results and Discussion

3.1. Music Retrieval Task Performance

The task completion time (the average time it took participants to complete a task successfully) for, respectively, the iPod, voice interface B and voice interface C systems was, respectively 39 (SD = 17.3), 47 (SD=13.8) and 25 (SD=12.6) seconds (this does not include the reward time). There was a main effect of condition [$F(2,23.269)=8.777$, $p<0.01$, using the Greenhouse-Geisser correction for sphericity]. As predicted, the multiple entry/turn touch (A) and voice (B) interfaces were associated with longer task completion times than the single turn voice interface (C). Post hoc paired t-tests indicated that the task completion time for voice interface B was significantly longer than

for voice interface C [t(16)=4.692, p<.001]. The results suggest some benefits of a single turn voice interface over the multiple turn stateful and system-paced speech interface. Still, the time on average that it takes to complete a task using even the best voice system (C) is longer than is recommended (15 s, Blanco et al., 2005; Green, 1999).

3.2. Glance Durations at Interface

The total eyes off the road time per task was computed for each participant. This analysis includes all tasks, both successfully completed and unsuccessful tasks. However, it does not include the 10 s reward period after successful completion of a task. On average the total time a participant spent with his/her eyes off the road for the iPod, voice interface B, and voice interface C interfaces was, respectively, 17.3 (SD =11.8), 13.3 (SD=18.4), and 10.7 (SD=12.8) seconds per task. (Note that there was one participant who had spent an extremely long time with his or her eyes off the road for all three interfaces – over one minute. Removing this participant from the analysis, one gets standard deviations for the three interfaces of, respectively, 4.2, 8.3 and 5.0.) A repeated measures ANOVA indicated that these differences were significant [F(1.57,25.11)=4.413, p<0.03, using the Greenhouse-Geisser correction for sphericity]. Post hoc paired t-tests indicated that participants spent more total time in each task with their eyes off the road using the iPod than they did voice interface C [t(16)=4.064, p<.05]. The difference between the iPod and voice interface B was not significant. This pattern of results is consistent with our predictions.

We computed the number of glances away from the forward roadway for the three systems of between 1.0 and 1.4 s (*short* glances), between 1.5 to 1.9 s (*medium*), and equal to or more than 2 s (*long*). The average number of short glances per task for the

iPod, voice interface B, and voice interface C was, respectively, 4.1 (SD=1.4), 2.2 (SD=2.2), and 2.0 (SD=1.2) (Figure 3). This number was significantly larger with the iPod than either voice interfaces B [$t(16)=3.019$, $p<0.01$] or C [$t(16)=4.148$; $p<0.01$]. The number of medium duration glances per task for the iPod, voice interface B and voice interface C was, respectively, 1.5 (SD=0.7), 0.7 (SD=0.8), and 0.5 (SD=0.5). This number was also significantly larger with the iPod than either voice interfaces B [$t(16)=2.922$, $p<0.05$] or C [$t(16)=5.644$, $p<0.001$] Finally, the number of long duration glances per task for the iPod, voice interface B and voice interface C was, respectively, 0.9 (SD=0.6), 0.4 (SD=0.5), and 0.5 (SD=0.8). This number was significantly larger with the iPod than voice interface B [$t(16)=3.076$, $p<0.01$]. In no case were there significant differences between the number of glances to voice interfaces B and C. In summary, the number of short, medium and long glances away from the forward roadway for the touch interface was about double the number for the two voice interfaces.

We also looked at the control sections to determine whether the frequency of long glance durations when drivers were not engaged in a secondary task differed from the frequency of long glance durations when they were engaged in a secondary task (we had complete data for only 16 of 17 participants in the control sections). Two 20 second periods were chosen in the control drives, one in the suburban and one in the urban section and short, medium and long glance durations classified. Twenty seconds was chosen because it was the time it took drivers on average to traverse the longest straight sections of roadway. During a 20 second period in the control sections, on average there were 0.160 long glances. In order to make a meaningful comparison between the frequency of glance durations in the control section and the frequency of glance durations

in the experimental sections, we need to normalize the frequency for a given participant using a given interface by the time that it took on average for participants to complete tasks using the interface (e.g., if it took on average 40 s for a participant to complete the iPod tasks and the participant made on average four long glances, then we need to multiply four by the fraction 20/40 to get the estimated number of long glances in a 20 s period equivalent to the control section). Doing the above, we find that on average during a 20 s period, the participants using the iPod and voice interface B and C had, respectively, 0.75, 0.29, and 0.37 glances away from the forward roadway. There was a significant effect of condition [$F(2.72,40.73)=6.26$, $p < .001$, Greenhouse-Geisser correction for sphericity). The frequency of long glances in the control section (0.160) was significantly smaller than this frequency when participants were using the iPod [$t(16)= 3.84$, $p < .01$], but not when participants were using either voice interface. The frequency of long glances was greater when participants were using the iPod than it was when participants were using either voice interface B [$t(16)=3.280$, $p < .01$] or voice interface C [$t(16)=2.366$, $p < .01$]. There was no difference in the frequency of the long glances in voice interfaces B or C. In summary, the frequency of long glances was greater with the touch interface than it was with either of the two voice interfaces. However, the frequency of long glances with either voice interface did not differ significantly from this frequency in the control sections.

Finally, we computed the total number of glances per task away from the forward roadway, irrespective of the duration. This average is 13 (SD=6), 9 (SD=5), and 8 (SD=5) for the iPod, voice interface B and voice interface C, respectively (this includes only successfully completed tasks). If we consider all tasks (successes and failures) then

the average number of glances away from the road per task equals 13, 11, and 8 respectively. The main effect of condition is significant [$F(2,23.3)=5.85$, $p=0.011$] with the Greenhouse-Geisser correction]. Additional post hoc comparisons were carried out. The number of glances per task away from the forward roadway, irrespective of the duration, was smaller when participants were using interface C than it was when using the iPod [$t(17)=2.984$, $p < .01$]. The number of glances in the iPod and voice interface B also differed [$t(17)=2.526$, $p < .05$]. The number of glances away from the forward roadway in interface B and interface C did not differ significantly.

3.3. Hazard Anticipation

We computed a measure of tactical scanning, in particular, the percentage of scenarios in which the driver looked for a threat that might materialize (e.g., looked for a pedestrian emerging in front of the truck, Figure 4). We found that for the control, iPod, voice interface B and voice interface C conditions 46%, 35%, 38% and 38%, respectively, of the participants looked for a threat that might materialize (unfortunately, only 12 of the 17 participants had complete eye tracker data in all four sections in the hazardous sections). There was not a significant main effect of condition and so no attempt was made to compare the conditions in any more detail. This result is somewhat surprising given that the participants spent more time with their eyes off the forward roadway when performing a music retrieval task. However, it may be that the lack of power due to a small sample is masking a statistically significant effect, though even then practically the size would be relatively small.

3.4. Vehicle Control

Measures were collected of the lateral deviation (root mean square error) around lane center in the four different condition: control (0.960 m), iPod (0.995 m), voice interface B (1.121 m) and voice interface C (0.843 m). A repeated measures ANOVA indicated that these differences were not significant.

3.5. Workload

Finally, we analyzed the NASA TLX workload scores. The average scores in the control, iPod, voice interface B, and voice interface C conditions were, respectively, 40.7 (SD=15.3), 58.2 (SD=18.2), 61.5(SD=22.0), and 44.5(SD=15.32). There was a main effect of condition [$F(2.24)=7.69$, $p < .001$, with the Greenhouse-Geisser correction for sphericity]. Additional post hoc comparisons were carried out. The workload of drivers in the control and voice interface C conditions did not differ significantly. Nor did the workload of drivers in the iPod and voice interface B differ significantly. All other pairwise comparisons were significant at the .01 level.

The workload scores are perhaps the most revealing of differences between single and multiple entry/turn interfaces. On the one hand, the iPod and voice interface B had equal (and high) workload scores. They were both multiple entry or multiple turn interfaces which made the most demands on attentional resources as users had to navigate the menu hierarchy, either visually or by voice. On the other hand, the control section and voice interface C had equal (and low) workload scores. Voice interface C requires only a single conversational turn and therefore makes fewer demands on attentional resources. The control section had no secondary workload and, correspondingly, has the smallest workload score..

4. Summary

An increasing number of complex tasks are being performed inside the cabin of an automobile, tasks which take much longer on average to perform than typical in-vehicle tasks such as adjusting the volume of the radio, lowering the temperature inside the car, or turning on the defrost. These tasks can create an increase in the likelihood of a crash if they lead drivers to take longer glances (Klauer et al., 2006; Horrey and Wickens, 2007) or too many glances (Blanco, Hankey and Chestnut, 2005) inside the vehicle, if they lead drivers to maneuver less safely (e.g., deviate more around lane center), or if they engage the driver for too long a period of time (Green, 1999). Music retrieval is one in-vehicle task which has been shown to increase the average durations of in-vehicle glances (Chisholm et al., 2008) and the deviation around lane center (Salvucci et al., 2007). In addition, when the music retrieval tasks are especially complex they can take longer than is considered safe and require more glances inside the vehicle than is considered safe (Chisholm et al., 2008).

In this context, it is clear that one cannot safely, efficiently or effectively use field or naturalistic observations to measure the impact of the different music retrieval systems on driver performance. Instead a driving simulator is ideally suited for the evaluation of such systems. First, from the standpoint of safety, there are no risks to drivers in the simulator comparable to the risks in the field. Glances away from the forward roadway for especially long periods of time in the field are potentially deadly; they are of little consequence for the safety of the participant in a driving simulator. Second, from the standpoint of efficiency, naturalistic observations could require 10s if not 100s of hours to gather the same amount of data that is gathered in a simulator. The information in the

simulator can easily be gathered in a fraction of the time. Finally, from the standpoint of effectiveness, it can be more difficult to conclude from experiments in the field that differences in the performance of drivers using the different in-vehicle music retrieval systems are a function of differences in the systems and not differences in the traffic at the time the different systems were evaluated. All can be controlled in a driving simulator and so this problem does not arise.

Given that driving simulators represent a platform in which one can safely, efficiently and effectively evaluate alternatives to the existing music retrieval systems, it makes sense to go forward with such evaluations. Music retrieval systems create risks in no small measure because they have typically used a touch interface. When used outside the vehicle, the touch interface has obvious advantages. But once placed inside the vehicle, the touch interface creates a real danger. One reasonable question in this context is whether a voice interface to music retrieval systems could lower the unsafe behaviors associated with use of such systems. Towards this end, the driving, eye and task behaviors of participants using the three music retrieval systems, one touch system (the iPod) and two voice systems, were compared. In addition, comparisons were made between the driving, eye and task behaviors of the participants both when using and not using the music retrieval systems.

In brief, it appears that voice interfaces can offer a real advantage over touch interfaces on some measures. Compared with touch interfaces, the two voice interfaces reduced the total time drivers spent with their eyes off the road, the number of long glances away from the forward roadway, and the total number of glances away from the roadway irrespective of duration. Moreover, when compared with driving in control

sections, participants using the voice interfaces do not anticipate fewer hazards, do not make more frequent prolonged glances, and do not deviate around lane center more. However, drivers using a voice interface may still be at increased risk on some measures. In particular, when compared with the control sections, although not significant, drivers using the voice interfaces did increase by a factor of about two the number of long glances away from the forward roadway. And, again although not significant, they were less likely to detect a hazard. Perhaps increases in power would have shown that these differences were real.

The voice interfaces also differed from one another. Participants using voice interface B (multi-turn) took longer to complete the task and judged their workload higher than drivers using voice interface C (single-turn). In fact, participants using voice interface B took longer to complete the task than did participants using the touch interface. And they judged their workload the same as the drivers using the touch interface. We argued that this was because voice interface B and the touch interface both required multiple turns/entries. Interestingly, there were no differences between drivers' assessment of their workload in the control sections and in voice interface C.

In summary, if appropriately designed the voice interfaces would appear capable of offering real advantages over touch interfaces on all measures of safety. And a single turn interface would appear to be better than a multiple turn interface. There is one last point which bears mention. Increasingly states are outlawing texting while driving. Texting requires a combination of touch and visual feedback. There may be little difference between the dangers posed by texting and the dangers posed by other devices which require touch and visual feedback like the iPod, among many others. Lawmakers

may need to take a more critical look at what is and is not outlawed. Our research suggest that any interface which requires a combination of touch and visual feedback many times during a typical drive is one which should be considered as potentially unsafe.

5. Acknowledgments

This research was funded in part by a grant from Mitsubishi Electric Research Laboratories in Cambridge, MA. The research was also supported in part by Grant Number 1R01HD057153-01 from the National Institutes of Health and in part by Equipment Grant Number SBR 9413733 from the National Science Foundation for the partial acquisition of the driving simulator. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

6. References

Blanco, M., Hankey, J.M., Chestnut, J.A., 2005. A taxonomy for secondary in-vehicle tasks based on eye glance and task completion time. In the *Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society*. Santa Monica, CA: Human Factors and Ergonomics Society, pp. 1975–1979.

Chan, E., Pradhan, A. K., Knodler, M. A, Pollatsek, A. and Fisher, D. L. (January 2008). Empirical Evaluation on a Driving Simulator of the Effect of Distractions Inside and Outside the Vehicle on Drivers' Eye Behaviors. *Proceedings of the 87th Transportation Research Board Annual Meeting* CD-ROM, TRB, National Research Council, Washington, D.C.

Chisholm, S.L., Caird, J.K. and Lockhart, J. (2008). The effects of practice with MP3 players on driving performance. *Accident Analysis and Prevention, 40,* 704–713.

Crundall, D., & Underwood, G. (1998). Effects of experience and processing demand on visual information acquisition in drivers. *Ergonomics*, *41*, 448–458.

Dingus, T.A., Antin, J.F., Hulse, M.C., Wierwille, W.W. (1989). Attentional demand requirements of an automobile moving-map navigation system. *Transportation Research A*, *23A* (*4*), 301–315.

Fisher, D. L., Laurie, N. E., Glaser, R., Connerney, K., Pollatsek, A., Duffy, S. A. and Brock, J. (2002). The Use of an Advanced Driving Simulator to Evaluate the Effects of Training and Experience on Drivers' Behavior in Risky Traffic Scenarios. *Human Factors, 44,* 287-302.

Green, P. (1999). The 15-second rule for driver information systems. In: Proceedings of the ITS America 9th Annual Meeting. ITS America,Washington, DC.

Green, P. (2007.) Where do drivers look while driving (and for how long)? In: Dewar, R.E., Olson, R. (Eds.), *Human Factors in Traffic Safety* (2nd ed). Tucson, AZ: Lawyers & Judges Publishing, pp. 57–82.

Greenberg, J., Tijerina, L., Curry, R., Artz, B., Cathey, L., Grant, P., Kochhar, D., Kozak, K. , and Blommer, M. (2003). Driver distraction: evaluation with an event detection paradigm. *Transportation Research Record, 1843*, 1-9.

Horrey, W. J. and Wickens, C. D. (2007). In-Vehicle Glance Duration Distributions, Tails, and Model of Crash Risk. *Transportation Research Record, 2018*, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 22–28.

International Standards Organization. (2002). *Road Vehicles—Measurement of Driver Visual Behaviour with Respect to Transport Information and Control Systems. Part 1. Definitions and Parameters, ISO Committee Standard 15007-1*. Geneva, Switzerland: International Standard Organization.

International Standards Organization. (2007). *Road vehicles -- Ergonomic aspects of transport information and control systems -- Occlusion method to assess visual demand due to the use of in-vehicle systems, ISO Committee Standard 16673*. Geneva, Switzerland: International Standard Organization.

Klauer, S.G., Dingus, D.R., Neale, T.A., Sudweeks, J., Ramsey, D.J. (2006). *The impact of driver inattention on near-crash/crash risk: an analysis using the 100-car*

naturalistic study data (Rep. No. DOT HS 810 594). National Highway Traffic Safety Administration, Washington, DC.

Lee, S. E., Olsen, E. C. B., & Simons-Morton, B. (2006). Eyeglance Behavior of Novice Teen and Experienced Adult Drivers. *Transportation Research Record, 1980*, 57-64.

McKnight J. A., & McKnight S. A., (2003) Young novice drivers: Careless or clueless. *Accident Analysis and Prevention*, *35*, 921–925

Muttart, J., Fisher, D. L., Knodler, M. and Pollatsek, A. (2007). Driving Simulator Evaluation of Driver Performance during Hands-Free Cell Phone Operation in a Work Zone: Driving without a Clue *Transportation Research Record, 2018*, 9-14.

Olsen, E. C. B., Lerner, N., Perel, M., & Simons-Morton, B. G. (2005). In-car electronic device use among teens. *TRB Annual Meeting CD-ROM*.

Pollatsek A, Narayanaan V, Pradhan A. K. , and Fisher DL. The Use of Eye Movements to Evaluate the Effect of PC-Based Risk Awareness Training on an Advanced Driving Simulator. *Human Factors* In Press.

Salvucci, D.D., Markley, D., Zuber, M. and Brumby, D. P. (2007). iPod Distraction: Effects of Portable Music-Player Use on Driver Performance. *CHI*, San Jose, California.

Wang, J. S., Knipling, R. R., and Goodman, M.J. (1996). The role of driver inattention in crashes: New statistics from the 1995 crashworthiness data system. In *40th Annual Proceedings of the Association for the Advancement of Automotive Medicine*, Vancouver, British Columbia.

Weinberg, G and Kondili, D. (2008) Display Style Considerations for In-Car Multimodal Music Search. *IADIS International Conference on Interfaces and Human Computer Interaction (IHCI)* Amsterdam, The Netherlands, 323-328. (accessed 8/5/2008, http://www.merl.com/publications/TR2008-038/.)

Wikman, A., Nieminen, T., & Summala, H. (1998). Driving experience and time-sharing during in-car tasks on roads of different width. *Ergonomics, 41*(3), 358-372.

7. Figure Legends

Figure 1. iPod Interface: Hierarchical Structure

Figure 2. Urban and Suburban Perspective Views. (Top panel: urban scenario. Bottom panel: suburban scenario.)

Figure 3. Distribution of Glances Away from the Forward Roadway

Figure 4. Plan View of Truck Crosswalk Scenario. (Yellow oval represents area from which a hidden risk – e.g., a pedestrian – could materialize. Red circle represents area which driver should actively scan. Gray vehicle is driver.)

8. Tables

Table 1. Three Music Retrieval System Evaluated

Music Retrieval System	Display Size	Touch Interface	Primary Input Demands	Number of Entries/Turns
A. iPod	2.5-inch LCD	Touch-sensitive click-wheel	Touch/Visual	Multiple
B. Commercial aftermarket in-dash unit	7-inch LCD	Wheel-mounted remote	Speech (dialog-based)	Multiple
C. SpeakPod prototype	7-inch LCD	Wheel-mounted remote	Speech (query-based)	Single

9. Figures

Figure 1. iPod Interface: Hierarchical Structure

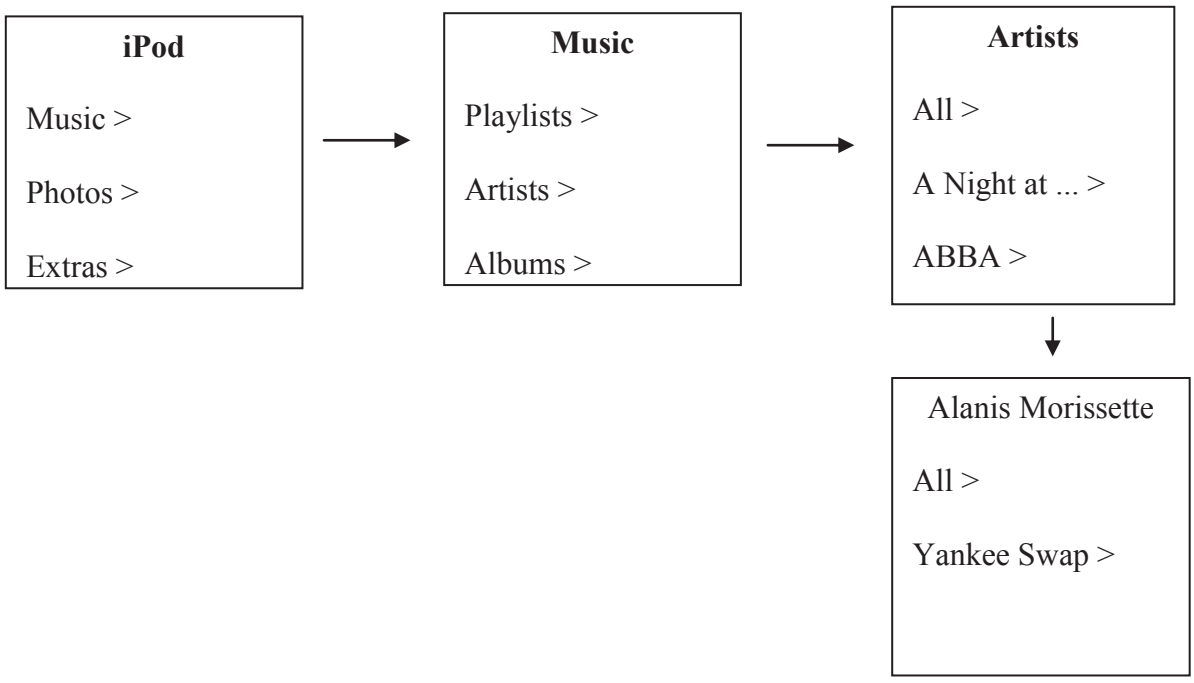

Figure 2. Urban and Suburban Perspective Views. (Top panel: urban scenario. Bottom panel: suburban scenario.)

Figure 3. Distribution of Glances Away from the Forward Roadway

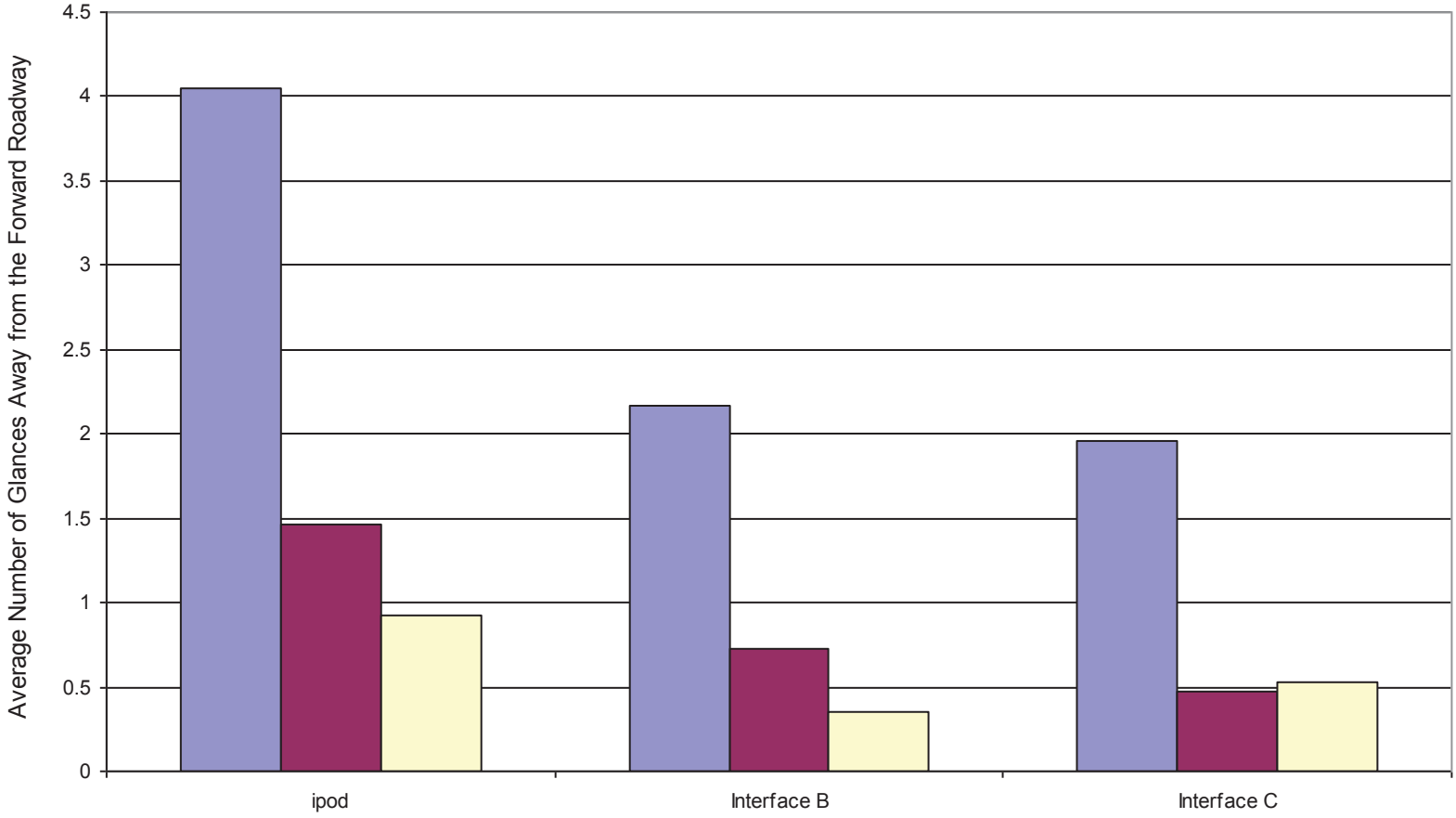

Figure 4. Plan View of Truck Crosswalk Scenario. (Yellow oval represents area from which a hidden risk – e.g., a pedestrian – could materialize. Red circle represents area which driver should actively scan. Gray vehicle is driver.)