APPLYING SPINDEX AUDITORY CUES WHILE DRIVING AND PERFORMING A SECONDARY SEARCH TASK

A Thesis
Presented to
The Academic Faculty

by

Thomas M. Gable

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

Georgia Institute of Technology
August, 2015

APPLYING SPINDEX AUDITORY CUES WHILE DRIVING AND PERFORMING A SECONDARY SEARCH TASK

Approved by:

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

Dr. Richard Catrambone
School of Psychology
Georgia Institute of Technology

Dr. Audrey Duarte
School of Psychology
Georgia Institute of Technology

Dr. Frank Durso
School of Psychology
Georgia Institute of Technology

Date Approved: May 18, 2015

ACKNOWLEDGEMENTS

I would like to thank my advisor Bruce Walker for his guidance through this process as well the rest of my committee, Richard Catrambone, Audrey Duarte, and Frank Durso for all of their great feedback. I would also like to thank all of those in the Sonification Lab that assisted with parts of the project for their great help. Finally I would like to thank my friends, my family, and of course Mary Dixon for encouragement and support throughout this process.

Portions of this work were supported by a National Science Foundation Graduate Research Fellowship (DGE-1148903).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This thesis investigated the impact of applying “spindex” text to speech (TTS) auditory cues in a long-list searching task on a cell phone while driving as compared to a visuals-only interface. Previous research has found that when using advanced auditory cues (i.e., spindex), both participants’ visual dwell time off the road and subjective workload is lower than when using visuals-only displays. The current study expanded on previous research by investigating the impact of these cues through two factors of distraction – workload and willingness to engage – as well as investigating the use of the Visual Auditory Cognitive and Psychomotor (VACP) predictive workload scale. Previously investigated workload measures of visual behaviors, subjective workload, primary and secondary task performance, and preferences were supplemented with additional measures via physiological detection and VACP as a predictive measure of workload. The newly added factor of willingness to engage was investigated via the inclusion of two different driving difficulties (hard or easy), by modifying the roadway type (city or highway). Results support previous findings of lower workload when novice users employ the spindex-TTS cues compared to visuals-only as seen through increased dwell time on the driving task, less glance frequency off the driving task, lower subjective demand, and higher perceived performance, but no conclusive results were seen in regards to willingness to engage. In addition, the patterns of workload predictions from the VACP measure matched well with data collected during the experiment. These results and their implications for the application of spindex-TTS cues as well as the future measurement of willingness to engage and use of the VACP scale as a predictor of workload are discussed.

CHAPTER 1: INTRODUCTION

The National Highway Traffic Safety Administration (NHTSA) (2013) reported that in 2010, 26,000 police-reported crashes were linked to drivers using a device or in-vehicle controls. These secondary interactions within the vehicle, or in-vehicle dual tasks, mean the driver is attempting to balance the completion of two goals (driving and another non-driving task) using two different interfaces, sources of information, and inputs. Due to the common undertaking of these secondary interactions, research investigating driving while performing other in-vehicle tasks is central to the automotive community.

Past studies investigating in-vehicle dual tasks have focused on communications technology tasks such as talking on the phone (Alm & Nillsson, 1995) or texting (Drews, Yazdani, Godfrey, Cooper, & Strayer, 2009). Most of these studies conclude that simply adding a basic secondary communications task increases workload and decreases driving performance. Meanwhile, recent advances in technology have expanded the types of secondary tasks a driver can complete, far beyond that of a basic phone call or text. In-vehicle technologies (IVTs) allow drivers to complete driving and non-driving related tasks such as driving navigation, music selection, video viewing, and many others, and can be completed on devices both integrated into the vehicle (i.e., infotainment units) and brought into the vehicle by the driver (i.e., cell phones). Although built-in IVTs are often designed with the driving situation in mind, brought-in devices are usually designed to be the primary task at the time of interaction and are consequently created for visual interaction. Therefore, when these tasks are taken into the car it means an individual is performing two tasks designed to be primary tasks, often both tasks heavily reliant on visual interaction. However, this heavy dependence on visual interaction with interfaces has not been abandoned for built-in interfaces either, as designs for many types of integrated IVTs continue to follow the trends in computing of visual heavy interfaces.

If we are to allow drivers to complete the all too common dual task in the vehicle with advancing IVTs, and do so safely, researchers and developers must create and implement superior forms of interaction with the devices used in the vehicle. We must look past visual based interfaces while also ensuring we investigate more of the common tasks completed in a vehicle than simply communications technologies. The present paper was an effort to investigate the effects on a dual task situation of applying a multimodal, advanced auditory cue called spindex, to a list-search interface. The study measured the effects of the dual task situation on workload through assessing driving and list search performance, visual behaviors, and subjective and physiological measures of workload as compared to a standard interaction method of a visuals-only interface. The study also explored the indirect modification of drivers' willingness to engage and the influence this had on their strategies of interaction for each type of interface. Finally the study explored the use of a predictive measure of workload to investigate the accuracy of said predictions to collected experimental data.

Driver Distraction, Workload and Willingness to Engage

When considering driving and secondary tasks "driver distraction" is often the first term discussed. This term however brings with it a large debate, as "driver distraction" has been defined in a multitude of ways (Young, Regan, & Hammer, 2007). Lee, Young, and Regan (2008) reviewed numerous descriptions of distraction and came to a widely accepted definition, stating, "driver distraction is a diversion of attention away from activities critical for safe driving toward a competing activity" (p. 34). Though this definition encompasses many activities as distracting, it does not attempt to break down the individual factors involved in "driver distraction," making them unavailable for independent investigation and therefore making any attempt to measure the factors of distraction un-diagnostic. A paper published by NHTSA uses a similar definition but also states it is more helpful to examine the four categories of distraction – visual, auditory, cognitive, and biomechanical – on their own (Ranney, Mazzae, Garrott, & Goodman, 2000). Whereas this breakdown brings forth four modalities within distraction that could be independently measured, possibly making the results of measurement diagnostic and informative for future redesign of interfaces, finding consensus on the ideal method of measuring distraction within each factor proves difficult, as actual "driver distraction" is not quantifiable. Instead, researchers will often investigate the factors that potentially influence distraction, workload and "willingness to engage" (Ranney et al., 2000).

The first of these factors, workload, has many definitions but is summarized well by Hart (2006) as "a term that represents the cost of accomplishing mission requirements for the human operator" (p. 1). Workload, much like distraction, is often broken down into four modalities of visual, auditory, cognitive, and biomechanical (psychophysical) demand (Keller, 2002). Mitchell (2000) points out that whilst no one theory of workload has consensus within the research community, one of the more common theories is Multiple Resources Theory (MRT), "a theory of multiple task performance" (Wickens, 2002, p. 1). MRT is a theory of human information processing based on the idea that the human brain uses "multiple" channels to independently process information across different modalities, and that the “resources” being used are limited but can be distributed across tasks (Wickens, 2002). MRT is particularly useful for application within the space of driving and secondary tasks, as “the value of multiple-resources lies entirely in its ability to account for performance in the ‘overload’ situation, where the operator is called to perform two or more tasks at one time” (Wickens, 2002, p. 2). This idea of overload is often referred to as crossing the “red-line,” the point at which the amount of workload goes over the total amount of available resources, and where performance on one or both of the tasks begins to decrease. For a full review of MRT see Wickens (1981; 2002). Finally, it is important to measure workload, as “the principal reason for measuring workload is to quantify the mental cost of performing tasks to predict operator and system performance” (Cain, 2007, p 4-3), information which can then be used to inform designers and researchers of what interfaces to use in the vehicle.

The second factor in Ranney et al.’s (2000) definition of distraction, “willingness to engage” is a term that “refers to the conscious or unconscious decision processes involved in electing to carry out secondary tasks while driving” (p. 2). The authors go on to break down the factors that affect willingness to engage, listing driver experience, vehicle or display design, environmental factors such as weather, situational factors such as the urgency of the task to be completed, and difficulty of the task (Ranney et al., 2000). This idea of willingness to engage is particularly useful to consider with dual task situations in driving and can help researchers to control for many variables that may influence results in different studies. Researchers can also vary some of these factors to modify willingness to engage and then investigate how these changes manifest through the use and performance with an interface, an important factor within driving and secondary task research.

It is through this lens of distraction – workload and willingness to engage – that the current paper focuses its approach to the study of dual task performance with in-vehicle technologies. Within workload the focus will be the four modalities of visual, auditory, cognitive, and psychomotor, attempting to measure and predict as many of these factors as possible. Meanwhile factors influencing willingness to engage will be modified through driving difficulty to see the effects of these changes on use of the interfaces, and controlling those that could affect the willingness to engage in undesired ways.

Measuring Workload in Driving Research

Similar to the variety of definitions available for workload, there are multiple approaches to measuring it (Miller, 2001). The goal of measuring workload "is to quantify the mental cost of performing tasks to predict operator and system performance" (Cain, 2007, p 4-3). Whilst trying to be diagnostic about what modality workload is coming from is an ideal situation, many of the modalities of workload influence the others and make this difficult. Brookhuis and de Waard (2001) offer three basic ways of measuring drivers' workload: subjective measures, physiological measures, and task performance. Whereas not diagnostic in what modality workload is being added to and not precise as to how much workload was used, driving performance is a straightforward and nonintrusive approach to see the effect of a dual task situation in driving (Miller, 2001; Wickens, 2008). Other types of task performance can also be used as detectors of workload such as measuring performance on a secondary task, or adding another, tertiary task, both of which are meant to probe the "residual capacity" of the user, or the amount of workload that is still available after the primary task is allocated the amount it is required (Wickens, 2008). Those who want a direct measurement of workload will often measure subjective workload via questionnaires, or through objective measures of workload through the collection of physiological data.

Other times researchers will attempt to investigate different modalities of workload independently. Both cognitive and physical (psychomotor) workload can be measured in this way via a subjective measure, the NASA-Task Load Index (TLX) by looking at the subscales that are combined to determine overall workload (Hart, 2006).

One modality of workload that is particularly important within the driving environment and often evaluated on its own is that of visual load. Investigating the distribution of visual behaviors during the task can be very informative of how taxing a dual task situation is and how it may affect safe driving, as visual distribution is an important resource within driving (Engström, Johansson, & Östlund, 2005; SAE J-2396). Unfortunately, a method for measuring auditory workload does not currently exist, and therefore cannot be measured independent of the other modalities (Nees & Walker, 2011). There is however, a way to try and predict auditory workload along with the visual, cognitive, and psychomotor modalities though not often applied in the driving literature.

Driving Performance

Driving performance is an easily applicable measure to determine outcomes of a dual task situation, as you are evaluating the direct thing you are investigating and it is not intrusive to the participant (Miller, 2001). There are many performance measures that can be collected in driving research, often depending on the driving task being performed and the resources available. In an attempt to standardize the usage and reporting of these measures the Society for Automotive Engineers (SAE) has been compiling a document to provide researchers with guidance in their use, SAE J2994 (SAE, 2013). The document breaks driving performance measures into four main families via vehicle based measures or responses of the driver on either longitudinal or lateral control, which can all be measured and reported in combination or separate from each other to gain knowledge of the driver's behaviors.

Longitudinal control measures are ways to determine the distance or time between the driver's vehicle and a vehicle in front of them, with increases in distance or time between the two vehicles or reaction time seen alongside rises in workload, particularly with visually demanding tasks (Angell et al., 2006; Young, Regan, & Lee, 2008). Vehicle related measures of this type are often used in studies that involve a following task and include distance measures such as distance gap, or distance headway, and time variants such as time gap (SAE, 2013; see e.g. Alm & Nillsson, 1995; Drews, et al., 2009). Driver centric measures include those such as brake response time (Angell et al., 2006; Young et al., 2008; e.g. Lamble, et al., 1999). Means of these types of measures are most commonly reported, but standard deviations can also be used (Young et al., 2008).

Vehicle centric lateral control measures, also referred to as lane keeping measures, or lane maintenance measures, look at the driver's ability to keep the car in the lane (Angell et al., 2006). Most researchers apply one of two major subsets of lateral control measures: lateral position measures, a family of continuous measures of lane-keeping which increases in value with the level of workload; or lane/roadway departures and their related measures, a family of measures for the discrete and infrequent event of exiting the lane in some form that also increases as workload goes up (Angell et al., 2006; SAE, 2013; Young et al., 2008). Lateral position measures are more widely collected due to their continuous nature, with some of the common ways of determining the variable being lateral lane position, and the standard deviation of lateral position (SDLP) (Angell et al., 2006; SAE 2013; Young et al., 2008; see e.g. Drews, et al., 2009; Engström et al., 2005; Lamble, Kauranen, Laakso, & Summala, 1999; Son, Park, & Oh, 2012; Zhao, Brumby, Chignell, Salvucci, & Goyal, 2013). Though measures of lane or roadway departures are measured less frequently due to the infrequency of occurrence, they can be useful in some situations and include measures such as roadway departures, lane departures, and other variations. (Angell et al., 2006; SAE, 2013; see e.g. Drews, et al., 2009). An example of a lateral control response measure is steering reaction time, with decreases in time being seen with increases in workload (Angell et al., 2006; SAE, 2013; Young et al., 2008).

Secondary and Tertiary Task Performance

Primary task performance is a direct measure of the outcome of adding a secondary task to driving, but another way to use task performance to see the influences of a dual task situation is through the secondary task performance or the addition of a tertiary task. These measures reflect the "residual capacity" of a driver, or the workload capacity that went unused in completing the primary task, or in the case of a tertiary task, the capacity left after the dual task completion (Wickens, 2008). As with driving performance these measures simply show the researcher if the tasks went over the "red-line," but are not necessarily informative into how much over, nor what modality may have been the cause.

Determining the measures to collect from secondary task performance is very much based on what secondary task is being investigated in the research. Often, as it is in real life, the tasks are visual-manual and have the basic measures of time to complete a task, accuracy in completing the task, and reaction time to the secondary task, all of which are seen to get worse as the workload of the dual task situation increases (e.g. Gable, Walker, Moses, & Chitloor, 2013; Horrey, Wickens, & Consalus, 2006; Lansdown, Brook-Carter, & Kersloot, 2004; Lee, Roberts, Hoffman, & Angell, 2012). Similarly to secondary tasks, tertiary task can also be added in an attempt to see the changes in available resources. These types of tasks are also usually based on time and accuracy, common ones being that of the peripheral detection task (PDT) or the (standardized) method of detection reaction time (DRT) (Angell et al., 2006; Ranney, Baldwin, Smith, Mazzae, & Pierce, 2014).

Subjective Measures of Workload

Task performance can be an informative measure but it only investigates if either the primary task workload is above the "red line" or how much "residual capacity" is left for the driver to use on the secondary or tertiary task (Wickens, 2008). Instead, researchers can apply methods to determine levels of workload across all tasks such as through the use of subjective measures. Researchers often investigate drivers' subjective workload by simply asking participants for their ratings of the task difficulty or how much effort they had to apply to the task (Lewis-Evans, De Waard, & Brookhuis, 2011; Recarte & Nunes, 2003). There are however, a number of standard subjective measures that researchers use in driving including multidimensional scales such as the subjective workload assessment technique (SWAT), the workload profile (WP), the NASA-TLX, and the driving-specific driving activity load index (DALI), as well as unidimensional scales such as the rating scale mental effort (RSME) (Paxion, Galy, & Berthelon, 2014). One of the more well known and widely used subjective measures of cognitive load is the NASA-TLX, which measures six subscales of workload including effort, temporal demand, physical demand, frustration, performance, and mental demand (Hart & Staveland, 1988; Lansdown et al., 2004). One reason that it is so widely used is because it can predict task performance and can be relatively easy to implement (Paxion et al., 2014). Researchers will often use the single numerical value output by this measure as a measure of total workload but can also look at each raw subscale independently of each other to investigate particular factors, known as the raw TLX (RTLX) (Hart, 2006). These scales have been used in many studies and are often used to base other measures of workload off of, however their subjective nature leads some to look for other methods of measurement.

Objective Measures of Workload

In more recent years, when researchers are investigating overall measures of workload, but want to avoid issues with subjective measures, they use objective, physiological measures of workload due to their growing availability of measurement. Methods commonly used in the driving field include heart rate (HR), heart rate variability (HRV), inter-beat intervals, electroencephalogram (EEG) activity, respiration rate, alertness monitoring, and skin conductance level (SCL) (Brookhuis & de Waard, 2010; Engström et al., 2005; Mehler, Reimer, & Coughlin, 2012; Mehler, Reimer, Coughlin, &

Dusek, 2009). One of the first physiological measures to be used to detect workload changes in human factors literature was HRV, with decreases in variability being seen with increases of working memory (Aasman, Mulder, & Mulder, 1987). More recently, Mehler et al. (2012) tested the efficacy of two other popular physiological measures, HR and SCL, after determining they were the more effective measures from a larger range of factors (Mehler et al., 2009). In the 2012 study, participants drove and performed a cognitively demanding auditory and vocal memory task. Results showed that the two measures increased almost linearly as cognitive demand increased, with HR increasing slightly more linearly and with a faster response time than SCL at higher levels of cognitive load. These results point to the use of HR and HRV as relatively reliable and simple way of determining workload through physiological measures.

Visual Workload

Another approach to measuring workload is to gauge the effects of workload through visual behaviors. To investigate visual workload through the measurement of participants' allocation of visual resources, researchers can apply eye-tracking technologies. Studies often apply a number of different types of measures in eye tracking, and standards have been developed for research within the driving context (e.g., SAE, 2000). Some of the common terms that are used in discussing visual load and behaviors include fixation – the "alignment of the eyes so that the image of the fixated target falls on the fovea for a given period of time;" and glance – "the time from the moment at which the direction of gaze moves toward a target (e.g., the interior mirror) to the moment it moves away from it" (SAE, 2000, p. 7). Some of the commonly applied measures in visual behavior research are that of glance frequency, dwell or fixation time, and percent dwell time (SAE, 2000; see Gable et al., 2013; Horrey et al., 2006). For an in-depth list of standard definitions on driver visual behavior see SAE J2396 (SAE, 2000).

Predicting Workload

The measurement of workload is key to knowing if interfaces are detracting from drivers' abilities to perform their primary task safely. However, the ability to predict workload could prove even more valuable for future research and development of in-vehicle interfaces. In addition, being able to measure the auditory modality of workload independently of the other modalities of workload could prove useful when looking at multimodal interfaces. Little research has been done in predicting workload within the driving domain but a few options do exist. One of these options, and the option most suited for this investigation, as it based on MRT, is the visual auditory cognitive psychomotor (VACP) scale (McCracken & Aldrich, 1984). The VACP scale focuses on breaking down the term "workload" into its four measureable parts of visual, auditory, cognitive, and psychomotor load (Keller, 2002). A user of the scale creates a task analysis of the activity on which workload needs to be predicted, has raters assign specific values (up to 7) that coincide with anchor statements on each of the four resources that apply to each part of the task, and then adds the scores up within each resource across the parallel tasks (Mitchell, 2000). As more demand is placed on one modality of workload, higher scores are estimated, but if the load is distributed across modalities, then the total predicted workload is lower, as expected in MRT (Keller, 2002). Often these calculated predictions are then paired with a task network-modeling program to create a model of workload for the task, which has been shown to be useful, particularly in predicting situations where a system or interface may put too much load on the users (Cain, 2007). Previous work has found that the scale correlates well with subjective workload ratings and also predicts performance variances (Mitchell, 2000).

Willingness to Engage

Although workload is commonly measured in driving research, willingness to engage is often not considered, despite its possible large implications on driver safety. While not done within the driving context, Fu and Gray (2006) studied strategies within a computer based dual task and found that participants allowed for tradeoffs between completing two tasks, determined by the cost of gathering information and the utility of that information in completing the overall goal. A model they created to explore the space output similar data to those collected from the participants. Whereas the result from these tradeoffs was "suboptimal performance" they found it was stable performance; but important to this discussion, heavily reliant on the environment the task was being completed in (Fu & Gray, 2006). These results point to the importance of environment in the way individuals modify their strategy in dual task situations or as the term being used in this document, their willingness to engage.

Within driving and secondary task research driver experience, vehicle factors, environmental factors, situational factors, and task characteristics could all have major influences on a driver's willingness to engage, and therefore affect the results of a study (Ranney et al., 2000). To investigate one or many of these factors researchers must attempt to control or at least limit the effect of the other factors at hand. Driver experience is a factor that can be easily controlled via a within subjects design. The vehicle factor, such as display design is a major factor of interest in most in-vehicle dual task studies, as the goal is to investigate the way different interfaces interact with driving performance, so that factor is changed when the interfaces are manipulated throughout a study. Display design also influences the factor of task characteristics as it influences how tasks are completed, however the difficulty of the tasks within each interface should be controlled as much as possible to ensure the tasks are of similar difficulty if they were on the same interface. The last two factors, situational factors and the environmental factors, can play very interesting parts in research within secondary tasks and driving.

Brumby, Davies, Janssen, and Grace (2011) investigated the issue of situational factors by modifying the priority for the secondary task when drivers were interacting with a visuals-only interface and a visuals-plus-auditory interface, finding that by emphasizing the priority of the secondary tasks participants dramatically changed their strategy of interaction with the interface. Prior to the instructions to change emphasis on the secondary task participants had relied more on the auditory cues, however after the priority of the secondary task was changed they abandoned the auditory feedback and used primarily visuals to search, as it was faster. Such strategic tradeoffs between completing two tasks and the resulting "suboptimal performance" on one task or the other could be considered together to explain why a task important to the driver, or one they felt they should finish quickly, would cause the driver to direct more visual attention toward the secondary task. This change in willingness to engage and therefore in driver strategy with visual attention is an interesting finding and should be investigated further but also considered in future research as a necessary control.

Environmental factors have also been investigated in driving research. Kun, Brumby, and Medencia (2014) investigated the influence of willingness to engage, what they called participant strategies, on use of an interface by modifying the difficulty of the drive, therefore changing the environmental factors. Kun et al. studied this through having participants drive a city and a highway condition and perform a secondary task. The different drive types, city and highway, were considered to be difficult and easy driving conditions respectively due to higher amounts of visual distractions, more navigational movements, narrower lanes, and other factors occurring in a city drive that made it more difficult. Measuring participants' visual behaviors showed that drivers modified their visual interactions with the secondary task interfaces, having glance durations away from the road and towards the secondary task for shorter periods of time during the more difficult city drive, and for longer periods of time during the easy highway drive. The researchers also reported that the participants had better lane keeping performance during the more difficult driving task, possibly due to more time with their eyes on the road, but also possibly due to narrower lanes, leaving less room for safe variation within the lane itself. Although the researchers did not report secondary task performance it could be expected that such visual attention differences, and better performance on the driving task imply a lower level of residual capacity left to be used in the harder, city drives. This increase in driving difficulty modified the drivers' willingness to engage, which they clearly saw through the visual behavior changes, but the effects of this workload change were not seen in the workload metrics, possibly due to a decrease in residual capacity since the primary task was harder in those conditions.

Other researchers have also used measurements of visual attention on the road as well as a secondary task to determine the driver's changes in willingness to engage. Research focusing on these visual behaviors within changing driving difficulties found a number of variables that increase as the difficulty of the driving task increases, such as more dwell time on the center of the road, shorter glances away from the road, and higher dwell time on the road than on the secondary task (Victor, Harbluk, & Engström, 2005). These results agree with those found by Kun et al. (2014), but they also found that the participants had no change in the glance frequency away from the road during these harder conditions, simply changes in glance durations (Kun et al., 2014). In addition Victor et al. (2005) found that when the complexity of a visual secondary task increases, the drivers tend to have a higher glance frequency toward the secondary task and less at the road, as well as more concentrated on the center of the road when they do look at the driving task. When investigating auditory secondary tasks they found that as the gaze concentration on the middle of the driving task was found to be different from the baseline, it was no different across difficulties of the secondary task itself. These results hint at possible differences between visual and auditory tasks and the influence they can have on the visual resources allocation of the driver and their willingness to engage with such interfaces.

These studies into the strategies and the effects of performing a dual task situation with differing levels of difficulty and importance of the driving task point to interesting considerations when doing studies related to in-vehicle tasks and should be controlled as much as possible. More specifically, disparities in the driving tasks could modulate differences in the secondary task performance, making it harder to complete a visual-manual task in such an environment and allowing differences between good and bad interfaces to be seen more easily in the difficult, city driving task, than an easy, highway task, due to changes in the willingness to engage.

It is important to consider the basis behind using some of the workload measures and how changing the difficulty of the primary driving task can influence results. As stated before, many of the workload measures determine the "residual capacity" after completion of the primary driving task, so if the driving task is very easy a large amount of residual capacity remains to be used and adding an easy or a hard secondary may lead to no differences in the secondary task performance due to a floor effect (Wickens, 2008). The opposite situation may also take place where a driving task is very hard, leaving no residual capacity, meaning no differences may be found in different types of secondary tasks due to a ceiling effect. This means that the difficulty of a driving task must be carefully considered during the method creation of a study due to the effect it has on residual capacity but also how this in turn will modulate willingness to engage, and the interaction of these two factors. This issue does however, leave an interesting question to researchers regarding how participants might respond in these situations and how they compensate or change strategies when the willingness to engage is manipulated.

Driving and List Searching

Applying combinations of these techniques can help researchers to see the effects of different interfaces. One type of interface interaction that is investigated less often than communication tasks but is quite distracting in the vehicle, due to high visual demand, is list searching. Drivers perform this task in situations such as exploring a song list or looking for a contact on built-in or brought-in devices. The extent of this occurrence within the vehicle is difficult to determine as it is involved in many tasks but in a 2001 study, the authors reported that simply interacting with the radio/cassette/CD interface was a contributor to about 11.4 percent of all distraction related crashes tracked by the study, while the use of a cell phone only contributed to about 1.7 percent of crashes (Stutts, Reinfurt, Staplin, & Rodgman, 2001). In many of these situations of list searching, the user applies recognition memory instead of recall, and therefore, popular in-vehicle speech interfaces cannot be used.

Browsing a song list on an mp3 player or smartphone is a good examples of a list-searching task undertaken in the car. Research has found that using mp3 devices when driving decreases driving performance as compared to not doing a search task (Harvey & Carden, 2009). Similar driving decrements have also been found in other studies, as well as increases in dwell time on the secondary task and PRT (a tertiary task) when performing a difficult mp3 task as compared to no task and driving (Chisholm, Caird, & Lockhart, 2008). In addition these negative effects on driving performance and visual behaviors do not appear to significantly lessen with practice or with the application of after-market controllers, which are sold to help people interact more safely with these devices in their vehicles (Chisholm et al, 2008; Lee, et al., 2012).

Applying Auditory Displays

In situations where visual demand is high within the vehicle cockpit, such as a list-searching task, Nees and Walker (2011) recommend the application of auditory displays. They state that visual displays in the vehicle have a fundamental shortcoming: They require high levels of visual attention when ideally the driver should have visual attention focused on the driving task. Though this would seem to be adding the same amount of workload to the driver, just through a different modality, in this case the total is not always equal to the sum of its parts. MRT describes the allocation of resources across modalities of input, or “cross-modal time-sharing,” as leading to better performance on tasks than applying the same amount of workload over one modality, or "intra-modal," likely due to less overall workload (Wickens, 2002). This means that by having a task performed primarily via one modality (i.e., the visual modality of the driving task) and another task performed primarily with another modality (i.e., using the auditory modality for list searching) an individual will have overall better performance on the two tasks than if they attempted to perform both tasks visually.

Research seems to support these theories. Results of one study showed lower dwell time off the primary driving task when using a multimodal interface as compared to a visuals-only interface (Chisholm et al, 2008). Liu (2001) tested workload, driving performance, and hazard detection (PRT) in a study comparing a multimodal display (auditory and visuals) to a visual-only display and found that the multimodal interface led to better performance on hazard detection, navigation and driving tasks, and lower workload. This could be inferred to mean more time was spent visually on the primary task. Another study found that two novel auditory interfaces were preferred by drivers over the visuals-only interface, and that participants had better driving performance, and lower perceived workload, although task completion times were slower for longer tasks (Sodnik, Dicke, Tomažič, & Billinghurst, 2008). Similar increased times to completion were found when a visual display was compared to a multimodal display while driving; however, the multimodal display again decreased the risk created by interacting with the interface (Zhao et al., 2013).

Although these results are promising for auditory or multimodal interfaces, the time it takes to complete an action is of particular concern when considering the adoption of these types of interfaces. The results of research discussed earlier by Brumby et al.

(2011) showed that the slower interaction for auditory displays than when using a visuals-only display negatively affected drivers' use of the system due to increasing their willingness to engage visually with the interface. This increased urgency to complete the secondary task should be considered as a modification of the willingness to engage for different drivers or for different tasks in real life. If a driver has a very important list searching task to complete, or feels it is very important, they may not be willing to wait to use these auditory cues and may abandon them altogether, reverting back to visual interaction with the interface and negating the positive effects that an auditory interface can have on drivers. Instead researchers must find and apply faster methods of auditory interfaces to these situations.

Advanced Auditory Cues

To address the issue of slow auditory feedback, advanced speech-based auditory cues for list navigation, such as a "spindex" or "spearcons" were developed (Jeon & Walker, 2011; Walker, Nance, & Lindsay, 2006). A spearcon is a brief sound produced by speeding up a spoken phrase, even to the point where the resulting sound may no longer be comprehensible as a spoken word, and can be very useful for short, well known menus (Walker, Nance, & Lindsay, 2006). A spindex on the other hand (i.e., speech index) is a short non-speech auditory cue based on the pronunciation of the first letter of each menu item and is particularly useful in alphabetical lists (Jeon & Walker, 2011). Spindex cues are usually followed up by spearcons or Text To Speech (TTS) to be even more informative, and are often made to be interruptible to allow for rapid movement through a list. These cues are used to enhance a typical auditory menu, which would otherwise consist only of spoken menu items. Jeon, Walker, and Srivastava (2012) investigated the effects of these cues on a search task: Participants performed a primary search task on a mobile phone by flicking, wheeling, or tapping, while receiving feedback via visuals plus advanced auditory cues, or with just visuals. Results indicated that when participants heard auditory menus that contained advanced auditory cues they had significantly faster search times and lower subjective workloads, compared to menus with no enhanced auditory cues using the same input methods (Jeon et al., 2012).

These advanced auditory cues have also been applied to auditory menus in the driving context. In one such study, the authors found decreased subjective cognitive workload and item selection time, as well as preferences for the auditory system when participants completed the search task on an in-vehicle interface while performing a driving like-task (Jeon, Davison, Nees, Wilson, & Walker, 2009). Analogous results were found in a similar study done in a mid-fidelity simulator (Jeon et al., 2015). However, in that recent study no new measures were examined other than the higher-fidelity driving simulator, and the driving task was not realistic.

Gable et al. (2013) took this research further, using eye-tracking glasses, moving the search task from an infotainment interface to a mobile phone, and varying the driving task to be the standard lane-change driving task (Mattes, 2003). As seen in Figure 1, results showed decreased percent dwell time toward the primary (driving) task in the dual-task list-searching conditions, as compared to the driving-only condition. However, when the spindex cue was added to the interaction, participants' dwell time on the secondary task significantly decreased compared to the visuals-only interaction. These results were found with no significant differences seen in participants' performance of

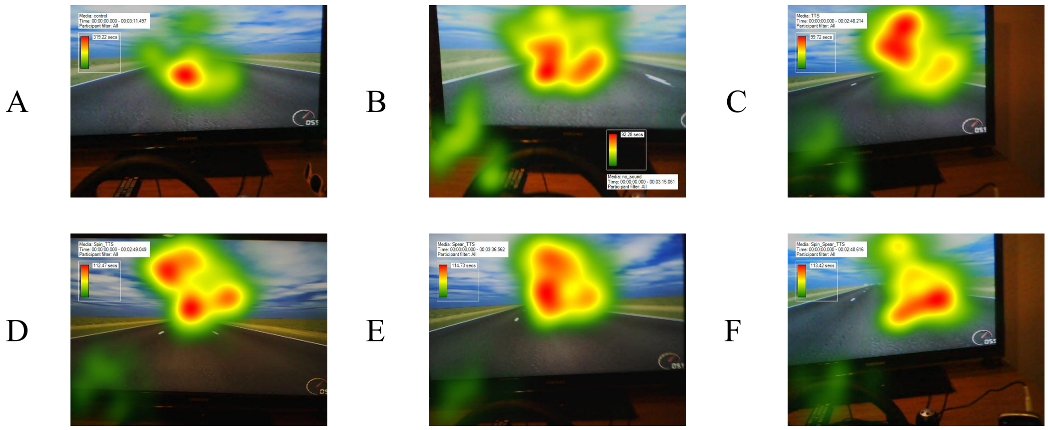


Figure 1. Heat maps of eye gaze fixation points from Gable et al. 2013 for left handed users. Conditions are ordered from left to right: (A) no search task, (B) visuals-only, (C) TTS only, (D) spindex+TTS, (E) spearcon+TTS, (F) spindex+spearcon+TTS. Green/yellow clouds in the bottom corner represent number of fixations spent outside of the primary task; the more yellow/red, the more time the eyes were focused in that area.

either the driving task or the secondary search task, and no significant workload differences between the spindex and visuals-only conditions as rated by the participants. The study did not, however, investigate the effects of these cues in a realistic driving scenario, gather physiological measures of cognitive load, nor attempt to predict the workload that the drivers would encounter. It also left questions open regarding the willingness to engage, when people decided to use the auditory cues, and if they might use them more in more visually demanding driving conditions.

Present Study

The current study aimed to extend the line of research regarding the outcomes of novices using advanced auditory cues on a mobile phone list search task while driving. In particular, it was a follow up to Gable et al. (2013) with a number of extended research questions. The primary extensions of this research included the use of a more realistic driving scenario, extended driving performance measures, the addition of physiological measures of cognitive load through heart rate and heart rate variability, the addition of a manipulation of willingness to engage via a high workload/low workload driving task variable, and the investigation into predicting differences in workload for different types of interfaces via the VACP scale. Prior measures of list-search task performance, visual behaviors, subjective workload, and preferences were again collected. The study focused on two dual task conditions of driving and searching for a song with visuals-only or with visuals plus spindex-TTS. It also included a baseline drive for each drive type/difficulty (city or highway) to investigate the effects of differing the drive types before the application of the secondary tasks and to determine if the drive types modulated performance or willingness to engage.

Hypotheses

The hypotheses of the current study are listed below. They are broken into the measures being used to investigate the hypotheses to allow for better flow through the following sections.

Secondary Task

H1 – Participants would have equal or better search task performance with the novel spindex-TTS cues than when using well practiced visuals-only interface as seen through mean search times, songs found, and accuracy. This was expected due to theoretical reasoning that participants would be able to continue to search for the desired song while keeping their visual attention on the road when using spindex-TTS cues, not forcing a time sharing procedure of the search task and driving that the visuals only condition would create. It was also expected as per previous research (Jeon et al., 2015).

H2 – Participants' search task performance via the same measures would improve with practice for the spindex-TTS condition, whereas the visuals-only interactions would not. This was expected due to previously high levels of abilities with the visuals-only interaction that participants will already have since they use that modality for searching everyday and have less room to improve as compared to interaction with the spindex-TTS interface, a result that has also been seen in previous work (Jeon et al., 2012).

H3 – Manipulation of the drive difficulty would increase participants' willingness to engage on the highway drive and this would be seen through more songs found in the easy highway drive than the difficult city drive. This hypothesis was made due to the expected balancing of workload and therefore suboptimal performance on the two tasks when the total workload increases, with more effort being given to the primary task and therefore causing the secondary task to receive less effort when the total workload increases and vice versa.

Driving Performance

H4 – Adding the spindex-TTS condition would have no negative effect on participants' driving performance as compared to the visuals-only condition, seen through lateral and longitudinal control metrics. This was hypothesized due to the expected lack of additional workload, if not the decreasing of workload with the application of the spindex-TTS cues, and has also been seen in previous work with primary task performance and these cues (Jeon et al., 2015).

Visual Behaviors

H5 – When using the spindex-TTS auditory cue, participants would have more dwell time on the primary driving task and lower glance frequency off of the driving task than the visuals-only interface. This hypothesis was made based on previous work (Gable et al., 2013) and the theoretical expectation that by applying auditory cues, participants would use those cues instead of their visual attention to search the list.

H6 – As experience with the spindex-TTS cues increases, visual behavior measures will show more reliance on the auditory modality of information gathering, manifesting itself through increased dwell time on the primary task and decreased glance frequency away from primary task as compared to the visuals-only interface. Similarly to H5, this was hypothesized due to the expectation of participants using the auditory cues instead of visual cues when auditory cues were made available, but also builds on that hypothesis and expects participants to become more comfortable and rely more on the auditory cues as they gain more experience with them.

H7 – Differences in willingness to engage would be seen between the drive difficulties through differences in participants' reliance on the auditory cue when driving the more difficult, city drive, manifested through interactions across drive type and condition, with more visual attention on the road in the difficult city drive than the highway conditions, particularly when using the spindex-TTS auditory cue. This hypothesis was made based on the theoretical implications of willingness to engage that by adding more workload to a drive individuals will be less willing to engage with an interface and this change in willingness will be seen through less visual interaction with an interface. It is expected to be particularly strong in the auditory conditions as people can rely even more on the auditory cues instead of visuals and would be expected to take advantage of that modality even more in these difficult conditions than the easy condition where the drivers can afford to use more of the visual workload for the non driving task.

Workload

H8 – Participants would report lower total workload via NASA-TLX when using the spindex-TTS cue as compared to the visuals-only interaction and no higher workload in the spindex-TTS condition for mental or physical workload. This was hypothesized due to both previous research (Jeon et al., 2015) and the theoretical implications of MRT that would suggest lower total workload when applying the demand over multiple modalities of input, i.e. over visual and auditory modalities in the spindex-TTS condition, instead of only visuals.

H9 – The city drive type would be found to have significantly higher total workload than the highway drive, as seen through the NASA-TLX ratings, and these differences would stay intact once the secondary tasks were added. This was hypothesized due to the expectation of the city drive having a higher level of workload than the highway drive based on the work previously done within the area of willingness to engage.

H10 – Participants would show lower levels of objective workload through lower levels of heart rate and heart rate variability when interacting with the spindex-TTS interface as compared to the visuals-only interface. Again the theoretical implications of MRT suggest lower workload and therefore this should also be seen through objective measures.

Perceived Performance and Preferences

H11 – Participants would assign higher values when using the spindex-TTS cue than the visuals-only condition on perceived performance on the search task and the driving task through ratings of being efficient at the secondary task, being effective at the driving task, and being a safe driver. In addition, participants would state that the sound was effective for the task and that it was helpful for the task. Previous work has shown participants rate the cues as helpful (Gable et al., 2013; Jeon et al., 2015) and the same is expected here.

H12 – Participants would report preferring the spindex-TTS condition to the visuals-only condition. Previous work has shown that participants prefer the spindex-TTS cue to a visuals-only condition (Jeon et al., 2015).

H13 – Participants would report being more comfortable using their cell phone or mp3 player in their car if it had spindex-TTS cues in it than their level of comfort using their current system of visuals-only in their car. This was expected due to previous findings of individuals preferring the spindex-TTS cue to a visuals-only cue as well as the perceived helpfulness, leading to this expected outcome.

VACP

H14 – The pattern of results seen in the VACP scores would match the pattern of results in the collected data. These patterns would be evident between total VACP score and NASA-TLX total score as well as between VACP modality scores and their associated collected measures such as the RTLX scores of mental and physical workload and eye tracking data. This was expected due to the findings of previous work (Mitchell, 2000) and that the measures would hold as a reliable tool in this context as well.

CHAPTER 2: METHOD

Participants

After removal of participants with low eye tracking quality (5 were below a criterion of 50% tracked) the remaining sample was composed of 26 students (12 male, 14 female; mean age = 20, SD = 1.7) from Georgia Tech. All participants were self-selected through a research database and received partial class credit for participating in the study. Participants were required to have had a valid driver's license in the United States for a minimum of 2 years to participate in the study. The resulting sample reported a mean of 4.3 (SD = 1.6) years driving. All participants in the study were also required to report normal or corrected to normal vision, hearing, and mobility, no history of attention deficit disorder (ADD) and were told to avoid performing any strenuous exercise or caffeine intake for two hours prior to the study (to avoid any artifacts of these activities in the heart rate measurements).

Apparatus

Primary Task

The driving task was the primary task, and was performed on a quarter cab version of the National Advanced Driving Simulator (NADS) MiniSim running software version 2.0. A simulator was chosen due to its ability to collect many measures of driving performance and to allow for control of the primary task variables as well as environmental variables, allowing for more consistent data than in a real car. The simulator was composed of three 42" plasma monitors to display the visuals to the participants, who were seated in a car seat and controlled the simulator using a steering wheel with force feedback, gas and brake pedals, and gear shifter. The system included a

2.1 audio system to present driving scenario related sounds and an LCD screen to display the instrument panel.

The driving tasks were created using the NADS Interactive Scenario Authoring Tool (ISAT). There were a total of six driving scenarios, three city drives, and three highways drives. The drive types were created in an attempt to manipulate driving difficulty and therefore influence willingness to engage via environmental factors, with the city drive being more difficult due to more turns, stops, and traffic throughout the drive. Note that it is not being claimed here that a drive simply being done in the city makes it more difficult than a drive being done on the highway, as the opposite can easily happen based on traffic and other factors, but instead were used to help modulate difficulty based on stereotypes of the driving scenarios and done in accordance with previous research. All six of the drives were following tasks where drivers were told to follow the vehicle they started behind at the distance initially seen between the two vehicles at the beginning of the drive before they started moving (50 feet, front bumper to rear bumper for both drive types). Each drive took a total of just over six minutes to complete, with the initial 20 seconds composed of being in a parked position to see the desired distance gap and starting up the vehicle to the driving speed (no secondary tasks were given during this start-up period). The three city and three highways drives were along the same respective city or highway route but had varying traffic for each drive. The lead vehicle in the drives also had different variations in its speed and lane deviation during each drive. In the city drives the lead vehicle's average speed was set to 30 miles per hour (MPH) with a randomized distribution based on a normal curve of 2.5 MPH, set to change every 10 seconds. In the highway drives the lead vehicle's average speed was set to 50 MPH with the same randomized distribution based on a normal curve of 2.5 MPH, set to change every 10 seconds. The lead vehicles in both the city and highway drives had small, random changes in lane deviation based on the ISAT software's standard setting based on a ramp model of deviation. The city drives had a number of turns, stops and curves along a 1-lane city road, whereas the highway drives all followed a fairly strait path down a 2-lane highway. Traffic was present in both drive types but the other vehicles never impeded the driver's view or following of the lead vehicle and never appeared to the driver to be on a collision course or present any hazard to the driver or lead vehicle.

Measures of driving performance included vehicle-based metrics of both lateral and longitudinal control. The longitudinal measures included mean and standard deviation of distance gap (front gap), distance gap – defined as per SAE J2944 as the distance (in feet) between the driver's front bumper and the lead car's rear bumper (SAE, 2013). The lateral control measures included mean and standard deviation of lateral lane position – defined as per J2944 as the lateral distance (in feet) from the longitudinal centerline of the driver's vehicle to the midpoint of the lane, and mean and standard deviation of SDLP – defined as per J2944 as the distribution of the lateral lane position (in feet) (SAE, 2013). During each drive, the simulator collected and stored these driving measures at a 60 Hz rate as well as any collisions and other driving measures not used in the analysis.

Secondary Task

The secondary task was a search task on a cell phone. The task was completed on a Google Nexus One HTC1 Android smartphone running Android OS version 2.3.6. The

3.75-inch resistive touch screen displayed a list of 150 popular songs from 2009. Participants interacted with the list through kinetic flicking, an interaction technique that has been shown to be highly un-optimized for driving, but continues to be a standard interaction mode for touch-screen phones and mp3 devices (Lasch & Kujala, 2012). Each participant was allowed to choose the hand he or she wanted to use for the search task before the study started. That arm was then placed on an armrest and was kept there for the duration of each drive. Participants used only the other hand to perform the driving task.

During the experimental conditions, after the initial start up of the vehicle, announcement of a song name through a set of speakers prompted participants to search for a song. Once they found the song they selected it by pressing on the song name on the phone screen. Announcements occurred at random intervals between 15 and 45 seconds after the last search was completed until the driving task was completed. Participants' performance on the task was measured through the average time to find a correct song, number of songs searched, number of correct songs selected, and accuracy of their search performance.

Auditory Cues

During one of the search task conditions, participants used a spindex cue (speech index), a type of advanced auditory cue, paired with text-to-speech (TTS) to assist in the search task. The spindex cue was a non-speech auditory cue based on the pronunciation of the first letter of each menu item (Jeon & Walker, 2011). The cue was based on a male voice and was interruptible as the participant was going down the list.

Workload

Subjective workload was measured through NASA-TLX after all driving scenarios. The participants performed the NASA-TLX on a resistive touch screen next to the simulator while referring to the definitions for each measure. As per the standard NASA-TLX procedure, participants rated each variable on a scale and then performed the paired comparisons, stating which factor was more important in their performance of the task (Hart & Staveland, 1988).

Objective workload was measured through the physiological measures of heart rate (HR) and heart rate variation (HRV). These were measured using the NeXus-10 physiological monitoring and biofeedback platform. This system transmitted data via the wireless Bluetooth transmitter attached to the body to a laptop running BioTrace software for storage and output of the data. An EXG Ground lead with a Micro-coax connector and two EXG Sensor leads (2 KHz bandwidth) with Brushed Aluminum medical grade ODU connectors connected participants to the Bluetooth interface via a modified lead II configuration as seen in Mehler et al. (2012). This configuration had the ground lead placed under the left clavicle, negative lead placed under the right clavicle, and the positive lead on the left lower side of the ribs. Prior to placing the silver/silver chloride pre-gelled disposable electrodes (Stens ecg/eeg electrode) in these locations to connect the leads via a snap connection, the skin was cleaned with isopropyl alcohol wipes.

Visual Behaviors

The visual behaviors of the participants were measured using Tobii eye-tracking glasses (version 1). The glasses used monocular infrared reflection to track wearers' pupil movement at 30 Hz. The device also recorded what the user was viewing with a front-facing scene camera so that gaze data can be overlaid onto the recorded visual scene for later data analysis. The measures of dwell time on the primary task – the sum of all fixations and saccades on the driving task – and glance frequency off the primary task – the number of glances to the driving task – were collected in this study as per SAE J2396 (SAE, 2000).

Other Measures

Participants completed three surveys over the duration of the study. Two of these surveys (condition questionnaires, Appendix A) were the same survey given once for each type of auditory cue; they asked the participants' opinions about their perceived performance on the cell phone search task and the driving task. A study questionnaire (Appendix B) was given to collect demographics such as age, gender, and years of experience driving, as well as the frequency with which they drive and interact with in-vehicle technologies such as infotainment interfaces and mobile phones. The study survey also asked whether they preferred having the spindex+TTS cue or having no auditory cue, if they were distracted during the research due to anything external to the study, and if they recognized any of the song names.

VACP

The VACP scale was employed in this study to explore the measure's ability to predict workload in an in-vehicle dual task situation. In particular, the focus of the exploration was on the scale's ability to measure and differentiate levels of the four mental resources that combine to create workload: visual, auditory, cognitive, and psychomotor workload. This breakdown of each piece of demand is of specific interest in this study due to the focus on auditory cues and MRT. The exploration of applying the

VACP scale in this manner is different from what it was created to be used for, as a method to inform models of workload, but the scale has been found to correlate with subjective workload ratings from participants in military applications as well as having good predictive validity (Mitchell, 2000).

Procedure

Experimental Session

Participants were self-selected through a database at Georgia Tech via an advertisement for the study (Appendix C). This advertisement ensured potential participants knew the criteria that were required to be in the study. Upon arrival at the study room, participants were first asked to review the criteria for the study and state if they met all the criteria. If they did not meet all the criteria they were not used as a participant and no credit was given. The participants that met the criteria completed consent forms (Appendix D) and reviewed the instructions of the study with the experimenter (Appendix E). Participants then completed the Georgia Tech Simulator Sickness Screening Protocol (GTSSSP) as described in Gable and Walker (2013). In short, the participants filled out a questionnaire regarding their current physical feeling, drove for two minutes in the simulator, and then filled out the same questionnaire. If during the drive the participants reported feeling sick or if after the drive the questionnaire revealed increase in sickness feeling they were released with credit being assigned to them. If participants did not display effects of simulator sickness they were then shown how to fit themselves with the heart rate monitor by the experimenter with a set of instructions and a diagram (Appendix F). The experimenter then left the room and the participants placed the monitor pads on themselves and the experimenter checked to make sure they were working once she re-entered the room. The Tobii glasses were then calibrated to the participant who was then seated in the simulator as seen in Figure 2, with the armrest on the side they requested.

Once participants were fitted with the physiological sensors they began the first block of the study. In total there were four experimental blocks: a city and a highway drive with the secondary search task using a visuals-only display; a city and a highway drive with the secondary search task using visuals and the spindex-TTS cue; and two baseline blocks of a city and a highway drive where the participant just drove without a secondary task. These six blocks were counterbalanced across participants using a Latin square. When participants began an experimental block, they were first introduced to the mobile phone list-searching task, and to the corresponding feedback being used in that

Figure 2. A photograph of the participant setup for the experiment. The participant is seated in the simulator wearing the Tobii eye trackers and the heart rate monitor (not visible) and is also holding the mobile phone used in the study where the list search visual interface can be seen.

block, if it was present. Participants had time to get used to the system and once they stated they were ready they began the driving task. During each drive the participants were instructed to follow the vehicle in front of them at the distance seen at the beginning of the drive and to drive the simulated vehicle as they would any normal car, attempting to keep good lane keeping and driving safely. If it was an experimental block, the phone periodically displayed and read out the song name they needed to find. In an attempt to control factors that might influence willingness to engage, participants were instructed to focus 80 percent of their resources on the primary (driving) task and 20 percent on the secondary (list-search) task.

After each block, participants filled out the NASA-TLX, and after their second drive for each search task condition they completed the corresponding condition questionnaire. During this time the participants' physiological measures had a chance to revert back to baseline and when the participants were ready to begin the next condition they let the experimenter know. Participants continued to perform the remaining blocks (with breaks) and when finished with all conditions they completed the final questionnaire. Once that was completed they removed the heart rate monitor and eye-trackers, followed by a debriefing on the study (Appendix G), their release from testing, and the assignment of their research credit.

VACP Ratings Collection

The VACP ratings were collected separately from the experimental data collection. First, a task analysis was created for the three tasks: driving (Appendix H), searching for a song on the phone with no auditory cues (Appendix I), and searching for a song on the phone with spindex+TTS cues (Appendix J). Five experts in the area of human factors were then asked to assess and rate the workload on each of the four mental resources for the three tasks using the VACP scale. Each rater was asked to assign values for the sub-tasks within the three task analyses. They did so on an Excel sheet while referring to a hard copy of the task analysis and the VACP workload resource unique seven-point scales and anchors (Appendix K). The order of the task analysis each expert rated was randomized using a Latin square. The averages of each implementable task were then calculated within the Excel sheet across the modalities, across each sub-task. After rating all of the steps in one task, raters moved on to rating the second and third tasks.

Data Organization, Design, and Analysis

This study was a 2 x 2, within subjects, full factorial design with baseline conditions for each drive type that were compared against each other to see the difference between the two drive types before applying a secondary task. These baseline measures were first analyzed via a paired t-test across the high and low workload (city and highway respectively) drives. These analyses then served as a point of information during discussion of the results into whether differences that were expected between the two drive difficulties (drive types) were created or not. All of the experimental data were analyzed via two-way full factorial within-subjects ANOVAs with Huynh-Feldt corrections for sphericity. This approach allowed for the investigation of main effects between the two experimental conditions (visuals-only, spindex-TTS), and the two drive types (city, highway), as well as the interactions between conditions and drive types. All analyses in this research were done at an alpha of 0.05 with Bonferroni corrections being made when investigating significant interactions, setting alpha to .0125.

Preferences and questionnaires were also analyzed, with the condition questionnaires being investigated via separate paired t-tests where applicable. The VACP measures inter-rater reliability was measured using intraclass correlations and then simply compared using descriptive statistics to look at patterns as compared with subjective workload and visual behaviors data.

Prior to analysis, the driving data were pulled from the simulator database and filtered through the MiniSim MatLab interface to output the variables desired. Secondary task performance data were collected and saved on the phone for each condition, and the paper forms containing the demographics and preference information were input into a database. The eye-tracking recordings were separated in the Tobii software into conditions, and the areas of interest (AOIs) were placed over the driving task for all conditions. These AOIs allowed for glance frequency and dwell time data to be determined for the primary task. The time-stamps of the heart rate data were used to separate them into the correct conditions for summary calculations in MatLab. In this summary procedure, all values more than two times the standard deviation of the mean for each condition for each participant were removed and then the averages were calculated of that block. NASA-TLX data were output by the program for each condition. Before analysis all experimental data were analyzed for outliers and outliers were removed. These outliers were defined as any data value less than or greater than the outlier range, calculated by determining the interquartile range (IQR) multiplied by 1.5 and then subtracting this value from Q1 and adding the value to Q3. VACP was analyzed for inter-rater reliability and then combined to create condition workload scores.

CHAPTER 3: RESULTS

Search Task Performance

Four measures were used to determine search task performance: mean search time for the songs correctly selected, mean number of songs searched, mean number of correct songs selected, and accuracy. The data were analyzed on these four measures by drive type as well as across block order (i.e., first time with spindex or second) to investigate practice effects. The means and standard deviations of these measures can be seen in Tables 1 and 2 and are analyzed in the subsequent sections.

Table 1

Search Task Performance Descriptive Data for the Conditions by Drive Type

	City		Highway		Condition Mean	
Condition	M	SD	M	SD	M	SD
Mean Search Time						
Visuals-only	36.74	6.14	35.92	3.25	36.33	4.17
Spindex-TTS	37.45	5.60	37.37	7.18	37.41	3.71
Drive Type Mean	37.10	4.71	36.65	4.29	-	-
Mean # Songs Searched						
Visuals-only	9.21	1.34	9.25	0.99	9.23	1.00
Spindex-TTS	8.54	1.18	8.83	1.37	8.69	1.13
Drive Type Mean	8.88	1.15	9.04	0.97	-	-
Mean # Correct Songs Found						
Visuals-only	9.04	1.43	9.39	0.84	9.22	0.97
Spindex-TTS	8.22	1.17	8.00	1.71	8.11	1.28
Drive Type Mean	8.63	1.15	8.70	1.06	-	-
Mean Accuracy						
Visuals-only	100	0.00	100	0.00	100	0.00
Spindex-TTS	97.03	7.03	91.88	10.03	94.50	6.58
Drive Type Mean	98.50	3.49	95.90	5.03	-	-

Note. Search time is reported in seconds and accuracy is reported in percent.

Mean Search Time for Drive Type

As seen in Figure 3, the two-way ANOVA revealed that participants had no significant differences in mean search time across the two experimental conditions, $F(1.00, 22.00) = 2.32$, $p = .142$, $\eta 2 = .095$. The main effect of drive type was also found to be non significant, F(1.00, 22.00) = 0.15, p = .699, η2 = .007, as was the interaction between drive type and condition, F(1.00, 22.00) = 0.08, p = .777, η2 = .004.

Mean # Songs Searched for Drive Type

Figure 4 displays the results from the two-way ANOVA for songs searched, which revealed a significant main effect of condition with the participants searching for more songs in the visuals only condition, F(1.00, 23.00) = 10.15, p = .004, η2 = .306. No

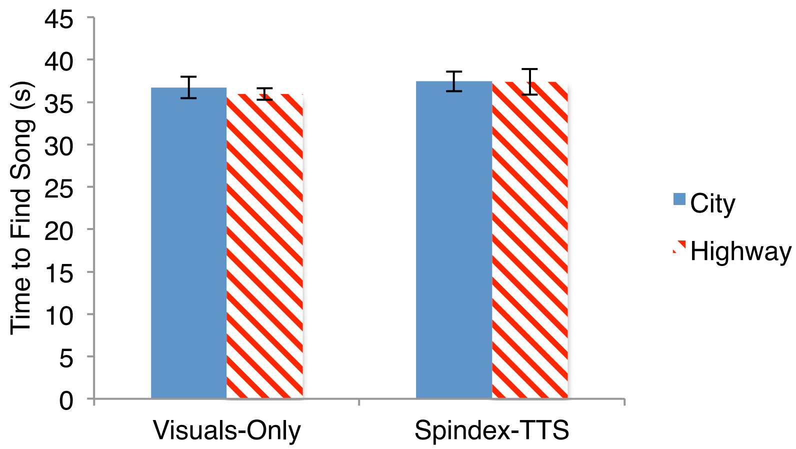


Figure 4. Graph of mean time to find a song across drive type for the two experimental conditions. Higher bars mean more time to find a song. Standard error is shown via error bars.

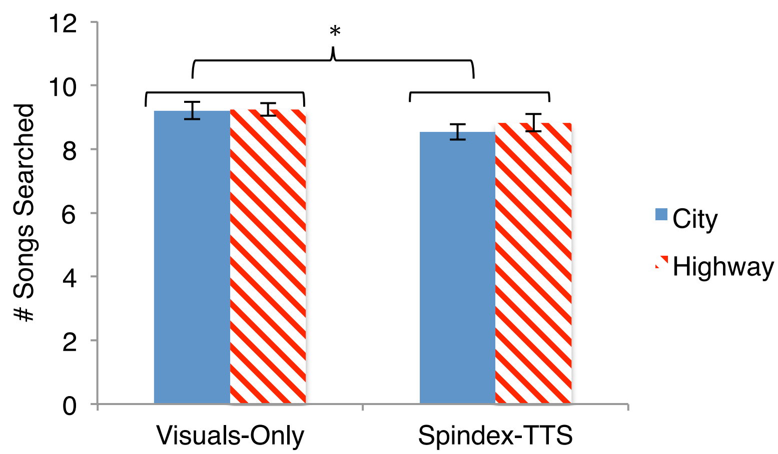


Figure 3. Graph of mean number of songs searched across drive type for the two experimental conditions. Higher bars mean more songs found on average per condition. Standard error is shown via error bars and significant differences are marked with “*”.

significant main effect of drive type, $F(1.00, 23.00) = 1.00$, $p = .328$, $\eta 2 = .042$, or interaction was found, $F(1.00, 23.00) = 0.44$, $p = .514$, $\eta 2 = .019$.

Mean # Correct Songs Found for Drive Type

The analysis of number of correct songs selected revealed a significant main effect of condition with the participants finding more songs in the visuals only condition, $F(1.00, 22.00) = 24.41$, $p < .001$, $\eta 2 = .526$. No significant differences in songs found were identified for the main effect of drive type, $F(1.00, 22.00) = 0.11$, $p = .744$, $\eta 2 = .005$, or interactions, $F(1.00, 22.00) = 1.93$, $p = .178$, $\eta 2 = .081$. These values are visualized in Figure 5.

Mean Accuracy for Drive Type

Figure 6 displays participants' accuracy on the search task across the two drive types. Accuracy of the participants was found to be significantly higher during the visuals only condition, $F(1.00, 15.00) = 10.45$, $p = .006$, $\eta 2 = .411$. No differences in accuracy were found in the main effect of drive type, $F(1.00, 15.00) = 3.79$, $p = .071$, $\eta 2 = .202$, or in the interaction, $F(1.00, 15.00) = 3.79$, $p = .071$, $\eta 2 = .202$.

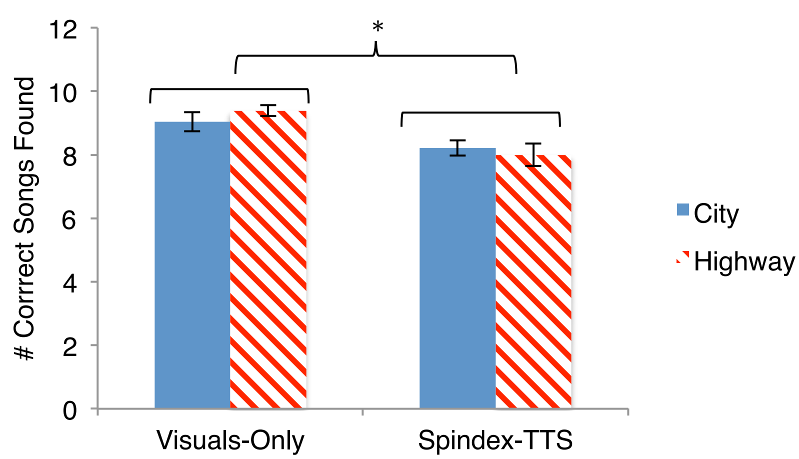


Figure 5. Graph of number of correct songs found across drive type for the two experimental conditions. Higher bars mean more songs found on average per condition. Standard error is shown via error bars and significant differences are marked with "*".

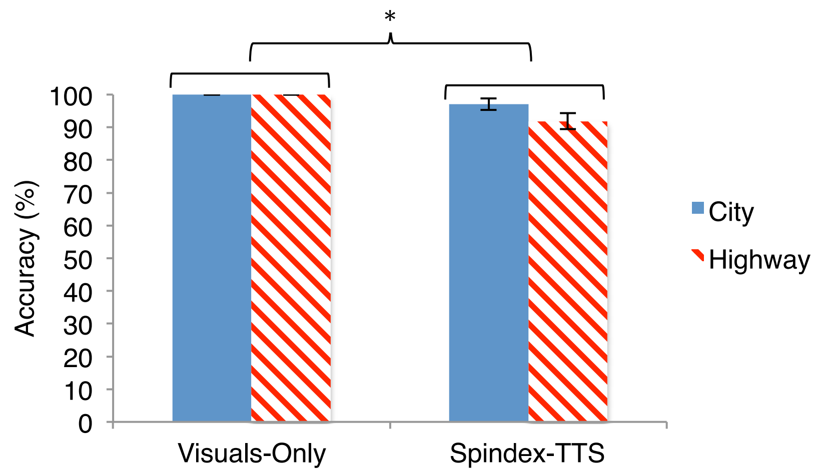


Figure 6. Graph of song finding accuracy across drive type for the two experimental conditions. Higher accuracy is visualized with higher bars. Standard error is shown via error bars and significant differences are marked with "*".

Table 2

Search Task Performance Descriptive Data for the Conditions by Block Order

	First Block		Second Block		Condition Mean	
Condition	M	SD	M	SD	M	SD
Mean Search Time						
Visuals-only	37.40	5.94	35.64	3.88	36.52	4.26
Spindex-TTS	38.96	4.14	36.54	4.16	37.75	3.44
Search Block Mean	38.18	4.62	36.09	3.21	-	-
Mean # Songs Searched						
Visuals-only	9.00	1.32	9.40	0.96	9.20	0.99
Spindex-TTS	8.48	1.12	8.84	1.41	8.66	1.12
Search Block Mean	8.74	1.13	9.12	1.01	-	-
Mean # Correct Songs Found						
Visuals-only	8.88	1.33	9.42	1.10	9.15	1.01
Spindex-TTS	7.92	1.41	8.29	1.46	8.10	1.25
Search Block Mean	8.40	1.21	8.85	1.10	-	-
Mean Accuracy						
Visuals-only	100	0.00	100	0.00	100	0.00
Spindex-TTS	97.00	6.42	93.40	8.31	95.2	3.69
Search Block Mean	98.50	3.10	96.70	4.26	-	-

Note. Search time is reported in seconds and accuracy is reported in percent.

Mean Search Time for Block Order

When investigating practice effects on the search tasks the effect of condition was still not significant for the measure of mean search time, $F(1.00, 20.00) = 2.92$, $p = .102$, $\eta 2 = .122$. There was a difference however, in the order that participants experienced the conditions with the participants being significantly faster at finding songs during their second testing block with that condition, $F(1.00, 20.00) = 6.50$, $p = .019$, $\eta2 = .236$. No significant interaction was found, $F(1.00, 20.00) = 0.25$, $p = .626$, $\eta2 = .021$.

The two-way ANOVA for songs searched revealed the significant main effect of condition with the participants searching for more songs in the visuals only condition, $F(1.00, 24.00) = 10.96$, $p = .003$, $\eta2 = .314$. No significant main effect of block order, $F(1.00, 24.00) = 3.67$, $p = .068$, $\eta2 = .132$, or interaction was found, $F(1.00, 24.00) = 0.02$, $p = .885$, $\eta2 = .001$.

Mean # Correct Songs Found for Block Order

The analysis of correct songs selected for block order found the same significant main effect of condition as found in the drive type analysis of participants finding more correct songs in the visuals only condition., $F(1.00, 23.00) = 21.42$, $p < .001$, $\eta2 = .482$. The main effect of order, $F(1.00, 23.00) = 3.75$, $p = .065$, $\eta2 = .140$, and the interaction, $F(1.00, 23.00) = 0.30$, $p = .590$, $\eta2 = .013$, were both not significant.

Mean Accuracy for Block Order

As with the songs found measure, the same significant difference was found in accuracy for the main effect of condition, $F(1.00, 14.00) = 8.48$, $p = .011$, $\eta2 = .377$, but no differences were found across the main effect of order, $F(1.00, 14.00) = 3.41$, $p = .086$, $\eta2 = .196$, or in the interaction, $F(1.00, 14.00) = 3.41$, $p = . 086$, $\eta2 = .196$.

Search Task Performance Results Summary

Results of the analyses revealed no main effect of condition or drive type on search time, but when investigating practice effects it was found that participants were faster during their second interaction with the interface by an average of about 2 seconds. The main effect of condition was found to be significant with the participants' performance on the visuals-only interface having a higher average number of songs searched (0.5 more songs searched over 6 minutes), higher average number of correct songs found (1.11 more songs found over 6 minutes), and higher accuracy on the search task than with the spindex-TTS condition. These results imply that the novice users were worse at the searching task in regards to accuracy with the spindex-TTS condition and that they searched for slightly more songs with the no-sound condition, despite a lack of difference in search time. The result of a practice effect seems to show participants' performance increased slightly in the second block as compared to the first over the two conditions but no interactions.

Driving Task Performance

Four measures of driving performance were collected and analyzed in the current study: mean and standard deviation of distance gap, lateral lane position, and SDLP. The values for these measures can be seen in Tables 3 and 4 and analyses are reported in the following sections.

Table 3

Longitudinal Driving Performance Metrics for the Conditions by Drive Type

	City		Highway		Condition Mean	
Condition	M	SD	M	SD	M	SD
Mean Distance Gap						
Baseline	77.91	23.82	92.78	24.10	85.34	22.33
Visuals-only	85.98	20.20	108.67	26.35	97.33	21.23
Spindex-TTS	82.15	11.46	107.82	26.84	94.99	17.51
Drive Type Mean	82.01	14.74	103.09	21.91	-	-
SD of Distance Gap						
Baseline	27.40	11.33	26.31	11.48	26.86	9.28
Visuals-only	29.27	9.46	37.15	15.41	33.21	10.50
Spindex-TTS	29.20	7.31	35.43	12.88	32.31	8.90
Drive Type Mean	28.62	6.59	32.96	10.16	-	-

Note. The values reported in the table are in feet.

Mean Distance Gap

The baseline t-tests revealed that the baseline highway drive had a larger mean distance gap, $t(23) = 4.195$, $p < .001$. Although the two-way ANOVA between the two experimental conditions revealed no significant main effect of condition, $F(1,23) = 0.65$, $p = .430$, $\eta 2 = .027$, there was a significant main effect of drive type with the highway drive having a larger mean distance gap, $F(1,23) = 51.07$, $p < .001$, $\eta 2 = .689$, showing the continuation of the difference seen in the baseline conditions. No significant interaction was present, $F(1,23) = 0.32$, $p = .576$, $\eta 2 = .014$. These relationships are visualized in Figure 7.

SD Distance Gap

The baseline comparisons of SD of distance gap revealed no significant differences between the two baseline drives, $t(21) = 0.384$, $p = .705$. The two-way ANOVA revealed no differences in the main effect of condition, $F(1,21) = 0.181$, $p = .675$, $\eta 2 = .009$, but did show a significantly higher SD of distance gap in the highway

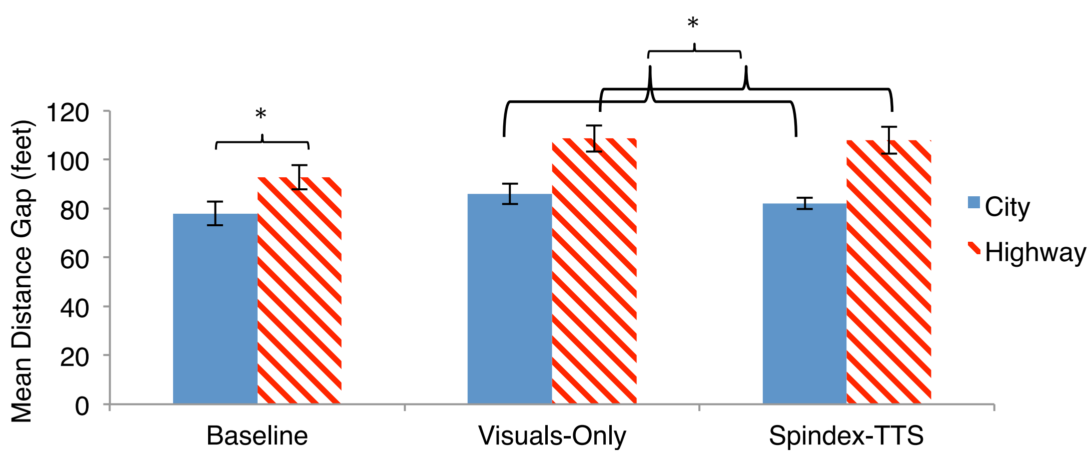


Figure 7. Graph of mean distance gap across drive types for the two experimental conditions and baseline. Larger gap means worse driving performance. Standard error is shown via error bars and significant differences are marked with "*".

drive type for the experimental conditions, $F(1,21) = 13.66$, $p = .001$, $\eta 2 = .394$, suggesting that the application of the secondary task created a difference in this measure.

Table 4

Lateral Driving Performance Metrics for the Conditions by Drive Type

	City		Highway		Condition Mean	
Condition	M	SD	M	SD	M	SD
Lateral Lane Position						
Baseline	0.82	0.16	0.89	0.25	0.85	0.19
Visuals-only	0.84	0.15	0.85	0.18	0.85	0.14
Spindex-TTS	0.84	0.19	0.94	0.22	0.89	0.16
Drive Type Mean	0.84	0.14	0.89	0.18	-	-
SD of Lane Position (SDLP)						
Baseline	0.64	0.09	0.64	0.11	0.64	0.07
Visuals-only	0.69	0.10	0.67	0.11	0.68	0.08
Spindex-TTS	0.67	010	0.70	0.14	0.69	0.10
Drive Type Mean	0.67	0.06	0.67	0.10	-	-

Note. The values reported in the table are in feet.

Lateral Lane Position

The baseline comparison revealed no significant differences between the baseline drives, $t(24) = 1.79$, $p = .086$. The two-way ANOVA also revealed no significant main effects of condition, $F(1,24) = 2.46$, $p = .130$, $\eta 2 = .093$, drive type, $F(1,24) = 1.58$, $p = .221$, $\eta 2 = .062$, or interaction, $F(1,24) = 3.42$, $p = .077$, $\eta 2 = .125$.

Standard Deviation of Lane Position

Visualizations of the SDLP can be seen in Figure 8. No differences between the two baseline drives were found, $t(23) = 0.18$, $p = .858$. No differences were found in the experimental analysis of SDLP either, with the ANOVA showing no main effects for condition, $F(1,23) = 0.14$, $p = .708$, $\eta 2 = .006$, drive type, $F(1,23) = 0.17$, $p = .687$, $\eta 2 = .007$, or interaction, $F(1,23) = 1.62$, $p = .216$, $\eta 2 = .066$.

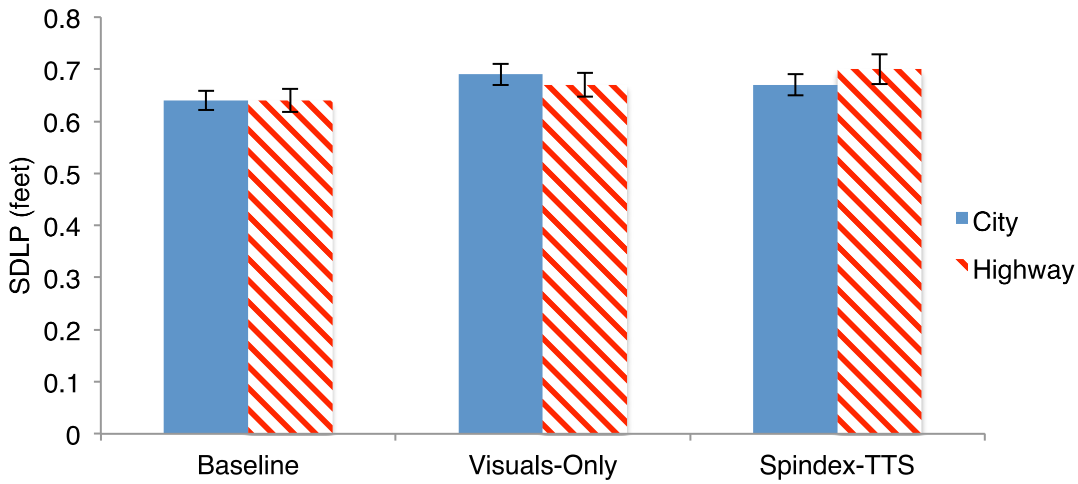


Figure 8. Graph of the standard deviation of lane position (SDLP) across drive type for the two experimental conditions and baseline. More deviation is representative of worse driving performance. Standard error is shown via error bars.

Driving Task Performance Results Summary

Results of the analyses of the driving data showed no significant main effects of interface condition. There were however, results regarding the baseline condition and main effects of drive type with the baseline t-test analysis and the ANOVA showing participants had a larger mean distance gap on the highway than the city drive. There was also a significant main effect of drive type for standard deviation of distance gap, with higher standard deviation occurring in the highway condition. These results imply that no differences in driving performance were present between the experimental conditions and while differences were there for the distance gap measures between the drive types, this would be expected between the two drive types due to speed differences.

Visual Behaviors

The visual behavior data of dwell time and glance frequency on the driving task can be seen in Table 5 and analyses in the sections following. It should be noted that while block order effects were investigated, no differences were seen in the visual behavior measures and these analyses were kept out of this document for brevity.

Table 5

Visual Behavior Descriptive Data for the Conditions by Drive Type

	City		Highway		Condition Mean	
Condition	M	SD	M	SD	M	SD
Dwell Time						
Baseline	300.45	41.23	301.23	39.03	300.84	36.36
Visuals-only	233.04	33.55	230.16	40.57	231.60	35.35
Spindex-TTS	250.74	51.76	238.00	42.97	244.37	43.60
Drive Type Mean	261.41	33.56	256.46	37.68	-	-
Glance Frequency						
Baseline	8.87	7.14	14.00	10.16	11.44	7.49
Visuals-only	42.57	23.13	42.78	21.79	42.67	21.42
Spindex-TTS	35.57	18.87	36.00	21.62	35.78	19.21
Drive Type Mean	29.00	13.35	30.93	15.66	-	-

Note. Values of time are in seconds.

Dwell Time

The baseline comparison found no difference in the total time participants' had their eyes on the road between the baseline drives, t(23) = 0.11, p = .911. The two-way ANOVA found a significant difference between the two experimental conditions with participants spending more total time with their eyes on the road in the spindex-TTS condition, F(1,25) = 5.40 p = .029, η2 = .178. The other main effect of drive type was not significant, F(1,25) = 3.77, p = .063, η2 = .131, and no significant interaction was found, F(1,25) = 0.78, p = .384, η2 = .030. A graph displaying this increased dwell time with the spindex-TTS interaction can be seen in Figure 9, and heat maps of dwell time on driving can be seen in Figure 10.

Glance Frequency

The baseline t-test revealed that fewer visits to the primary task occurring in the city drive, meaning the drivers took their eyes off the road more often in the highway baseline drive, t(22) = 2.69, p = .013. The two-way experimental ANOVA revealed a

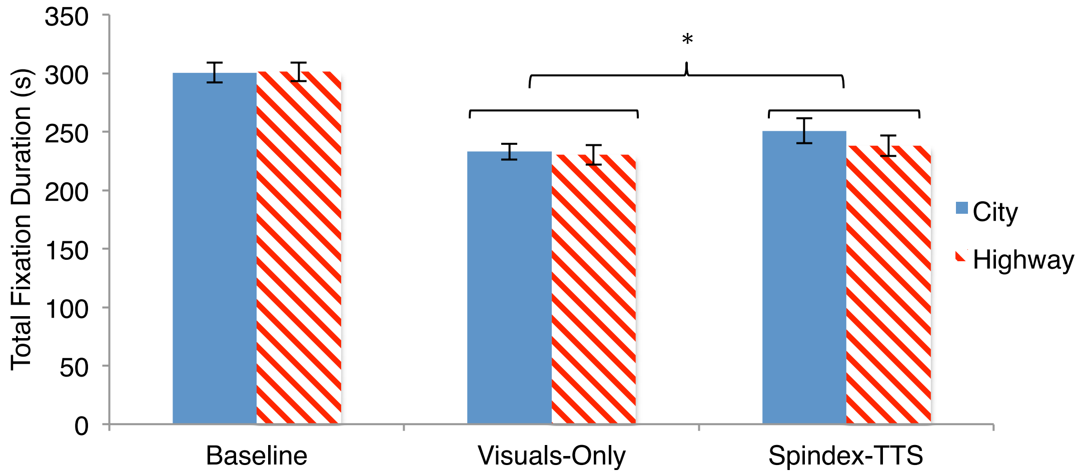


Figure 10. Graph of the dwell time on the primary task across drive type for the two experimental conditions and baseline. Higher bars represent more time eyes on the road. Standard error is shown via error bars and significant differences are marked with "*".

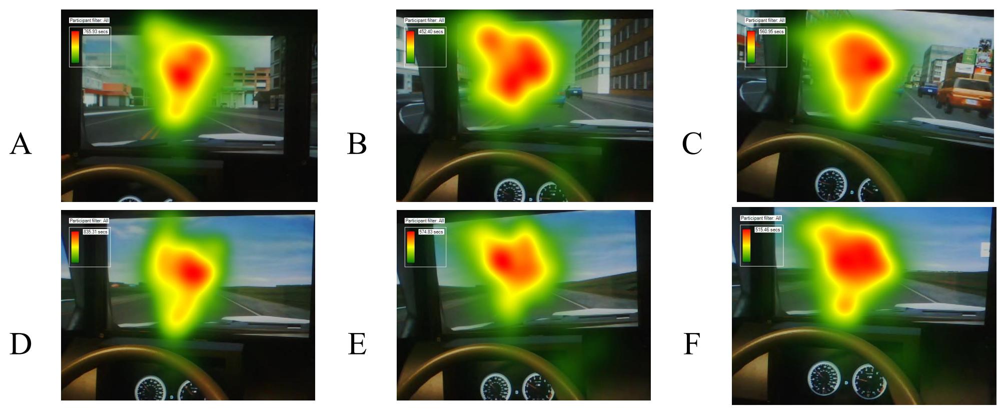


Figure 9. Heat maps of eye gaze fixation points for participants during the six drives. Drives are ordered from left to right: (A) city no search task, (B) city visuals-only, (C) city spindex-TTS, (D) highway no search task, (E) highway visuals-only, (F) highway spindex-TTS. Green/yellow clouds in the bottom corner represent number of fixations spent outside of the primary task; the more yellow/red, the more time the eyes were focused in that area.

significant main effect of condition with spindex-TTS having a lower number of visits to the primary task, therefore meaning they took their eyes off the driving task fewer times, $F(1,22) = 5.58$ $p = .027$, $\eta 2 = .202$. The main effect of drive type was not significant meaning the difference seen in the baseline conditions was not continued once the secondary task was applied, $F(1,22) = 0.02$, $p = .888$, $\eta 2 = .001$, and no interaction was found, $F(1,22) = 0.01$, $p = .946$, $\eta 2 = .000$. Figure 11 depicts these relationships.

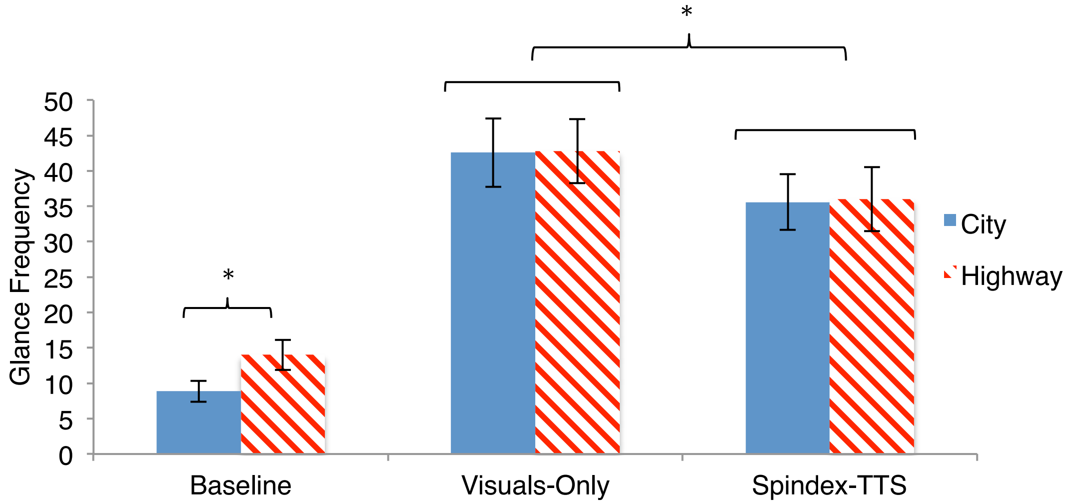


Figure 11. Graph of the glance frequency on the primary task across drive type for the two experimental conditions and baseline. Higher bars mean more glances away from the road. Standard error is shown via error bars and significant differences are marked with "*".

Visual Behavior Results Summary

The results of the visual behaviors analyses revealed that participants had more total time eyes on the road (dwell time) and less glances off of the road to the search task (glance frequency) in the spindex-TTS condition drives than the visuals-only condition drives. It was also found that in the baseline drives participants had a significantly higher glance frequency on the highway than the city, but this effect seemed to go away once the secondary task was added. These results point to spindex-TTS condition having lower visual workload than the visuals-only condition and the highway baseline drive having less visual workload than the baseline city.

Subjective Workload

The averages for NASA-TLX total workload as well as the raw measures of interest of mental and physical workload for each condition and drive type can be seen in Table 6 and the analyses are discussed in the subsequent sections.

Table 6

NASA-TLX Score Descriptive Values for Total, Mental, and Physical Workload

Condition	City M	City SD	Highway M	Highway SD	Condition Mean M	Condition Mean SD
Total Workload						
Baseline	38.26	16.17	34.00	16.20	36.13	14.92
Visuals-only	53.42	18.10	52.35	17.77	52.89	17.32
Spindex-TTS	50.00	17.68	47.19	16.82	48.60	16.19
Drive Type Mean	47.23	15.04	44.51	15.08	-	-
Mental Workload						
Baseline	27.31	17.04	25.77	17.70	26.54	16.09
Visuals-only	54.42	23.55	52.50	24.83	53.46	22.84
Spindex-TTS	52.50	20.99	46.15	21.13	49.33	19.86
Drive Type Mean	44.74	17.99	41.47	18.48	-	-
Physical Workload						
Baseline	28.54	17.84	19.79	13.06	24.17	12.76
Visuals-only	31.87	20.58	31.88	21.76	31.88	20.43
Spindex-TTS	31.46	19.25	28.75	16.44	30.10	16.25
Drive Type Mean	30.63	14.78	26.81	14.69	-	-

Total Workload

The comparison of the two baseline conditions was found to not be significant in regards to total subjective workload, $t(25) = 1.73$, $p = .095$. The two-way ANOVA for the experimental conditions revealed a significant main effect of condition with the spindex-TTS condition having lower workload than the visuals-only condition, $F(1.00, 25.00) = 5.52$, $p = .027$, $\eta2 = .181$. No significant main effect of drive type, $F(1.00, 25.00) =1.44$, $p = .241$, $\eta2 = .055$, and no significant interaction between condition and drive type were found, $F(1.00, 25.00) = 0.41$, $p = .528$, $\eta2 = .016$. These results are visualized in Figure 12.

Mental Workload

The mental workload means and standard deviations can be seen in Figure 13. The baseline comparison revealed no significant difference in mental workload between the two baseline drives, $t(25) = 0.60$, $p = .555$. The two-way ANOVA for the experimental conditions by drive type revealed no significant main effect of condition, $F(1.00, 25.00) = 2.44$, $p < .131$, $\eta 2 = .089$, but did reveal a significant main effect of drive type with higher mental workload being reported in the city drive than the highway, $F(1.00, 25.00) = 4.34$, $p = .048$, $\eta 2 = .148$. No significant interaction between condition and drive type was found, $F(1.00, 25.00) = 1.02$, $p = .321$, $\eta 2 = .039$.

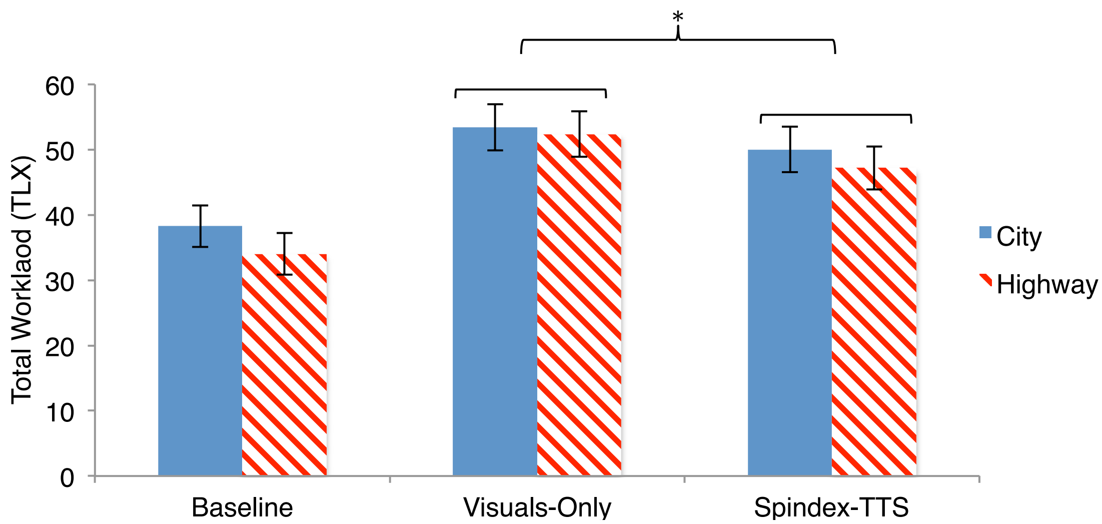


Figure 13. Graph of total workload via NASA-TLX across drive type for the two experimental conditions and baseline. Higher bars represent more workload. Standard error is shown via error bars and significant differences are marked with "*".

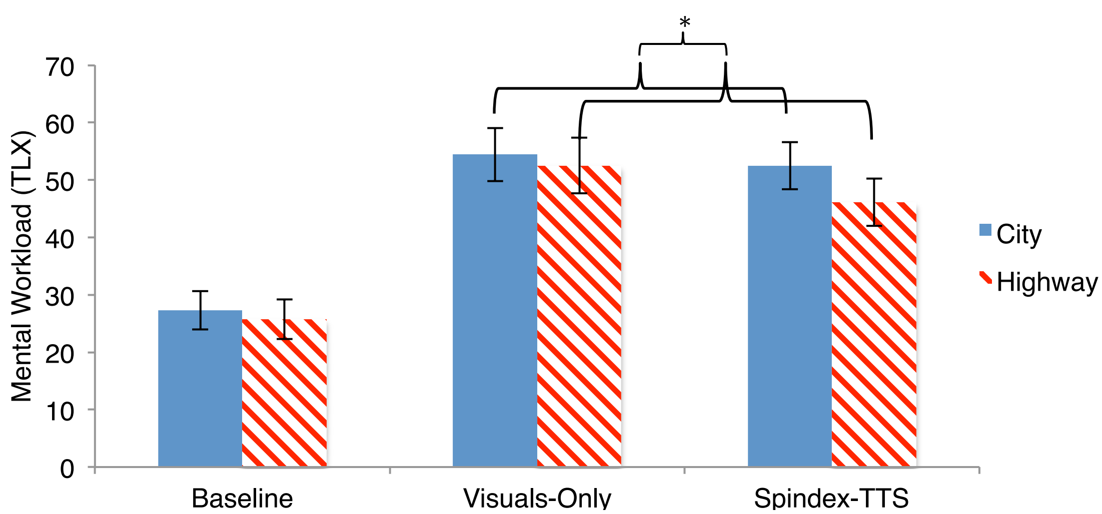


Figure 12. Graph of mental workload via NASA-TLX across drive type for the two experimental conditions and baseline. Higher bars represent more workload. Standard error is shown via error bars and significant differences are marked with "*".

Physical Workload

The baseline t-test revealed a significant difference between the two baseline drives, with the city drive proving to have higher physical workload than the highway, $t(24) = 2.37$, $p = .026$. The two-way ANOVA then revealed no significant difference between the two conditions, $F(1,25) = 0.44$ $p = .516$, $\eta2 = .019$, no main effect of drive type, $F(1,25) = 0.39$, $p = .537$, $\eta2 = .017$, and no interactions, $F(1,25) = 0.70$, $p = .412$, $\eta2 = .029$. This implies that by adding the secondary task the difference between the drive types is minimized. The values of physical workload across the conditions and drive types can be seen in Figure 14.

Subjective Workload Results Summary

Results of the subjective workload analyses show that participants had lower total subjective workload when interacting with the secondary task with the addition of spindex-TTS cues than in a visuals-only interaction. Drive type differences of lower subjective mental workload in the highway drive after the secondary task was applied are could infer some extra workload available in the highway drives. A significant difference

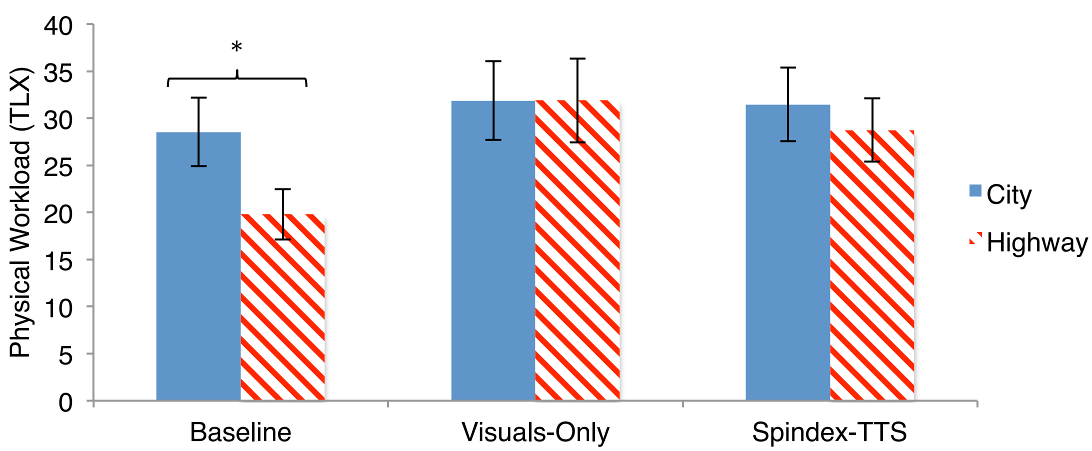


Figure 14. Graph of physical workload via NASA-TLX across drive type for the two experimental conditions and baseline. Higher bars represent more workload. Standard error is shown via error bars and significant differences are marked with "*".

between the two baseline drives in regards to physical workload, with the city being rated as higher prior to the application of the secondary task in the city is not surprising due to the need to make more maneuvers during the task but that difference did hold once the secondary task was applied.

Objective Workload

The values for heart rate and heart rate variability across the drives can be seen in Table 7 and the analyses are reported in the sections below.

Table 7

HR and HRV Descriptive Values

	City		Highway		Condition Mean	
Condition	M	SD	M	SD	M	SD
Average HR						
Baseline	75.22	6.91	74.84	7.32	75.03	7.00
Visuals-only	75.28	6.88	74.24	8.24	74.76	7.34
Spindex-TTS	74.98	7.32	74.44	6.91	74.71	7.00
Drive Type Mean	75.16	6.86	74.50	7.09	-	-
Average HRV						
Baseline	55.84	17.33	54.83	16.43	55.34	15.95
Visuals-only	54.86	15.18	55.01	15.49	54.94	14.78
Spindex-TTS	53.38	15.61	54.42	15.61	53.90	15.04
Drive Type Mean	54.69	15.10	54.75	15.01	-	-

Note. HR is reported in beats per minute.

Heart Rate

The comparison of highway and city baseline drives revealed no significant differences, $t(18) = 0.63$, $p = .537$. The two-way ANOVA revealed no significant main effect of condition, $F(1,18) = 0.01$ $p = .938$, $\eta2 < .001$, drive type, $F(1,18) = 2.47$ $p = .133$, $\eta2 = .121$, or a significant interaction, $F(1,18) = 0.20$ $p = .662$, $\eta2 = .011$.

Heart Rate Variability

The baseline t-test was also found to not be significant, $t(22) = 0.44$, $p = .667$. The two-way ANOVA revealed no significant main effect of condition, $F(1,22) = 0.69$ p

= .415, $\eta 2$ = .030, drive type, $F(1,22) = 0.17$ $p = .683$, $\eta 2 = .008$, or a significant interaction, $F(1,22) = 0.17$ $p = .683$, $\eta 2 = .008$.

Objective Workload Results Summary

No subjective workload differences were found in the statistical analyses performed within this study. Both HR and HRV were found to have large standard deviations and the values gathered during the study were also not showing large differences. Possible reasons for this are considered in the discussion section.

Subjective Performance and Preferences

Condition Questionnaire

The ratings of the two dual task conditions of visuals-only interaction and spindex-TTS that were gathered using the 6-point Likert-like condition questionnaire (Appendix A) and the results of the paired t-test analyses performed can be seen in Table 8. Only three of the questions were analyzed statistically, as the reference to the "sound" in the other three questions ("How effective was this sound for the task?"; "How functionally helpful was the sound?"; and "How annoying was this sound?") is not applicable to the visuals-only interaction. From these analyses it is clear that participants thought spindex improved their performance on the tasks. When asked how effective the participants were at the search task, they rated themselves significantly higher at the search task on the 6-point scale when using the Spindex-TTS interface as compared to the visuals only interface. When rating their own driving performance participants stated they were a significantly more effective driver at the task given to them when using the spindex-TTS cue than when using the visuals only interface. The participants also rated themselves as being a safer driver overall when they were using the spindex-TTS cue than when using the visuals only interface. The trends of ratings related to the “sound” in the spindex-TTS condition showed that participants thought it was effective, functionally helpful, but somewhat annoying.

Table 8

Descriptive Values of Responses to the Condition Questionnaire

	Visuals-only		Spindex-TTS		Analysis	
	M	SD	M	SD	t	p
Question						
How effective were you at the search task?	3.65	0.98	4.54	0.90	4.21	< .001
How effective were you at the driving task?	3.35	0.98	4.12	0.99	3.95	.001
How safe of a driver were you during the driving task?	2.96	1.15	3.50	1.03	3.89	< .001
How effective was this sound for the task?	1.73	1.00	4.77	1.24	-	-
How functionally helpful was this sound?	1.5	0.76	4.89	1.07	-	-
How annoying was this sound?	1.46	0.90	3.21	1.39	-	-

Note: Answers were on a 1-6 Likert-like scale from least to most (n=26). Analyses were not performed for the last three questions, as they did not make sense for the visuals-only condition where no sound was present but the values are still reported for completeness.

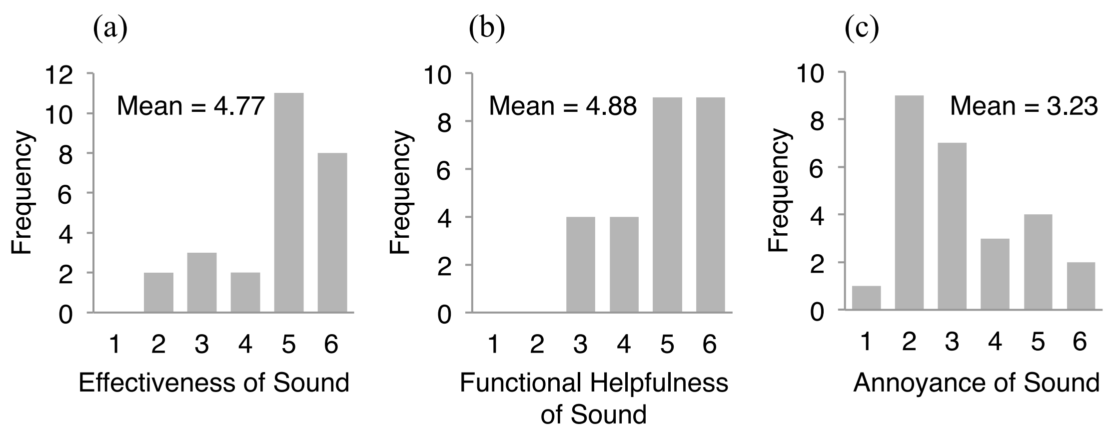


Figure 15. Graphs of the distribution of ratings for the spindex-TTS condition on the sound related questionnaire. The questions were: (a) “How effective was this sound for the task?”; (b) “How functionally helpful was the sound?”; and (c) “How annoying was this sound?”. This was done on a 1-6 Likert-like scale from least to most (n=26).

Experiment Questionnaire

In the final questionnaire, participants' data were found to show a moderate level of distraction from the study (10 participants reported having some sort of distraction). Of those participants who did report being distracted during the study, their write-in explanations included topics of signs/billboards and moving objects in the simulator (5 participants), lead car or other traffic behaviors (2 participants), hunger and personal issues/thoughts (3). It did not seem from the reported causes of distraction that thinking about the songs they were searching for was a source of distraction, and all participants (26) reported recognizing at least one of the song names during the search task. When comparing participants' responses regarding their comfort level using a cell phone or external mp3 player while driving on a six point scale before the study ($M = 3.73$, $SD = 1.12$) and if they were to use a spindex-TTS cue ($M = 4.12$, $SD = 1.07$), it was found via a paired samples t-test that there was no significant difference $t(25) = 1.41$, $p = .170$. Despite this lack of differences 72% of participants preferred the spindex-TTS cue (18) over the visuals only interface (7) (one participant did not respond to this question).

Subjective Performance and Preferences Results Summary

Results of the analyses found that participants rated the spindex-TTS interaction as better in all perceived performance and preference questions than the visuals-only where it was applicable. More people also preferred the spindex-TTS interaction, although it was not rated higher than the participants' current methods of list searching

VACP

Inter-rater reliability for the scores assigned by the raters was calculated for the VACP ratings via three separate two-way random, absolute agreements, average-measures intraclass correlations (ICC) (McGraw & Wong, 1996). These analyses were done to look for agreement across the ratings given for all the sub-tasks that raters made for each task analysis. The results of these analyses revealed that the driving task ratings had an ICC = .554, the visuals-only search task ratings an ICC = .753, and the spindex-TTS search task ratings an ICC = .565. According to Cicchettti (1994) this meant the driving task had fair reliability, the visuals-only search task had excellent reliability, and the search task spindex-TTS had fair reliability. When all of the subtask ratings for each task analysis were analyzed together the value was calculated to be ICC = .831, falling into the excellent reliability range. Although these values were informative to the agreement on the absolute values assigned by the raters, the consistency of values assigned by the raters could also be informative in this exploration. Therefore, three separate two-way random, consistency, average-measures ICCs were also performed (McGraw & Wong, 1996). These correlations revealed that the driving task values assigned by the six raters had an ICC = .819, the visuals-only search task ratings an ICC = .896, and the spindex-TTS search task ratings an ICC = .813, all of which would be considered excellent reliability. All of the ratings were also analyzed as a group, revealing an ICC = .939, also considered to be an excellent reliability. These results mean that the absolute values of the task analyses were somewhat reliable depending on the task analysis, but all of the raters were consistent with their agreements of workload differences in their ratings.

To determine the modality and total condition VACP workload values for the two experimental conditions the task analyses for the driving task and the two search tasks had to be added together. This was done by first determining the values for each task analysis by averaging the six rater's VACP values for each subtask across each modality, creating a rating for each modality for each task with the task analyses. These values were then averaged across the task analysis to give the overall value for that task analysis for each modality. These modality values were then summed with the other modality averages for that task analysis to give the total workload value for each analysis. These means and standard deviations for each modality and the total scores for the three task analyses across the 6 raters can be seen in Table 9. Following these calculations the modality values and total workload for the two experimental conditions was calculated by simply adding each modality score for the driving task and the respective search task. These values can be seen in Figure 16 along with the values for the baseline condition.

Table 9

Descriptive Values of the VACP Ratings for the Modalities and Total Workload for the Three Task-Analyses

	Visual		Auditory		Cognitive		Psychomotor		Total	
Task Analysis	M	SD	M	SD	M	SD	M	SD	M	SD
Average Value										
Driving	2.73	1.32	0	0	4.01	0.96	1.58	0.32	8.32	2.36
Visuals-only	2.74	0.92	1.54	0.63	4.04	1.08	1.39	0.50	9.72	2.63
Spindex-TTS	2.01	0.56	2.09	0.56	4.12	0.94	1.43	0.47	9.64	1.96

VACP Results Summary

Analyses of the VACP data revealed that the raters had fairly high inter-rater reliability, particularly regarding consistency. The descriptive results of the predicted total and modality specific workload showed that the spindex-TTS condition was predicted to have a slightly lower total than the visuals-only condition, with both of those predicted to be higher than driving only. The spindex-TTS condition was also predicted to have a lower level of visual workload than the visuals-only condition, while having an increase of auditory workload. Meanwhile the cognitive and psychomotor modalities were predicted to have only slight differences between the two experimental conditions.

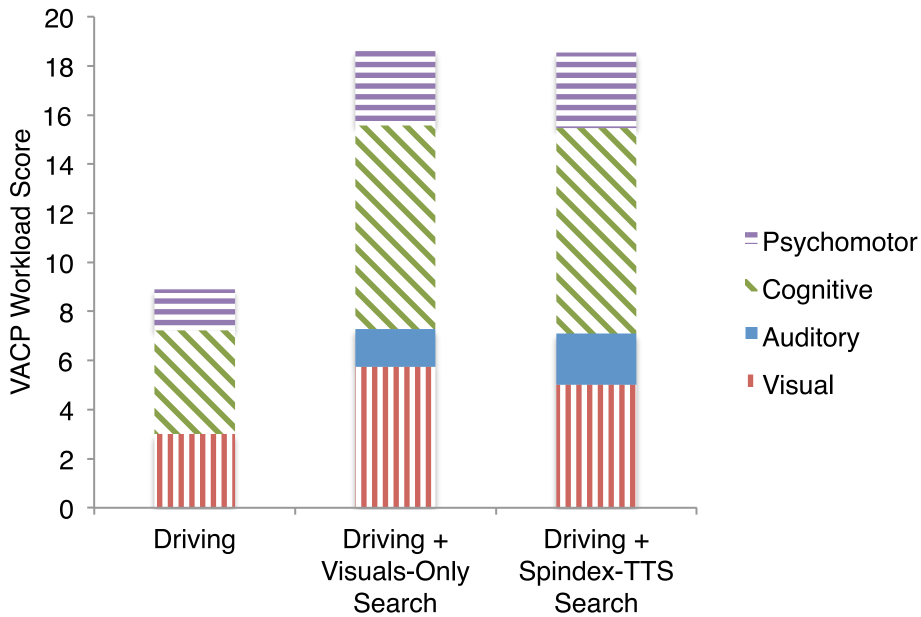


Figure 16. Graph of VACP ratings for each modality for the two experimental conditions and the baseline. The experimental condition values were determined by adding the driving averages across raters to the search task ratings averaged across rater.

CHAPTER 5: DISCUSSION

The present study investigated novices' use of spindex-TTS advanced auditory cues in a list-searching task on a cell phone while driving as compared to a visuals-only interaction. Through measures of list-searching and driving performance, visual behaviors, subjective and objective workload, preferences and perceived performance, and prediction of workload, the study inspected unpracticed participants' workload and willingness to engage when interacting with the two types of devices during two difficulties of driving environments. The research found that applying the spindex-TTS cues to the search task lead to novices being no worse in average time to find a song as compared to the standard, and commonly used visuals-only interaction method. There were slight differences discovered in the number of songs searched and found across the two conditions, with more songs being searched for and found in the visuals-only condition as well as significantly higher accuracy of song selection with that interaction type as compared to the spindex-TTS cues. No learning effects were found to be significant either, most likely due to the lack of enough trials as seen in previous work. No differences across the two conditions were found in driving performance, implying a lack of decrements on the driving task by applying these auditory cues. The visual behaviors results showed that dwell time was greater on the primary task, and glance frequency off of the driving task was lower when using the spindex-TTS cues, meaning participants had their eyes on the road for longer and took their eyes off the road less often when using the spindex-TTS cues. Workload results revealed significantly lower subjective workload but no differences were found in subjective physical and mental workload or any of the objective physiological measures between the two conditions.

Preferences and perceived performance found that participants favored the novel spindex-TTS condition over the normally applied visuals-only interaction, even though they reported it being slightly more annoying. The study also revealed an interesting outcome within the exploration of the use of VACP as a predictive workload measure, showing a similar pattern between the VACP workload ratings and the collected data. Finally, the outcomes of the manipulation of willingness to engage were found to have mixed results, where in some cases it seemed the manipulation of drive difficulty via the drive type might not have created the effect desired. These results and their applications to driving and secondary task research are discussed further in the sections below.

Hypotheses Outcomes

Secondary Task

Hypotheses related to the secondary search task were partially supported by the findings in the current research. H1 was supported in part by participants having no differences in average search time for the songs between the two conditions. However, the significantly higher number of songs searched for, found, and higher accuracy when using the visuals-only interface did not support H1 or previous research (i.e. Gable et al., 2013). These results of fewer songs found and decreased accuracy in the spindex-TTS condition could be due to a number of factors, one of which being the removal of outliers in the current study, which limited the accuracy of the collected visuals-only data to only including values that were at 100 percent accurate. Another factor that could explain the differences in accuracy, searches made, and songs found was the lack of practice with the spindex-TTS cues, which has been shown to increase performance with these cues (Jeon et al., 2012). Although H2 predicted increased performance with increased practice time, this difference was not found to be significant in the current study. Due to the fully mixed design of this study, order effects may not be able to explain the lack of these results, and it could be simply that participants did not have enough trials over the given time to significantly improve performance as seen in previous work. This may be due to the long period of time between searches in the study, which was added to make the study more realistic than constant searching, but may have limited the ability to see the effects of practice. Finally H3 of the manifestation of drive difficulty affecting the willingness to engage being seen through more songs found in the highway drive was not supported, as no differences were found across the drive types. This may be due to lack of differences between the drives, which will be discussed further later in this section.

Driving Performance

H4 was supported in the current study, as driving performance between the two conditions of spindex-TTS and visuals-only did not significantly differ from one another as seen in previous work (Gable et al., 2013; Jeon et al., 2015). These results show support for the idea that adding spindex-TTS cues does not make driving performance any worse than using visuals-only, although it also means it did not improve it.

Visual Behaviors

Results of the study supported H5, seen through the significantly higher dwell time on the primary task, and lower glance frequency away from the task in the spindex-TTS condition, supporting previous research (Gable et al., 2013). This meant that participants relied more on the auditory cues during the conditions with the spindex-TTS cues and applied less visual attention to said task as would be anticipated with ideal use of the system. The expectation of seeing even less time eyes off the road when using the spindex-TTs as practice increased, H6, was not supported by the results, possibly due to lack of trials and acclimation to the system, and further research that investigates the effects of practice with these cues should consider looking into this relationship. The differences in willingness to engage expected to be seen through the visual measures were not found, therefore not supporting H7 or previous research in the area that drove this hypothesis (Kun et al., 2014). This lack of difference, as with others relating to willingness to engage may be due to the lack of variance in difficulty between the two drive types, not forcing participants to modify their willingness to engage and therefore their outward behaviors.

Workload

NASA-TLX results supported H8 in regards to differences in total workload between the two conditions, with participants reporting lower total workload in the spindex-TTS condition than the visuals-only condition, as seen in previous work (Jeon et al., 2015). As referred to earlier in this section, the expected result of higher workload in the city drive being seen via the baseline comparison and again after application of the secondary task was not found in the present study, not supporting H9. This outcome may explain some of the lack of differences seen in regards to hypotheses relating to willingness to engage. This failure to significantly increase workload in the baseline drive, and sustain it through the application of another task, means that willingness to engage may have not been influenced via the environment as was attempted in the current research, although the use of city and highway drive types has been shown to create this desired manipulation in other work (Kun et al., 2014). This lack of evidence for manipulation of the willingness to engage via the use of different drive types should be carefully considered in future work. However, while not a hypothesis, differences in mental workload were found between the drive types within the two-way ANOVA, suggesting some differences between the drive types did exist after addition of the secondary tasks. H10, the expectation to see lower measures of objective workload via lower heart rate and heart rate variability in the spindex-TTS condition was not supported in this research. This may be in part due to the spacing between searches during driving, allowing for heart rates to begin to return to baseline, therefore increasing variability of the heart rate. However, the lack of results may also be attributable to a simple lack of large enough workload differences between the conditions due to the earlier mentioned ceiling effect of driving, therefore leaving too much residual workload available for the secondary task and not increasing workload enough to be detectable via heart rate measures.

Perceived Performance and Preferences

The results of the condition questionnaire supported H11 in that participants rated the spindex-TTS cue higher in all categories related to the condition questionnaire where it was comparable. The annoyance level of the spindex-TTS condition, while not compared statistically, did have an average rating of 3.2 out of 6 towards being annoying, which was not hypothesized but is an important piece of information to note for future work. However, participants also stated that the sound was functionally helpful and effective for the task. H12 was also supported in the results of the research, with more people preferring the spindex-TTS interaction to the visuals-only interaction. The hypothesis of participants being more comfortable using the spindex-TTS cue than using their current method of interaction with their cell phone in the car, H13, was not supported. This may have been due to lack of experience and therefore possibly comfort in using the spindex-TTS cues.

VACP

The hypothesis that similar patterns would be seen between the VACP ratings and the collected subjective workload and visual behavior data, H14, was partially supported. The pattern seen in the slightly lower predicted total workload for the spindex-TTS condition than the visuals-only condition matched with the overall workload differences seen in the collected data, although the predicted differences would be projected to be higher based on the collected results. The pattern of modality specific values seen in the VACP ratings actually matched well to the collected data, with the physical and cognitive modalities showing little difference between the two conditions, as also seen in the RTLX scores. The pattern of predicted scores seen in the visual modality between the two conditions also matched well with the longer dwell time and higher number of glances off the driving task in the visuals-only condition. This result supported previous research showing that actual data collected correlates well with predictions made by experts this study seemed to not entirely support these findings (Mitchell, 2000). While not all of the patterns were as strong as might be expected a number of factors could have affected the results seen here such as the raters low levels of experience with the VACP scale and the process of using it. In addition the actual tasks done in the experiment were not the exact task analyses that were assigned values by the raters. Instead the values were calculated by adding together results from separate task analyses. This shows that part of the usefulness of VACP, if it is found to be useful through more research, may be the flexibility of the measure.

Applications of Results and Future Directions

Though not all of the hypothesized results were found to be significant, the results in the current study add to previous support for the use of spindex-TTS cues for list-searching tasks in the vehicle. Although the number of songs search, number of tasks found, and accuracy may have been slightly decreased with the application of the spindex-TTS cues with these novice users, a difference that has been shown to decrease with more practice in previous work, no other results pointed to negative outcomes of applying the auditory cue. Instead results point toward spindex-TTS cues actually making the search task safer than the visuals-only condition through increased time eyes on the road and decreased glances off of the road, decreased total workload, and preferences showing strong belief that the cues helped them drive more safely. These results all point toward the possibly huge effect that applying these cues could have on the driving population.

Also of particular interest was the theoretical exploration of the factor of willingness to engage. It was somewhat unclear whether participants' willingness to engage was influenced by the manipulation of drive type in this study, but the theoretical implications of the factor are of considerable interest in this and related research. Whereas few values were different from each other between the two drives, if the difficulty of the drives had been made more variable, interesting results between the drive types might have appeared. Further relationships between the spindex-TTS cue and the visuals-only interaction might have been revealed if other factors between drive types had been controlled. Future work could instead focus on modifying factors that are specifically known to increase difficulty of driving tasks such as traffic levels, changes in speed, and other variables instead of simply the type of drive, which would also control for drive type variables not of interest. This area of willingness to engage, not commonly investigated within driving research, should be considered in future work within the driving context due to its possibly large influence on the outcome of future studies or in the future use of interfaces. Whether the research considers willingness to engage as a confound and attempts to control for it throughout the research, or investigates the possible interaction that willingness to engage has on the use of the interface, its consideration in future investigations will help to strengthen the applicable results of future driving research.

The use of the VACP measure in the current research is one of few studies to apply the scale to driving research in recent years. However, the possibility of being a reliable predictive measure, and the ability to investigate the four measures of workload independently of each other show its capabilities and give it strong promise for increased use in the space of driving research. This scale could be used as it is, or converted slightly to make it more directly applicable to the driving task by adding driving or driving related technology specific anchors. The scale could then be applied within the driving interface research community as a first step in development and research of in-vehicle interfaces. This could possibly allow companies and researchers to save money on participants until interfaces and designs are further along in development, allowing for more data to be collected on more focused and meaningful questions. This scale should be strongly considered in future work as a simple predictive measure of workload along with other data, or as a possible replacement for collection of large amounts of human subjects data if applied properly with the understanding that it needs to be investigated further.

This study found a number of interesting findings, but there were also a few questions that the research left unanswered, whether it was due to lack of measuring these factors, or limitations in the study design that did not allow for the data to be revealed. One of the larger limitations of the data was that of experience given to the participants with the novel spindex-TTS interface. As most people are used to using the touch screen, visuals-only, kinetic flicking interaction method on our phones and other display, very few people have ever even heard of advanced auditory cues, let alone used them. A lack of trials also seemed to prevent investigating learning effects in previous work (Gable et al., 2013) and although this study attempted to increase the number of trials with the interface, it seems that it was not enough to see the participants get used to the new interface. To rectify this issue future work should either focus on investigating the variable of practice with the cues and attempt to ensure the participants are at a high level of ability and comfort with the spindex-TTS cue before starting the research or try to push more trials into the research blocks via shortening of the time between search tasks or longer testing blocks.

One of the areas lacking in any results across all drive types were the physiological measures of workload. Both heart rate and heart rate variability were not found to be significant in any analyses. As can be seen in Table 7 it seems the measures changed very little across participants and had relatively high standard deviations. While previous studies have found differences in physiological measures of workload it can be difficult to measure and one must consider many factors when using such measures (Mehler et al., 2012). As stated previously one of these factors may have been that the time between search tasks was too large and allowed participants' heart rates to get back to baseline, and future work could decrease these time periods. Another factor that could be considered is that heart rate has been known in some instances to show very little change during the intake of visual information due to activation of both the parasympathetic and sympathetic systems, keeping the heart rate from increasing in these situations and therefore showing no differences in workload (Mehler et al., 2012). Mehler et al. (2012) recommend using SCL as a secondary measure for these instances and future studies should consider the integration of this measure to investigate objective workload further.

Another option for investigating objective workload in future work would be the use of a different physiological measure. Pupil size has been used both in driving (Palinko, Kun, Shyrokov, & Heeman, 2010; Recarte, & Nunes, 2003), and other areas of research (Granholm, Asarnow, Sarkin, & Dykes, 1996; Iqbal, Zheng, & Bailey, 2004). Most of these studies have found significant increases in pupil size, or pupillary response, as secondary task difficulty increases (Palinko et al., 2010), or as the task difficulty itself increases (Granholm et al., 1996; Iqbal et al., 2004) although a ceiling effect when an individual reaches the limit of his or her available cognitive resources has been found (Iqbal et al., 2004). One major concern when using the measure is that a pupillary light reflex occurs when a participant is exposed to different levels of luminance, however Kun, Palinko, and Razumenić (2012) reviewed this issue and introduced a weighting function that can be used if the luminance cannot be controlled within the experiment such as when participants would be looking between two interfaces. It should be noted that work seems to find the pupillary response is very fast, increasing and decreasing the pupil size very quickly, so future researchers many want to detect pupil size levels only during the search task periods instead of investigating an aggregate over the whole driving duration as was done with heart rate in this study.

Some of the most convincing results in the current research were in relation to the visual behavior data. Similarly, the same types of data were some of the strongest results in the study prior to this (Gable et al., 2013). Other studies in the space, particularly when investigating willingness to engage have also focused heavily on eye tracking metrics (e.g., Kun et al., 2014; Victor et al., 2005). This sensitivity of visual behavior measures, and the meaning that can be gleaned from them is clearly seen through the common use of these measures. Further use of these types of measures such as investigation of gaze concentration measures as detectors of workload may be of use in future research in the area and provide more information to researchers (SAE, 2000).

Increased dwell times on the road and decreased glance frequency off of the road are promising results and infer that the driver therefore had a better idea of what was going on within the driving task, but these results are not direct evidence of this assumption. To determine if this increased eyes on the road time actually makes the driver safer an important aspect of safe driving that was not discussed in this report must be measured, that of situation awareness (SA). As defined by Endsley (1995) SA is, “the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (p. 36). This definition implies a large effect of attention to the factors of SA but also implies that by measuring SA, researchers could glean more knowledge than through simple visual behavior measures. Therefore, it is recommended that future research investigate the SA of drivers while they use these types of cues to determine the cues' influence on participants' SA. Many options for measuring SA exist such as the Situation Present Assessment Method (SPAM) (Durso & Dattel, 2004) and the Situation Awareness Global Assessment Technique (SAGAT) (Endsley, 1995). However, as Schuett and Walker (2013) point out, these measures, while well accepted, are limited in their usefulness within investigations of auditory displays as they interfere with the verbal processing that would be necessary to complete the ongoing auditory or multimodal interaction. Instead a measure of hazard perception may be more applicable in this type of research. Multiple methods for measuring SA through hazard perception exist, either within a driving simulator or via a video-based technique (Gugerty, 2011; Horswill & McKenna, 2004). Performing such a task in a simulator would also allow for an opportunity to collect driving performance reaction time variables such as time to break and others discussed earlier, which would provide more even more interesting and applicable data. Whichever method is chosen, it should be considered carefully with what method will be used to have participants complete the search tasks as ensuring that the interactions occur at the same time would be important.

Conclusion

In all, this study found that the application of a spindex-TTS cue to a list-searching task while driving is almost entirely better or comparable to a visuals-only interaction method in regards to distraction via measures of workload and wiliness to engage. Outside of novice users' slightly lower search task performance, which has been shown to improve with practice, there were no drawbacks of applying the spindex-TTS method other than some slight annoyance by the sound. Measures of driving performance between the conditions, average time to a find a song, subjective physical and mental workload, and objective workload via heart rate measures all pointed to the spindex-TTS interaction being no worse than the visuals-only interaction, meaning no negative effects of applying the auditory cues. Instead, the application of the spindex-TTS cue actually showed significant improvements as compared to a visuals-only condition through the drivers' visual workload via more dwell time on the secondary task and lower glance frequency off the road, lower total subjective workload, and perceived performance and preferences. The study also found that a process of predicting workload assigned higher numerical workload values to the visuals-only condition, particularly in regards to the visual load, suggesting a possible overload of that modality. Due to the similar patterns between the predicted workload and actual visual behavior data collected during the experiment this is particularly interesting and suggests that the spindex-TTS cue would be a safer interface to use during a driving task than the visuals-only interaction method as it does not overload the visual modality nearly as much. Future work in the area is still necessary as discussed in sections above, however results continue to support the use of these auditory cues within the driving and list-searching context.

Aside from finding support for the use of spindex-TTS cues to improve drivers' wellbeing via allowing them to do a list-searching task in the car more safely, it is not so much the integration of these cues into the vehicle that makes this study so important, but the process that was done to find these results. This process of studying the different possible interfaces through the many methods available to researchers is something that all of those involved with in-vehicle interface design research should contemplate. So, while the findings regarding the auditory cues should be considered in future development of in-vehicle interfaces, it is more the process that was undertaken here that in-vehicle interface researchers should reflect upon in an effort to improve the safety of our roads.

APPENDIX A: CONDITION PREFERENCE QUESTIONNNAIRE

Participant #:
Condition #:

Please select your level of performance for each of the following:

[Perceived performance on the search task]

1. How effective were you at the search task?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
	Not at all Effective					Very Effective
2. How effective was this sound for the task?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
	Not at all Effective					Very Effective
3. How functionally helpful was this sound?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
	Not at all Helpful					Very Helpful
4. How annoying was this sound	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
	Not at all Annoying					Very Annoying

[Perceived performance on the driving task]

5. How effective were you at the driving task?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
	Not at all Effective					Very Effective
6. How safe of a driver were you during the driving task?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
	Not at all Safe					Very Safe

APPENDIX B: DEMOGRAPHICS AND PREFERENCE QUESTIONNAIRE

Participant #:

1. Gender: ☐Male ☐Female

2. Age:

3. Years that you have been driving:

4. Years that you have been playing video games:

5. Number of moving vehicle tickets or crashes in the past three years:

6. What was your comfort level for using a touch screen before the study?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
	Not at all Comfortable			Extremely Comfortable		
7. What was your comfort level for using a touch screen in a car before the study?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
	Not at all Comfortable			Extremely Comfortable		
8. What is your comfort level for using a cell phone/external mp3 player while driving?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
	Not at all Comfortable			Extremely Comfortable		
9. What would your comfort level be using a phone/external mp3 player while driving if you used Spindex?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
	Not at all Comfortable			Extremely Comfortable		

10. What kind of auditory cue type do you like best?

A. No sound: "……", "……",…
B. Attenuated Speech Index + TTS: "A Ada Gilmore", "a Allegra Seidner", "B Bruce Walker", "b Burl Rose",…

11. Did you recognize any of the song names? ☐ Yes ☐ No

12. Did you feel distracted during any of the study by anything other than the cell phone search task? ☐ Yes ☐ No

12B. If yes please explain:

Comments:

APPENDIX C: EXPERIMENT ADVERTISEMENT

Intended for posting on SONA

Georgia Tech Recruitment Text
Comparison of Auditory Representations for Menu Navigation While Driving

This study is comparing various auditory cues to improve the performance of menu navigation while driving. Participants will perform a simple driving task on a driving simulator while attempting to find song names on a touch screen cell phone and receiving different types of auditory cues. This study will take no more than around 1 hour and participants will receive 1 credit for their participation. Participants must have normal or corrected to normal vision and hearing, and have a valid drivers license with at least 2 years of driving experience. If vision correction is needed it MUST be in the form on contacts, corrective glasses are NOT allowed. Participants are also asked not perform any vigorous exercise or consume caffeine, nicotine, or any other stimulant for at least 2 hours prior to the study. If participants have any form of ADD we also ask they do not sign up for the study as we are looking at attention in the study and such issues could confound the data. Additionally, the participant must be open to wearing an eye tracking device as well as heart rate monitor connections on their persons for the study For this reason we ask that participants come wearing clothing that allows for ease of access to their torso for their own placement of the necessary connections.

If there are any questions please contact thomas.gable@gatech.edu

APPENDIX D: STUDY CONSENT FORM

Study: Comparison of Auditory Representations for Menu navigation while Driving
Principal Investigator: Dr. Bruce N. Walker (404-894-8265)
Location: School of Psychology, Coon (Psychology) Building, Georgia Institute of Technology
Duration of Each Session: 1.0 hour **Number of Sessions:** 1
Total Compensation: 1.0 credit hour
Approximate Number of Participants: 200
Participation limitations: Normal or corrected to normal vision and hearing and no mobility impairments. Participants must also have had a valid drivers license for 2 years and be wearing contacts on the day of the study if vision correction is necessary. Participants are also asked to have not performed any vigorous exercise or consumed caffeine, nicotine, or any other stimulant for at least 2 hours prior to the study and anyone with ADD is also asked to not take part in the study due to its focus on attention.

General: You are being asked volunteer for a psychological experiment research project. There are many ways to use sound to represent traditionally visual menu items, and we want to compare some of them with a various dual task situations. We would like to make these comparisons in a realistic scenario, such as using your cell phone or touch devices while driving. You will be asked to navigate through a long list using a unique combination of visuals and sound while performing the primary driving task. Your participation will help develop new audio cues for future real-world applications. It will also provide you with some experience in the conduct of research in psychology.

Study Purposes: This research is looking at how auditory cues can improve navigation through an electronic menu.

Procedures: You will navigate the menus to find the target items on the list by using a touch screen cell phone while you are performing a simple driving task. You will be asked to find the requested target name on the menu as quickly as possible while you keep driving with the given portion of your attention resource by an experimenter. After each condition, you will be asked to answer some questions. Finally, after completion of all conditions, you will be asked to fill out the demographic questionnaire and write your comments. During the study you will be asked to wear an eye tracking system and be connected to a heart rate monitor to measure your eye movements and heart rate while you perform the tasks.

Foreseeable Risks or Discomforts: This study is expected to involve no more than minimal risks associated with listening to sounds and driving a simulator. In some instances individuals do have a feeling of motion sickness when using the simulator. We will ask you to perform a short drive to test if you show any signs of sickness. In the event that you do experience any sickness at that time or at any point during the study we will ask you to sit in a stationary chair until the feeling subsides. At that point you will be debriefed and released from the study but will still receive your full credit for participating in the study.

Confidentiality: The following procedures will be followed to keep your personal information confidential in this study: The data that is collected about you will be kept private to the extent allowed by law. To protect your privacy, your records will be kept under a code number rather than by name. Your records will be kept in locked files and only study staff will be allowed to look at them. Your name and any other fact that might point to you will not appear when results of this study are presented or published. To make sure that this research is being carried out in the proper way, the Georgia Institute of Technology IRB will review study records. The Office of Human Research Protections may also look over study records during required reviews. Again, your privacy will be protected to the extent allowed by law.

Alternative Credit Option: Alternatives to participating in this study are provided by your course instructor. They include, but are not limited to, reading journal articles and writing a brief report based on the articles.

Injury/Adverse Reaction: Reports of injury or reaction should be made to Dr. Bruce Walker (404-894-8265). Neither the Georgia Institute of Technology nor the principal investigator has made provision for payment of costs associated with any injury resulting from participation in this study.

Contact Persons: If you have questions about this research, call or write Dr. Bruce Walker at 404-894-8265; School of Psychology, GA Tech, 654 Cherry Street, Atlanta, GA 30332-0170.

Statement of Rights: You have rights as a research volunteer. Taking part in this study is completely voluntary. If you do not take part, you will have no penalty. You may stop taking part in this study at any time with no penalty. If you have any questions about your rights as a research volunteer, call or write: The Institutional Review Board, Office of Research Compliance, 505 Tenth Street, Campus 0420. Phone: 404-894-6942; Fax: 404-385-2081.

Benefits: This study would contribute to people with normal vision as well as visually-impaired people in their use of electronic menus.

Signatures: A copy of this form will be given to you. If you sign below, it means that you have read the information given in this consent form, and you would like to be a volunteer in this study.

Participant's Signature:

____

Date: _____________________

Person Obtaining Consent:

____

Date: _____________________

APPENDIX E: STUDY INSTRUCTIONS

Thanks and Introduction
First of all, thank you for your participation in this study. We are members of Sonification Lab in school of psychology.

Overall Project
This experiment is a part of the auditory menus project, which is intended to enhance the use of electronic menus with auditory cues.

Purpose of Experiment
This research is comparing various auditory cues to improve the performance of menu navigation (secondary task) and a simple driving task (primary task) while users' attention focuses on the primary task.

Procedure
In this experiment, you will be asked to navigate through a list on a cell phone to find the requested target-items (song names) while you are driving a simulator.

<u>Driving Simulation</u>

The goal of the driving task is to drive down the road at the posted speed while maintaining good lane keeping and following the car in front of you at the distance seen at the beginning of the drive. You will be driving with one hand during the whole experiment. During the course of each drive, the lead car will change speeds and you should attempt to continually drive safely at the designated distance behind the vehicle.

You can control the simulated car as you would with any automatic transmission vehicle. We will give you few minutes to familiarize yourself with the simulator before we start the conditions as well to see if you have any adverse reactions. In the case that you do have adverse reactions to the simulation we will stop the research and then debrief you and release you from the study.

<u>Menu Navigation</u>

While you are driving, you will be searching for a song name on a cell phone. You will be able to hold the cell phone in whatever hand you choose and rest it on an armrest, but for the rest of the experiment you must use that same hand and drive with the other hand. Once the experiment begins you will hear a target song name randomly generated from the cell phone*. Upon hearing this, you will start to navigate the list on the phone by flicking through it. The list is alphabetized by the first letter of the song. You will be asked to move down the list as quickly as possible until you find the target song. Once you find the target, clicking on the song will signal that you have reached your final destination. Make sure that the target name disappears on the screen. Unlike the functionality of a typical device, the list will not wrap around after reaching the top or end of the list. You do not have to listen to the complete name before moving to the next item, instead, we would like you to move as quickly as possible through the list but without sacrificing accuracy. If you are sure an item is not your target item, feel free to move over that item as you navigate.

After the selection of the target, you can continue the driving task alone again. After a few seconds, you will hear a new target name, and the remaining procedure is the same as the previous case.
We will give you few minutes to familiarize yourself with this task before we start the conditions.
*Note: when each condition on the cell phone is being loaded you will hear multiple songs announced by the phone, we ask that you ignore these sounds and upon the start of each condition you select the first song to truly begin the measurement.

In this experiment, driving task is the primary task and the search task is secondary. You should attempt to drive safely first and then give attention to the search task. If you need to consider a numerical split of attention, you should focus on it as 80% of your resources on the driving and allocate 20% to your list navigation task.

For each condition you will be asked to drive for a predetermined period and during this time you should continually perform the two tasks to the best of your ability. You will perform one condition with no search task, and two conditions with a search task. At the end of each condition you will be asked to answer some questions for that condition and perform a workload questionnaire. Before each condition, the experimenter will set up for that condition, so DO NOT press any buttons on either screen until the condition begins. After each condition, you will be asked a few questions about how you handled the song search task. You may then be given a few minutes to rest before beginning the next condition.
Finally, after completion of three conditions, you will be asked to fill out the demographic questionnaire and write your comments, any suggestions or notions for this research.

You will also be asked to wear a pair of mobile eye-trackers and a heart rate monitor during this study. We will fit you with them at the beginning of the study and they will be used during each driving and searching condition (They will not be recording when you are filling out preference and workload questionnaires). We ask that you view the driving simulator and secondary task as you normally would. Since heart rate will be measured, we also have previously asked you if you have exercised or consumed caffeine or any stimulants in the least two hours before the study. If you have done either, please tell the experimenter now.

If you have any questions about the task you are being asked to perform, please ask the experimenter now.
If you change your mind to participate in this study, you can stop at anytime.
Now, let’s get started.

APPENDIX F: HEART RATE MONITOR PLACEMENT INSTRUCTIONS

Instructions for Applying Heart Rate Monitoring Leads

1) Using the included diagram, identify the 3 locations where the electrode pads will be applied. Two will be located just under the collarbones, preferably in the gap between the shoulder muscles. The other will be near the stomach over the bottom rib bone.

2) Gently wipe the 3 areas with a cotton swab to clear any dead skin.

3) Using an alcohol wipe, clean the 3 areas thoroughly and then let them dry to allow the electrode pads to stick cleanly.

4) Once the wiped areas are dry, place one pad in the center of each cleaned area. All pads are the same, and it does not matter which pads goes on which of the 3 areas.

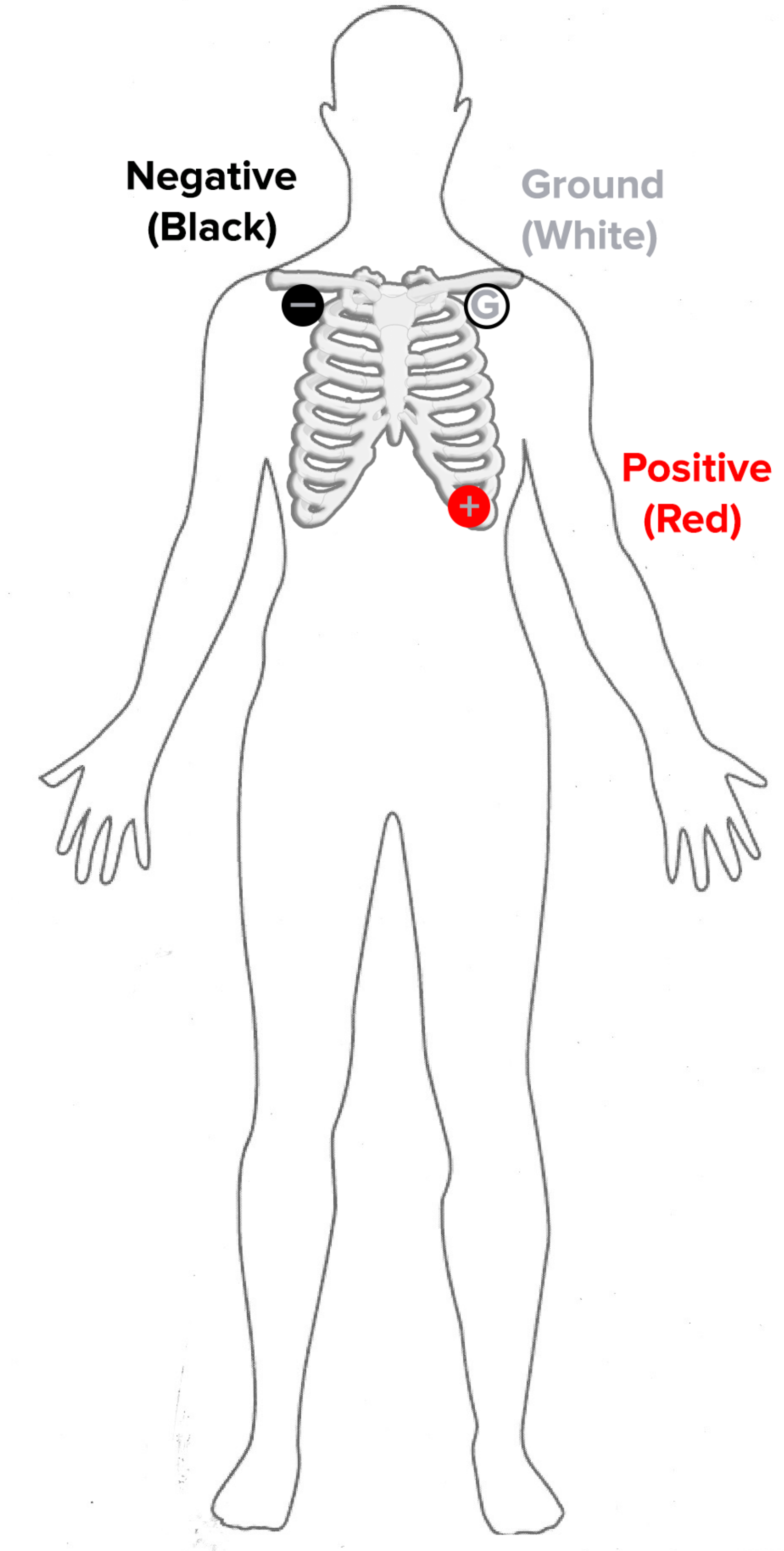
5) Grab the three wire leads for the heart rate system. They are labeled with twist ties as:

+	-	G
(Positive)	(Negative)	(Ground)

6) Snap each lead onto the proper electrode pad according to the included diagram. The leads should snap in with only a small amount of force.

7) Verify the location of the three electrodes. They should be free from all fabric, belts, and clothing and should not fall or peel off as you move. The leads should go under your clothing and out of the area by your belt.

8) Notify the experimenter that you are ready to continue.


Negative
(Black)
Ground
(White)
Positive
(Red)
−
G
+

APPENDIX G: DEBRIEF FORM

Thanks and Introduction
First of all, thank you for your participation in this experiment. We are members of Sonification Lab in school of psychology.

Overall Project
This experiment is a part of the in vehicle dual task project, which is intended to look at how different types of interfaces can improve navigation on in-vehicle technologies and ways to measure these differences.

Purpose of Experiment
The purpose of this experiment is to investigate cognitive load, visual attention, and driving performance when performing a secondary task (song searching). Auditory cues are believed to improve the performance of menu navigation (secondary task) and driving (primary task) while reducing the cognitive effort of users and increase visual attention towards the road.

Revelation of Experiment Condition
You performed 3 conditions of this experiment, one of which was a control where you did not perform a search task, one that incorporated the song search task, and a final condition that added auditory feedback called Spindex + TTS (text-to-speech).

Experimental Conditions:

1) Driving only
2) Driving while performing song search
3) Driving while performing song search – with auditory Spindex+TTS feedback

Meaning of Expected Results
Analysis of your driving performance will help reveal if a secondary task has any effect on driving. From the NASA TLX, which you did in every condition, we can infer how much overall task workload could be ameliorated by using the auditory cues. The eye-trackers will allow for us to see how the two tasks affected your visual attention and cognitive load. The heart rate monitor also allows us to assess your cognitive load. All of these may allow us to determine if the auditory cues made a difference in your visual attention and performance. If driving performance, song search performance, and/or cognitive load change between conditions, we can infer that the addition of auditory feedback may help reduce drivers' cognitive load.

Confidentiality and Anonymity
The results of your experiment will be used for only psychological study and never used for any other purposes. The data that is collected from you will be kept private to the extent allowed by law. To protect your privacy, your records will be kept under a code

number rather than by name. Your records will be kept in locked files and only research staffs will be allowed to look at them. Your name and any other fact that might point to you will not appear when results of this study are presented or published. To make sure that this research is being carried out in the proper way, the Georgia Institute of Technology IRB will review study records. Again, your privacy will be protected to the extent allowed by law.

Conclusion
All of the experiment procedures are finished. We very much appreciate your efforts again.

Contact Information
For further information of this research, contact:
Principal Investigator
Dr. Bruce Walker (bruce.walker@psych.gatech.edu)
Experimenters
Thomas Gable (thomas.gable@gatech.edu), Riley Winton (rjwinton@gatech.edu), & Brittany Corbett (bcorbett8@gatech.edu)

APPENDIX H: TASK ANALYSIS – DRIVING TASK

Driving Task

0.0 Maneuver simulator so car is in your lane and at specific distance behind lead vehicle

1.0 Keep instructed distance between you and the lead car

- 1.1 Maintain current gas pedal position with foot
- 1.2 Look at visual cues (lead car, road, etc.) presented by driving simulator monitors to estimate distance between you and the lead vehicle
- 1.3 Compare estimated following distance to mental representation of instructed distance
- 1.4 If following distance is less than or greater than instructed, adjust pressure on gas or brake pedal to correct following distance
- 1.5 Go back to 1.1

2.0 Keep ideal lane position

- 2.1 Maintain current steering wheel position with hand
- 2.2 Look at visual cues (lines on road, etc.) presented by driving simulator monitors to assess current lane position
- 2.3 Compare estimate of current lane position to mental representation of instructed lane position
- 2.4 If lane position is left or right of instructed lane position, adjust steering wheel to place car in instructed lane position
- 2.5 Go back to 2.1

APPENDIX I: TASK ANALYSIS – SEARCH TASK VISUALS-ONLY

Search Task - Visuals-only

0.0 Find goal song on cell phone

1.0 Determine goal song

- 1.1 Detect TTS from phone stating what song to find

2.0 Search list for song

- 2.1 Make ballistic movements in direction of correct song
  - 2.1.1 Make large flicks on screen with thumb at needed velocity to move through list in necessary direction towards the goal song
  - 2.1.2 Observe screen to determine current location in list
  - 2.1.3 Compare current location in list to goal song in memory
  - 2.1.4 If near song move to 2.2, if not return to 2.1.1
- 2.2 Make fine tuning movements to the correct song
  - 2.2.1 Make small flicks on screen with thumb at needed velocity to move through list in necessary direction towards the goal song
  - 2.2.2 Observe screen to determine current location in list
  - 2.2.3 Compare current location in list to goal song in memory
  - 2.2.4 If highlighted song is the correct one move to 3.0, if not return to 2.2.1

3.0 Select correct song

- 3.1 Use thumb to press on the song selection, viewing screen to ensure thumb presses on correct song

APPENDIX J: TASK ANALSYSIS – SEARCH TASK SPINDEX+TTS

Search Task - Spindex+TTS

0.0 Find goal song on cell phone

1.0 Determine goal song

1.1 Detect TTS from phone stating what song to find

2.0 Search list for song

2.1 Make ballistic movements in direction of correct song

2.1.1 Make large flicks on screen with thumb at needed velocity to move through list in necessary direction towards the goal song

2.1.2 Listen to auditory cues to determine current location in list

2.1.3 Compare current location in list to goal song in memory

2.1.4 If near song move to 2.2, if not return to 2.1.1

2.2 Make fine tuning movements to the correct song

2.2.1 Make small flicks on screen with thumb at needed velocity to move through list in necessary direction towards the goal song

2.2.2 Listen to auditory cues to determine current location in list

2.2.3 Compare current location in list to goal song in memory

2.2.4 If highlighted song is the correct one move to 3.0, if not return to 2.2.1

3.0 Select correct song

3.1 Use thumb to press on the song selection, viewing screen to ensure thumb presses on correct song*

APPENDIX K: VACP WORKLOAD SCALE

Scale Value	Scale Descriptor
Visual	
0.0	No Visual Activity
1.0	Visually Register/Detect (detect occurrence of image)
3.7	Visually Discriminate (detect visual differences)
4.0	Visually Inspect/Check (discrete inspection/static condition)
5.0	Visually Locate/Align (selective orientation)
5.4	Visually Track/Follow (maintain orientation)
5.9	Visually Read (symbol)
7.0	Visually Scan/Search/Monitor (continuous/serial inspection, multiple conditions)
Auditory	
0.0	No Auditory Activity
1.0	Detect/Register Sound (detect occurrence of sound)
2.0	Orient to Sound (general orientation/attention)
4.2	Orient to Sound (selective orientation/attention)
4.3	Verify Auditory Feedback (detect occurrence of anticipated sound
4.9	Interpret Semantic Content (speech)
6.6	Discriminate Sound Characteristics (detect auditory differences)
7.0	Interpret Sound Patterns (pulse rates, etc.)
Cognitive	
0.0	No Cognitive Activity
1.0	Automatic (simple association)
1.2	Alternative Selection
3.7	Sign/Signal Recognition
4.6	Evaluation/Judgment (consider single aspect)
5.3	Encoding/Decoding, Recall
6.8	Evaluation/Judgment (consider several aspects)
7.0	Estimation, Calculation, Conversion
Psychomotor	
0.0	No Psychomotor Activity
1.0	Speech
2.2	Discrete Actuation (button, toggle, trigger)
2.6	Continuous Adjustive (flight control, sensor control)
4.6	Manipulative
5.8	Discrete Adjustive (rotary, vertical thumbwheel, lever position)
6.5	Symbolic Production (writing)
7.0	Serial Discrete Manipulation (keyboard entries)

REFERENCES

Aasman, J., Mulder, G., & Mulder, L.J.M. (1987). Operator effort and the measurement of heart- rate variability. Human Factors, 29(2), 161-170.

Angell, L. S., Auflick, J., Austria, P. A., Kochhar, D. S., Tijerina, L., Biever, W., ... & Kiger, S. (2006). Driver workload metrics task 2 final report (No. HS-810 635).

Alm, H., & Nilsson, L. (1995). The effects of a mobile telephone task on driver behaviour in a car following situation. Accident Analysis & Prevention, 27, 707-715.

Brookhuis, K. A., & de Waard, D. (2001). 2.5 Assessment of Drivers' Workload: Performance and Subjective and Physiological Indexes. Stress, workload, and fatigue, 321.

Brookhuis, K. A., & de Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. Accident Analysis & Prevention, 42, 898-903.

Brumby, D. P., Davies, S. C., Janssen, C. P., & Grace, J. J. (2011, May). Fast or safe?: how performance objectives determine modality output choices while interacting on the move. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 473-482). ACM.

Cain, B. (2007). A review of the mental workload literature (Report Number RTO-TR-HFM-121-Part-II). Toronto, Canada: Defence Research and Development, Human System Integration Section

Chisholm, S.L., Caird, J.K., & Lockhart, J. (2008). The effects of practice with mp3 players on driving performance. Accident Analysis and Prevention, 40, 704-713.

Drews, F. A., Yazdani, H., Godfrey, C. N., Cooper, J. M., & Strayer, D. L. (2009). Text messaging during simulated driving. Human Factors: The Journal of the Human Factors and Ergonomics Society, 51, 762-770.

Durso, F. T., & Dattel, A. R. (2004). SPAM: The real-time assessment of SA. In Banbury, S., & Tremblay, S. (Eds.). (2004). A cognitive approach to situation awareness: theory and application. Gower Publishing, Ltd.137-154.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. Human Factors: The Journal of the Human Factors and Ergonomics Society, 37, 32-64.

Engström, J., Johansson, E., & Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. Transportation Research Part F: Traffic Psychology and Behaviour, 8, 97-120.

Fu, W. T., & Gray, W. D. (2006). Suboptimal tradeoffs in information seeking. Cognitive Psychology, 52, 195-242.

Gable, T.M. & Walker, B.N. (2013). Georgia Tech Simulator Sickness Screening Protocol. Georgia Tech School of Psychology Tech Report GT-PSYC-TR-2013-01.

Gable, T. M., Walker, B. N., Moses, H. R., & Chitloor, R. D. (2013, October). Advanced auditory cues on mobile phones help keep drivers' eyes on the road. In Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 66-73). ACM.

Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. Psychophysiology, 33, 457-461.

Gugerty, L. (2011). Situation awareness in driving. Handbook for driving simulation in engineering, medicine and psychology.

Hart, S. G. (2006, October). NASA-task load index (NASA-TLX); 20 years later. In Proceedings of the human factors and ergonomics society annual meeting (pp. 904-908). Sage Publications.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Human mental workload, 1, 139-183.

Harvey, A. R., & Carden, R. L. (2009). Driving error and anxiety related to iPod mp3 player use in a simulated driving experience. Perceptual and motor skills, 109, 159-167.

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

Horswill, M.S., McKenna, F.P., 2004. Driver's hazard perception ability: situation awareness on the road. In: Banbury, S., Tremblay, S. (Eds.), A Cognitive Approach to Situation Awareness. Ashgate, Aldershot, UK.

Iqbal, S. T., Zheng, X. S., & Bailey, B. P. (2004, April). Task-evoked pupillary response to mental workload in human-computer interaction. In CHI'04 extended abstracts on Human factors in computing systems (pp. 1477-1480). ACM.

Jeon, M., Davison, B.K., Nees, M.A., Wilson, J., & Walker, B.N. (2009, September). Enhance auditory menu cues Improve dual task perforce and are preferred with in-vehicle technologies. In Proceeding of the First International Conference on Automotive User Interfaces and Interactive Vehicular Applications, (pp. 91-98), ACM.

Jeon, M., Gable, T. M., Davison, B. K., Nees, M. A., Wilson, J., & Walker, B. N. (2015). Menu Navigation With In-Vehicle Technologies: Auditory Menu Cues Improve Dual Task Performance, Preference, and Workload. International Journal of Human-Computer Interaction, 31, 1-16.

Jeon, M., & Walker, B. N. (2011). Spindex (speech index) improves auditory menu acceptance and navigation performance. ACM Transactions on Accessible Computing (TACCESS), 3, 10.

Jeon, M., Walker, B. N., & Srivastava, A. (2012). "Spindex"(Speech Index) Enhances Menus on Touch Screen Devices with Tapping, Wheeling, and Flicking. ACM Transactions on Computer-Human Interaction (TOCHI), 19,14.

Keller, J. (2002, December). Human performance modeling for discrete-event simulation: workload. In Simulation Conference, 2002. Proceedings of the Winter (pp. 157-162). IEEE.

Kun, A. L., Brumby, D. P., & Medenica, Z. (2014, September). The Musical Road: Interacting with a Portable Music Player in the City and on the Highway. In Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 1-7). ACM.

Kun, A. L., Palinko, O., & Razumenić, I. (2012, October). Exploring the effects of size and luminance of visual targets on the pupillary light reflex. In Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 183-186). ACM.

Lamble, D., Kauranen, T., Laakso, M., & Summala, H. (1999). Cognitive load and detection thresholds in car following situations: safety implications for using mobile (cellular) telephones while driving. Accident Analysis & Prevention, 31, 617-623.

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

Lasch, A. & Kujala, T. (2012, October). Designing browsing for in-car music player: effects of touch screen scrolling techniques, items per page and screen orientation on driver distraction. In Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 41-48). ACM.

Lee, J. D., Roberts, S. C., Hoffman, J. D., & Angell, L. S. (2012). Scrolling and Driving How an MP3 Player and Its Aftermarket Controller Affect Driving Performance and Visual Behavior. Human Factors: The Journal of the Human Factors and Ergonomics Society, 54, 250-263.

Lee, J. D., Young, K. L., & Regan, M. A. (2008). Defining driver distraction. In M. A. Regan, J. D. Lee, & K. Young (Ed.), Driver distraction: Theory, effects and mitigation, 31-40. Boca Raton, FL: CRC.

Lewis-Evans, B., De Waard, D., & Brookhuis, K. A. (2011). Speed maintenance under cognitive load–Implications for theories of driver behaviour. Accident Analysis & Prevention, 43, 1497-1507.

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

Mattes, S. (2003). The lane-change-task as a tool for driver distraction evaluation. Quality of Work and Products in Enterprises of the Future, 57-60.

Mehler, B., Reimer, B., & Coughlin, J. F. (2012). Sensitivity of Physiological Measures for Detecting Systematic Variations in Cognitive Demand From a Working Memory Task An On-Road Study Across Three Age Groups. Human Factors: The Journal of the Human Factors and Ergonomics Society, 54, 396-412.

Mehler, B., Reimer, B., Coughlin, J. F., & Dusek, J. A. (2009). Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. Transportation Research Record: Journal of the Transportation Research Board, 2138, 6-12.

Miller, S. (2001). Workload measures. National Advanced Driving Simulator. Iowa City, United States.

McCrasken, J. H., & Aldrich, T. B. (1984). Analysis of selected LHX mission functions: Implications for operator workload and system automation goals (Technical Note ASI479-024-84). Fort Rucker, AL: Army Research Institute Aviation Research and Development Activity.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. Psychological methods, 1, 30.

Mitchell, D. K. (2000). Mental workload and ARL workload modeling tools (No. ARL-TN-161). Aberdeen Proving Ground, MD, US Army Research Laboratory, Human Research & Engineering Directorate.

Nees, M. A., & Walker, B. N. (2011). Auditory displays for in-vehicle technologies. Reviews of Human Factors and Ergonomics, 7, 58-99.

National Highway Traffic Safety Administration. (2013). Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices (Docket No. NHTSA-2010-0053). Washington, DC: US Department of Transportation National Highway Traffic Safety Administration (NHTSA).

Palinko, O., Kun, A. L., Shyrokov, A., & Heeman, P. (2010, March). Estimating cognitive load using remote eye tracking in a driving simulator. In Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications (pp. 141-144). ACM.

Paxion, J., Galy, E., & Berthelon, C. (2014). Mental workload and driving. Frontiers in psychology, 5.

Ranney, T. A., Baldwin, G. H. S., Smith, L. A., Mazzae, E. N., & Pierce, R. S. (2014, November). Detection response task evaluation for driver distraction measurement application. (Report No. DOT HS 812 077). Washington, DC: National Highway Traffic Safety Administration.

Ranney, T. A., Mazzae, E., Garrott, R., & Goodman, M. J. (2000, July). NHTSA driver distraction research: Past, present, and future. In Driver Distraction Internet Forum.

Recarte, M. A., & Nunes, L. M. (2003). Mental workload while driving: effects on visual search, discrimination, and decision making. Journal of experimental psychology: Applied, 9, 119.

SAE (2000). J2396 Surface Vehicle Recommended Practice, Definitions and Experimental Measures Related to the Specification of Driver Visual Behavior Using Video Based Techniques. Society of Automotive Engineers, Warrendale, Pa, USA.

SAE (2013). J2944 Operational Definitions of Driving Performance Measures and Statistics (Draft). Society of Automotive Engineers, Warrendale, Pa, USA.

Schuett, J. H., & Walker, B. N. (2013, September). Measuring comprehension in sonification tasks that have multiple data streams. In Proceedings of the 8th Audio Mostly Conference (p. 11). ACM.

Sodnik, J., Dicke, C., Tomažič, S., & Billinghurst, M. (2008). A user study of auditory versus visual interfaces for use while driving. International Journal of Human-Computer Studies, 66, 318-332.

Son, J., Park, M., & Oh, H. (2012, October). Sensitivity of Multiple Cognitive Workload Measures: A Field Study Considering Environmental Factors. In Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, ACM.

Stutts, J.C., Reinfurt, D.W., Staplin, L., Rodgman, E.A., 2001. The Role of Driver Distraction in Traffic Crashes. AAA Foundation for Traffic Safety, Washington, DC.

Victor, T. W., Harbluk, J. L., & Engström, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. Transportation Research Part F: Traffic Psychology and Behaviour, 8, 167-190.

Wickens, C. D. (1981). Processing Resources in Attention, Dual Task Performance, and Workload Assessment (No. EPL-81-3/ONR-81-3). Illinois University at Urbana Engineering-Psychology Research Lab.

Wickens, C. D. (2002). Multiple resources and performance prediction. Theoretical issues in ergonomics science, 3, 159-177.

Wickens, C. D. (2008). Multiple resources and mental workload. Human Factors: The Journal of the Human Factors and Ergonomics Society, 50, 449-455.

Young, K., Regan, M., & Hammer, M. (2007). Driver distraction: A review of the literature. Distracted driving. Sydney, NSW: Australasian College of Road Safety, 379-405.

Young, K. L., Regan, M. A., & Lee, J. D. (2008). Measuring the effects of driver distraction: Direct driving performance methods and measures. In M. A. Regan, J. D. Lee, & K. Young (Ed.), Driver distraction: Theory, effects and mitigation, 31-40. Boca Raton, FL: CRC.

Zhao, S., Brumby, D. P., Chignell, M., Salvucci, D., & Goyal, S. (2013). Shared input multimodal mobile interfaces: Interaction modality effects on menu selection in single-task and dual-task environments. Interacting with Computers, 25, 386-403.