Car crashes rank among the leading causes of death in the United States.

Measuring Cognitive Distraction in the Automobile

June 2013

AAA Foundation for Traffic Safety

Title

Measuring Cognitive Distraction in the Automobile *(June 2013)*

Authors

David L. Strayer, Joel M. Cooper, Jonna Turrill, James Coleman, Nate Medeiros-Ward, and Francesco Biondi *(University of Utah)*

Acknowledgements

About the Sponsor

AAA Foundation for Traffic Safety
607 14th Street, NW, Suite 201
Washington, DC 20005
202-638-5944
www.AAAFoundation.org

Founded in 1947, the AAA Foundation in Washington, D.C. is a not-for-profit, publicly supported charitable research and education organization dedicated to saving lives by preventing traffic crashes and reducing injuries when crashes occur. Funding for this report was provided by voluntary contributions from AAA/CAA and their affiliated motor clubs, from individual members, from AAA-affiliated insurance companies, as well as from other organizations or sources.

Table of Contents

Overview

Driver distraction from secondary in-vehicle activities is increasingly recognized as a significant source of injuries and fatalities on the roadway. The National Highway Traffic Safety Administration (NHTSA) has issued driver distraction guidelines to address visual and manual sources of distraction, but there are currently no published standards that explicitly and exclusively apply to cognitive distraction.

The goal of the current research was to establish a systematic framework for measuring and understanding cognitive distraction in the vehicle. In this report, we describe three experiments designed to systematically measure cognitive distraction.

The first experiment served as a control in which participants performed eight different tasks without the concurrent operation of a motor vehicle. In the second experiment, participants performed the same eight tasks while operating a high-fidelity driving simulator. In the third experiment, participants performed the eight tasks while driving an instrumented vehicle in a residential section of a city.

In each experiment, the tasks involved 1) a baseline single-task condition (i.e., no concurrent secondary task), 2) concurrent listening to a radio, 3) concurrent listening to a book on tape, 4) concurrent conversation with a passenger seated next to the participant, 5) concurrent conversation on a hand-held cell phone, 6) concurrent conversation on a hands-free cell phone, 7) concurrent interaction with a speech-to-text interfaced e-mail system, and 8) concurrent performance with an auditory version of the Operation Span (OSPAN) task. Each task allows the driver to keep his or her eyes on the road and, with the exception of the hand-held cell phone condition, hands on the steering wheel, so any impairment to driving must stem from cognitive sources associated with the diversion of attention from the task of operating the motor vehicle.

We used a combination of performance indices to assess mental workload, including reaction time and accuracy in response to a peripheral light detection task (the detection reaction task [DRT]: ISO, 2012), subjective workload measures from the NASA Task Load Index (NASA TLX: Hart & Staveland, 1988), and physiological measures associated with Electroencephalographic (EEG) activity and Event-Related Brain Potentials (ERPs) time-locked to the peripheral light detection task. We also obtained primary-task measures of driving in experiments using the driving simulator and instrumented vehicle.

We used these data to develop a rating system for cognitive distraction where non-distracted single-task driving anchored the low-end (Category 1), and the OSPAN task anchored the high-end (Category 5) of the scale. In-vehicle activities such as listening to the radio (1.21) or an audio book (1.75) were associated with a small increase in cognitive distraction, the conversation activities of talking to a passenger in the vehicle (2.33) or conversing with a friend on a hand-held (2.45) or hands-free cell phone (2.27) were associated with a moderate increase in cognitive distraction, and the speech-to-text condition (3.06) had a large cognitive distraction rating.

These findings can be used to help craft scientifically-based policies on driver distraction, particularly as they relate to cognitive distraction stemming from the diversion of attention to other concurrent activities in the vehicle. Some activities, such as listening to the radio or a book on tape, are not very distracting. Other activities, such as conversing with a passenger or talking on a hand-held or hands-free cell phone, are associated with moderate/significant increases in cognitive distraction. Finally, there are in-vehicle activities, such as using a speech-to-text system to send and receive text or e-mail messages, which produced a relatively high level of cognitive distraction. The data suggest that a rush to voice-based interactions in the vehicle may have unintended consequences that adversely affect traffic safety.

Introduction

Driver distraction is increasingly recognized as a significant source of injuries and fatalities on the roadway. In fact, NHTSA estimated that inattention accounted for 25 percent of all police-reported crashes (Ranney, Mazzae, Garrott, & Goodman, 2000; Wang, Knipling, & Goodman, 1996). Other estimates have suggested that inattention was a factor in as many as 35-50 percent of all crashes (Sussman, Bishop, Madnick, & Walker, 1985). More recently, data from the 100-car naturalistic driving study (Dingus et al., 2006) found that inattention was a factor in 78 percent of all crashes and near crashes, making it the single largest crash causation factor in their analysis. However, in each of these analyses, the classification of inattention was a catchall category encompassing a variety of phenomena, including fatigue, driving-related distractions (such as glances to mirrors during a merge), nonspecific eye glances away from the forward roadway, and distraction from secondary in-vehicle activities.
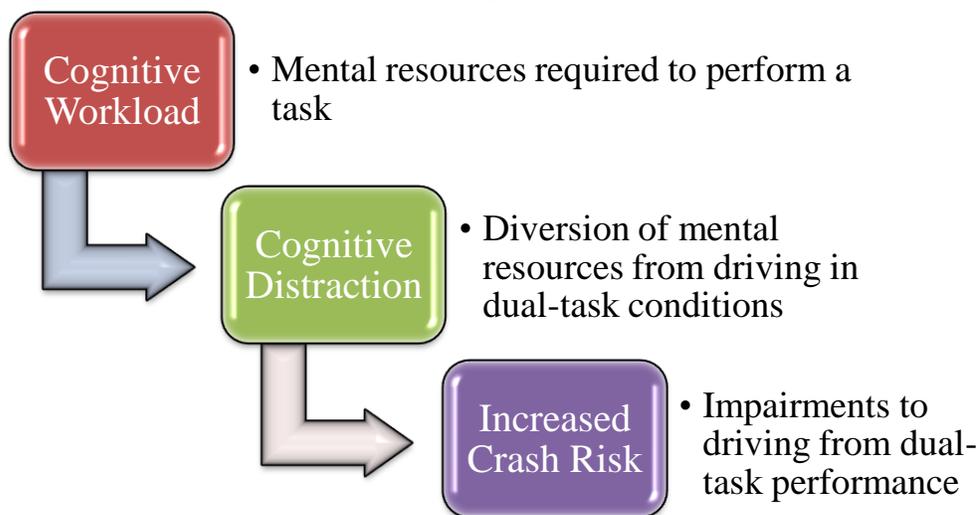
Distractions from secondary in-vehicle activities stems from a combination of three sources (Strayer, Watson, & Drews, 2011). Impairments to driving can arise from a competition for visual processing, an example being when a driver takes his or her eyes off the road to interact with a device. Impairments can also arise from manual interference, as in cases where drivers take their hands off the steering wheel to manipulate a device. Finally, cognitive sources of distraction occur when attention is withdrawn from the processing of information necessary for the safe operation of a motor vehicle. These three sources of distraction can operate independently, and interacting with different devices can result in impairments from one, two, or all three sources. The focus of this report is on developing a valid and sensitive tool for reliably measuring inattention arising from cognitive sources of distraction.

Standardized efforts to evaluate sources of distraction are not new. Indeed, NHTSA has issued driver distraction guidelines to address visual and manual sources of distraction (cf. the visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices entered into the Federal Register on March 15, 2012). Like other published standards, these guidelines specify a number of methods for evaluating the visual (and to a lesser extent, manual) demand of secondary task interactions. However, there are currently no published standards that explicitly and exclusively apply to cognitive distraction.

In fact, cognitive distraction is the most difficult of the three sources of distraction to assess because of the problems associated with observing what a driver's brain (as opposed to hands or eyes) is doing. Furthermore, changes in driving performance associated with cognitive distraction have been shown to be qualitatively different from those associated with visual distraction (Angell et al., 2006; Engström, Johansson, & Östlund, 2005). For example, visual distraction has been shown to increase the variability of lane position, whereas cognitive distraction has been shown to decrease the variability of lane position (Cooper, Medeiros-Ward, & Strayer, in press).

Figure 1 presents a framework for understanding the relationship between cognitive workload, cognitive distraction, and crash risk. For example, talking on a cell phone requires mental resources to perform the conversation task (i.e., the cognitive workload associated with task performance). Performing the cell phone conversation task while driving diverts attention from the driving task. Given the capacity limitations of human attention (Kahneman, 1973), the mental resources available for driving are inversely related to the cognitive workload of the concurrent secondary task (i.e., cognitive distraction is monotonically related to cognitive workload). As attention is diverted from the task of driving, the crash risk increases. Proxies of crash risk include increased brake reaction time (Brown, Lee, & McGehee, 2001; Caird, Willness, Steel, & Scialfa, 2008; Horrey & Wickens, 2006), failure to scan for potential hazards in the driving environment (Taylor et al., 2013), failure to notice objects in the line of sight (Strayer & Drews, 2007), and failures to stop at controlled intersections (Strayer, Watson, & Drews, 2011).[1]

Figure 1. Cognitive workload, cognitive distraction, and crash risk.

Cognitive Workload • Mental resources required to perform a task

Cognitive Distraction • Diversion of mental resources from driving in dual-task conditions

Increased Crash Risk • Impairments to driving from dual-task performance

In order to assess cognitive workload, prior experimental research has typically

---

[1]Note that a similar framework could be developed for visual workload (e.g., the visual demands of performing a task) that when paired with driving lead to visual distraction and result in an increased crash risk. For example, reading a text message requires the eyes to be directed to the cell phone for a certain duration. When paired with driving, this leads to visual distraction where the eyes are off the road (e.g., to read the text message, a competition for visual resources) which increases the crash risk (Victor, Harbluk, & Engström, 2005).

employed some combination of primary-task and secondary-task behavioral measures, physiological measures (e.g., neurological, cardiovascular, and ocular), and subjective workload assessments. Prior research in aviation psychology has used these measures to assess the cognitive workload of pilots (Kramer, Sirevaag, & Braun, 1987; Sirevaag et al., 1993). For example, Kramer et al. (1987) examined pilots' workload by comparing their flight performance during takeoff, level flight, holding a heading, and landing. In this study, flight performance, subjective measurements, and brain-based physiological measures all reflected changes in mental workload as the primary task of piloting became more difficult.

Following the lead from aviation psychology, Strayer, Drews, and Johnston (2003; Strayer & Drews, 2007) used an eye tracker in conjunction with an incidental recognition memory paradigm to determine what information in the driving scene participants attended to while operating a motor vehicle. The study found that participants were more than twice as likely to recognize objects encountered in the single-task driving condition as when they were driving and concurrently talking on a hands-free cell phone. Even when the participants' eyes were directed at objects in the driving environment for the same duration, they were less likely to remember them if they were conversing on a cellular phone. Strayer and Drews (2007) suggested that using a cell-phone induces a form of *inattention blindness,* whereby the cell phone conversation diverts attention from processing the information necessary to safely operate a motor vehicle.

Logically, the principal measures of most driver distraction research focus on the primary task of driving. These often include analyses of steering, throttle, and brake inputs, as well as their effects on lateral and longitudinal control. Surprisingly, the effects of cognitive distraction on these primary measures are somewhat subtle and often contradictory. Indeed, two highly cited meta-analyses of cognitive distraction indicated that it does not reliably affect basic lateral or longitudinal control, but that it does reliably degrade reaction time measures (Caird et al., 2008; Horrey & Wickens, 2006).

To evaluate cognitive driver distraction, reaction time is typically measured using sudden onset stimuli (such as a braking lead vehicle or a flashing dashboard light) that require an immediate response from drivers. Results are often interpreted as an indication of a drivers' ability to quickly and safely respond to the sudden appearance of a threat. Reaction time measures show a great deal of consistency, regardless of whether drivers are responding to a lead braking vehicle, to peripherally flashing lights, or to the appearance of unexpected objects. In all cases, drivers engaged in secondary in-vehicle activities are slower to react than drivers who are paying attention to the roadway. Because of the consistent sensitivity of reaction time measures, a new effort is being considered by the International Standards Organization (ISO) to standardize a protocol for reaction time measurement while driving (ISO, 2012). This technique will be discussed in more detail below.

Cognitive distraction can also be measured through a variety of physiological techniques. Among these, direct measures of brain activity may be the most compelling. One approach that shows high promise is to use time-locked signals of

Electroencephalographic (EEG) activity, referred to as Event-Related Brain Potentials (ERPs). This technique provides a window into the brain activity that is associated with responses to imperative driving events (e.g., brake lights on a lead vehicle). Using this technique, Strayer & Drews (2007) found that the brain activity associated with processing the information necessary for the safe operation of a motor vehicle was suppressed when drivers were talking on a cell phone. (See Appendix C for a brief overview of the ERP methodology used in the current research.) These data help to explain why drivers using a cell phone fail to see information in the driving scene and why their response time to scene events is slowed; they do not encode it as well as they do when they are not distracted by the cell phone conversation. In situations where the driver is required to react quickly, the ERP data suggest that those drivers using a cell phone will be less able to do so because of the diversion of attention from driving to the phone conversation.

A second key physiological measure of cognitive distraction involves the examination of eye movements. This measure is important because, like reaction time, it has direct implications for driver safety. Unlike other measures, eye glance coding can be obtained in a fairly non-intrusive manner, coded either in real-time or after-the-fact, and is sensitive to variations in cognitive demand. Across a variety of experimental settings, the general finding is that when drivers engage in a cognitive secondary task distraction, they tend to look away from the forward roadway less often, exhibiting a form of *tunnel vision* (Harbluk & Noy, 2002; Harbluk, Noy, Trbovich, & Eizenman, 2007; McCarley et al., 2004; Recarte & Nunes, 2000; Recarte & Nunes, 2003; Reimer, 2009; Reimer, Mehler, Wang, & Coughlin, 2012; Sodhi & Reimer, 2002). This includes a reduction in the number of glances to side mirrors, the rear-view mirror, the vehicle instrument panel, and even safety critical roadside objects, such as hidden crosswalks or cross traffic threats. In fact, data from the 100-car naturalistic driving study found that active scanning of the driving environment led to fewer crashes and near crashes (Dingus et al., 2006). To the extent that cognitively distracted drivers fail to scan their environment for potential threats, safety will be compromised (Taylor et al., 2013). The measurement of cognitive distraction through eye glance behavior links it to the visual system, suggesting that cognitive and visual distraction may be inseparably linked at some level.

It is important to note that the demonstrations of inattention blindness provide a pure measure of cognitive distraction because participants' eyes were on the road and they were not manually manipulating the phone in dual-task conditions. However, one shortcoming of the literature on cognitive distraction is that it has often assessed various secondary tasks in a piecemeal fashion. While many forms of cognitive distraction have been evaluated (e.g., listening to the radio, talking on a cell phone, talking to a passenger, interacting with a speech-to-text system), no single study has yet analyzed a comprehensive set of common real-world tasks using the same experimental protocol. A number of studies have demonstrated the sensitivity of cognitive distraction metrics to gradations in artificial task difficulty (Mehler, Reimer, & Coughlin, 2012; Reimer, Mehler, Pohlmeyer, & Coughlin, 2006), yet sensitivity to gradations in real-world cognitive tasks has not been clearly established.

A second knowledge gap with respect to cognitive distraction is that there is no comprehensive way for assessing the distraction potential of any single activity and relating that to the distraction potential of other in-vehicle activities. What is needed is a comprehensive method for assessing secondary task cognitive distraction and a method to integrate the results into a simple, meaningful metric. This metric would allow researchers to make definitive statements about how one source of cognitive distraction compares to another. While it is clear that activities such as conversing on a cell phone degrade certain aspects of driving, it is not clear how to interpret the magnitude of the findings. Is the cognitive distraction of cell phone conversation so severe that it is clearly incompatible with safe driving, or is it sufficiently benign that it is nearly indistinguishable from listening to the radio?

In this report, we present the results from three experiments designed to systematically measure cognitive workload in the automobile. The first experiment served as a control in which participants performed eight different tasks without the concurrent operation of a motor vehicle. In the second experiment, participants performed the same eight tasks while operating a high-fidelity driving simulator. In the third experiment, participants performed the eight tasks while driving an instrumented vehicle in a residential section of a city.

In each experiment, the order of the eight tasks was counterbalanced and the tasks involved 1) a baseline single-task condition (i.e., no concurrent secondary task), 2) concurrent listening to a radio, 3) concurrent listening to a book on tape, 4) concurrent conversation with a passenger seated next to the participant, 5) concurrent conversation on a hand-held cell phone, 6) concurrent conversation on a hands-free cell phone, 7) concurrent interaction with a speech-to-text interfaced e-mail system, and 8) concurrent performance with an auditory version of the Operation Span (OSPAN) task. The OSPAN task is a complex span task developed by Turner and Engle (Turner & Engle, 1989) that requires participants to simultaneously perform a math and memorization task. It was chosen to anchor the highest level of cognitive workload.

Each task allows the driver to keep his or her eyes on the road, and with the exception of the hand-held cell phone condition, both hands on the steering wheel, so that any impairment to driving must stem from cognitive sources associated with the diversion of attention from the task of operating the motor vehicle. Based upon prior research (Strayer, Watson, & Drews, 2011), these tasks were hypothesized to reflect increasing levels of cognitive workload and were selected because they are representative of the type of activities commonly engaged in while operating a motor vehicle (Stutts et al., 2003).

In each of the experiments, we used a combination of performance indices to assess mental workload, including reaction time and accuracy in response to a peripheral light detection task (the DRT task: ISO, 2012), subjective workload measures from the NASA Task Load Index (NASA TLX: Hart & Staveland, 1988), and physiological measures associated with EEG activity and ERPs time-locked to the peripheral light detection task. We also obtained primary-task measures of driving in experiments using the driving simulator and instrumented vehicle. By combining these different measures of cognitive workload (see Figure 2), we provide a more comprehensive

assessment than would be afforded by using only one technique (Gopher & Donchin, 1986; Sirevaag et al., 1993).

Figure 2. The four categories of dependent measures that were combined to generate the cognitive distraction metric.
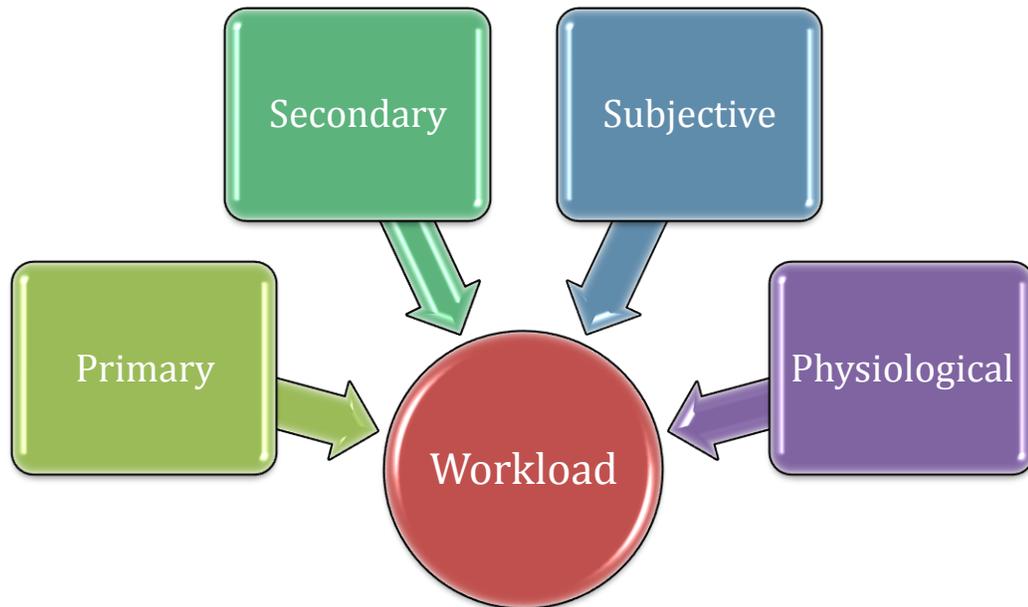
After describing the methods and results of each study in greater detail, we report a meta-analysis that integrates the different dependent measures across the three studies to provide an overall cognitive distraction metric for each of the concurrent secondary tasks. In particular, we used these data to develop a rating system for cognitive distraction where non-distracted single-task driving anchored the low end (Category 1), and the OSPAN task anchored the high end (Category 5) of the scale. For each of the other tasks, the relative position compared to the low and high anchors provides an index of the cognitive workload for that activity when concurrently paired with the operation of a motor vehicle.

Experiment 1: Baseline Assessment

Experiment 1 was designed to provide a baseline assessment of the eight tasks described above. In this controlled assessment, participants were seated in front of a computer monitor that displayed a static fixation cross, and they performed the conditions without the added task of driving. The objective was to establish the *cognitive workload* associated with each activity and to thereby predict the accompanied *cognitive distraction* from performing that activity while operating a motor vehicle (cf. Figure 1).

Experiment 1 Methods

Participants: Thirty-eight participants (20 men and 18 women) from the University of Utah participated in the experiment. Participants ranged in age from 18 to 30 years, with an average age of 22.2 years. All had normal neurological functioning, normal or corrected-to-normal visual acuity, normal color vision (Ishihara, 1993), a valid driver's license, and were fluent in English. Participants' years of driving experience ranged from 2.5 to 14.5, with an average of 6.9 years.[2] All participants owned a cellular phone and reported that they used their phone regularly while driving. They were recruited via University-approved flyers posted on campus bulletin boards. Interested individuals contacted an e-mail address for further information and to schedule an appointment. Eligible participants reported a clean driving history (e.g., no at-fault crashes or history of traffic violations).

Materials: Subjective workload ratings were collected using the NASA TLX survey developed by Hart and Staveland (1988). After completing each of the conditions, participants responded to each of the six items on a 21-point Likert scale ranging from "very low" to "very high." The questions in the NASA TLX were:

> a) *How mentally demanding was the task?*
> b) *How physically demanding was the task?*
> c) *How hurried or rushed was the pace of the task?*
> d) *How successful were you in accomplishing what you were asked to do?*
> e) *How hard did you have to work to accomplish your level of performance*?
> f) *How insecure, discouraged, irritated, stressed, and annoyed were you?*

Equipment: Cellular service was provided by Sprint. The cellular phone was manufactured by Samsung (Model M360) and the hands-free earpiece was manufactured by Jawbone (Model Era). Participants dialed a friend or family member and the volume for both the cellular phone and the hands-free earpiece was adjusted prior to the task.

NaturalReader 10.0 software was used to simulate an interactive messaging service with text-to-speech features. Participants indicated friend names prior to beginning the study. These names were entered into a template containing generic e-mail and text messages (e.g., "*Text from _____:'Hey! Let's meet for lunch sometime this week. When are you free?'*"). Participants were given a short list of commands (i.e., *Repeat, Reply, Forward, Delete, Next Message*) that were used in order for the messaging program to respond. The NaturalReader program was controlled by the experimenter who reacted to the participants' verbal commands, mimicking a speech

---

[2] Our research cohort is representative of a young experienced driver with approximately seven years of driving experience. In contrast, research suggests that novice drivers (Caird, Chisholm, Edwards, & Creaser, 2007) or older drivers (Strayer & Drews, 2004) are likely to experience greater levels of workload than our research cohort because the task of driving is more attention demanding (i.e., consumes more resources for novice and older drivers), and because of capacity and processing speed declines with senescence (Salthouse, 1996; West, 1996). Consequently, the workload estimates obtained with our research cohort provide a conservative estimate of the workload experienced by novice or older drivers when they interact with the same in-vehicle systems.

detection system with perfect fidelity. If a participant did not use the correct command, the NaturalReader program would not continue.

Hosted on a 32-bit research laptop, NeuroScan 4.5 software was used to collect continuous EEG in the experiment. The EEG was recorded using a NeuroScan32-electrode NuAmp amplifier. The EEG was filtered online with a low pass filter of 50hz and a high pass filter set to DC with a sample A/D rate of 250 Hz. The DRT software communicated with the NeuroScan system via a parallel port connection to create event markers associated with the continuous EEG. These event markers allowed for offline stimulus-locked analysis of the EEG recordings (i.e., the DRT stimuli were used for the creating of time-locked ERPs). The EEG was first visually inspected for artifact and any sections with excessive noise from movement or electronic interference were removed. Next, the influence of blinks on the EEG was corrected using ocular artifact rejection techniques (Semlitsch, Anderer, Schuster, & Presslich, 1986) and the data was epoched 200ms before to 1200ms after the onset of the green target light. These epochs were then filtered with a bandpass, zerophase shift filter of 0.1 to 10 Hz. Finally, events that exceeded an artifact rejection criterion of 100 μV were rejected and the remaining events were averaged to obtain one subject's average waveform for each condition in the experiment.

Procedure: Prior to their appointment time, participants were sent a general demographic survey. Upon arrival at the lab in the Behavioral Sciences building, participants read and signed the University of Utah IRB approved consent document and the research team placed an EEG cap on the participant and ensured cap fit. Measuring EEG involved using a cap with built-in electrodes configured based upon the International 10–20 system (Jasper, 1958). Dry sponges (QuickCell™ cellulose-based electrodes manufactured by Compumedics) were placed in each electrode location in preparation for cap use. Saline was applied to the sponges so that they expanded to make contact with the surface of the participant's head, with all impedances below 10kΩ. A reference electrode was placed behind the left ear on the mastoid bone and electrode site FP1 served as the ground. Electrooculogram (EOG) electrodes were placed at the lateral canthi of both eyes (horizontal) and above and below the left eye (vertical) to track eye movements and record eye blinks for later data processing. Participants' field of view and normal range of motion were not impeded when wearing the EEG cap.

Participants were asked to complete eight different 10-minute conditions that were chosen to provide a range of cognitive workload. These conditions were counterbalanced across participants using a Latin Square design. The participants were seated an average of 65cm from a computer screen displaying a fixation cross. Participants were asked to look forward and avoid moving the eyes during the completion of each task.

Described here in hypothesized ascending order of cognitive workload, the single task condition was selected to provide a baseline of cognitive workload (i.e., no concurrent secondary task). In the second condition, participants were allowed to select a radio station to which they normally listen when driving. Depending on the participant's selection, the live radio broadcast was a mix of music and talking.

Before the condition began, participants selected the station and adjusted the volume to a comfortable level. To avoid the influence of manual manipulations, they were not allowed to change the station once they began the recording session.

In the third condition, participants choose from three audio book excerpts. They selected from portions of Chapter 1 from *The Giver* by Lois Lowry, portions of chapter 20 from *Water for Elephants* by Sara Gruen, or portions of Chapter 11 from *Harry Potter and the Sorcerer's Stone* by J. K. Rowling. Once again, all manual adjustments to volume were made before the condition began. Participants were informed that at the end of the audio book, they would take a simple five-item quiz about the events of their chosen audio book. This quiz was to ensure that participants attended to the story; across the three experiments reported, the accuracy on the quiz averaged 86 percent.

Conditions four through six involved different forms of conversation. In each of the conditions, the interlocutors were asked to speak and listen in equal proportions (i.e., 50% speaking and 50% listening). The fourth condition entailed conversation with the experimenter seated behind the participant. Participants wrote down a few conversation topics at the beginning of the study. Experimenters would ask the participant to start telling an interesting story from the list and then helped to maintain an engaging conversation by asking questions about the story or by responding with a story of their own.

The fifth condition required the participant to call a friend or family member and talk with that person on a handheld cellular phone. The participant confirmed that the conversation partner could talk for the duration of the condition (approximately 10 minutes). The call was initiated and the volume was adjusted before the condition began. Because the microswitch for the DRT task (see below) was attached to the left thumb, participants held the phone with their right hand. Most participants indicated that this was the hand they normally used to hold their hand-held device.

Similarly, the sixth condition was a conversation with either the same or a different friend or family member, but it occurred via the hands-free Bluetooth earpiece. Participants indicated in which ear they wished to use the hands-free earpiece. The adjustable earpiece was selected to fit the participant's unique ear size and shape, and then the call was initiated and the volume was adjusted before beginning the condition. As with condition five, the participant confirmed in advance that the conversation partner could talk for the duration of the condition

For condition seven, the participant interacted with a text-to-speech program, NaturalReader 10.0, which simulated speech-based e-mail and text messaging services. Participants interacted with the program as if it were a fully automated system. Perfect speech recognition capabilities were implemented using the "Wizard-of-Oz" paradigm (Kelley, 1983; Lee, Caven, Haake, & T. L. Brown, 2001) in which the participant's speech was actually being secretly entered into the computer by the experimenter with perfect fidelity. Prior to beginning to condition, the participant was familiarized with the program's basic commands, which were: *Repeat*, *Reply*, *Forward*, *Delete*, and *Next Message*. The participant completed a simple tutorial to become familiar with how the commands functioned.

The final condition provided the highest level of cognitive workload: solving simple math problems and remembering words. An auditory version of the OSPAN task required participants to solve sets ranging from two to five math problems and remember as many words in the correct serial order. The math and memory problems were read aloud by the experimenter and the participant spoke his or her answers, which were recorded by the experimenter. Participants were given a short example of the OSPAN task before beginning the condition.

The OSPAN task creates a challenging dual-task condition (Sanbonmatsu, Strayer, Medeiros-Ward, & Watson, 2013). Participants completed an auditory version of the OSPAN task developed by Watson and Strayer (2010) in which they attempted to recall single syllable words in serial order while solving mathematical problems. In the auditory OSPAN task, participants were asked to remember a series of two to five words that were interspersed with math-verification problems (e.g., given "[3 / 1] – 1 = 2?" – "cat" – "[2 x 2] + 1 = 4?" – "box" – RECALL, the participant should have answered "true" and "false" to the math problems when they were presented and recalled "cat" and "box" in the order in which they were presented when given the recall probe). The mathematical problems could be repeated as many times as the participant required. Measures of memory and math performance were recorded for later analysis.

In each of these conditions, participants also performed the DRT task (ISO, 2012). The DRT task presented red or green lights every three to five seconds via a head-mounted device. Red lights were presented 80 percent of the time and green lights were presented 20 percent of the time. Both the color of the light and the interval between trials (e.g., 3-5 seconds) was randomized (i.e., this is a 20/80 oddball with stimuli presented in a Bernoulli sequence with an interstimulus interval of 3-5 seconds). Using a go/no-go design, participants were instructed to respond to the green light as quickly as they could by depressing a microswitch that was placed on the participants' left thumb, but to not respond to the red lights. The lights remained illuminated until a response was made or one second had elapsed.

The DRT device is modeled after an ongoing ISO effort to generate guidelines for measuring cognitive distraction (ISO, 2012). A red/green LED light was mounted on the participant's head via a headband (see Figure 3). The headband did not interfere with the continuous EEG/EOG data collection or with a clear field of view. The light was adjusted to an average 15° to the left and 7.5° above the participant's left eye. Response reaction time was recorded with millisecond accuracy.

Experiment 1 Results

Detection Reaction Time (DRT): The DRT data reflect the manual response to the red and green lights in the peripheral detection task. The reaction time (RT) and accuracy data for the DRT task are plotted in Figures 4 and 5 (Appendix B), respectively. RT for correct responses (i.e., green light responses) was measured to the nearest msec and the accuracy data were converted to the non-parametric measure of sensitivity, A', where a response to a green light was coded as a "hit," non-responses to a red light were coded as a "correct rejection," non-responses to a

green light were coded as a "miss," and responses to a red light were coded as a "false alarm" (Pollack & Norman, 1964)[3]. A repeated measures Analysis of Variance (ANOVA) found that RT increased across condition, $F(7, 259) = 33.87$, $p < .01$, partial $\eta^2 = .48$, and that A' decreased across condition, $F(7, 259) = 7.62$, $p < .01$, partial $\eta^2 = .17$.

Figure 3. The DRT device used in the current research.

NASA Task Load Index (TLX): The data for the six NASA TLX subjective workload ratings are plotted in Figures 6-11 (Appendix B). In each of the figures, the eight conditions are plotted across the abscissa and the 21-point Likert scale workload rating is represented on the ordinate, ranging from "very low," 1, to "very high," 21. The subjective workload ratings increased systematically across the conditions, with the notable exception of physical workload that remained relatively flat (with a noticeable bump in the hand-held cell phone condition). Given that the conditions were selected to allow the drivers to keep their eyes on the road and their hands on wheel, the physical workload ratings are consistent with expectation.

A series of repeated measures ANOVAs found that NASA TLX ratings increased for mental workload, $F(7, 259) = 83.12$, $p < .01$, partial $\eta^2 = .69$, physical workload, $F(7, 259) = 3.33$, $p < .01$, partial $\eta^2 = .08$, temporal demand, $F(7, 259) = 28.74$, $p < .01$, partial $\eta^2 = .44$, performance, $F(7, 259) = 14.92$, $p < .01$, partial $\eta^2 = .29$, effort, $F(7, 259) = 64.87$, $p < .01$, partial $\eta^2 = .64$, and frustration, $F(7, 259) = 33.79$, $p < .01$, partial $\eta^2 = .48$.

Physiological Measures: Figures 12-19 (Appendix B) present the grand average ERP waveforms obtained in Experiments 1-3 at the midline Parietal electrode site (Pz) that were time-locked to the onset of green lights in the DRT task. In each of the

---

[3]Formally, A' measures the average area under the receiver operating characteristic curve (Parasurman & Davies, 1984) and is computed as A' = 1.0-0.25*((p(false alarm)/p(hit)) + (1-p(hit))/(1-p(false alarm))).

figures, the amplitude in microvolts is cross-plotted with time in msec. A close inspection reveals a well-defined P2-N2-P300 ERP component structure in Experiment 1 that becomes noisier in the driving simulator and instrumented vehicle studies (to be discussed in Experiments 2 and 3, respectively). We focused on the P300 component of the ERP because of its sensitivity to cognitive workload, and we measured both its peak latency and the amplitude.

In Figure 20, P300 peak latency, measured as the point in time of maximum positivity in a window between 400 and 700 msec, is plotted for each of the conditions in the experiment. An ANOVA found that P300 latency systematically increased across the conditions, $F(7,259)=13.80$, $p < .01$, partial $\eta^2 = .27$. The P300 amplitude was quantified by computing the average area under the curve between 400 and 700 msec. Figure 21 (Appendix B) plots P300 amplitude as a function of condition. An ANOVA found a main effect of condition, $F(7, 259) = 4.02$, $p < .01$, partial $\eta^2 = .10$; however, there was no systematic pattern in this effect.

Figure 20. P300 Peak Latency from Experiment 1.
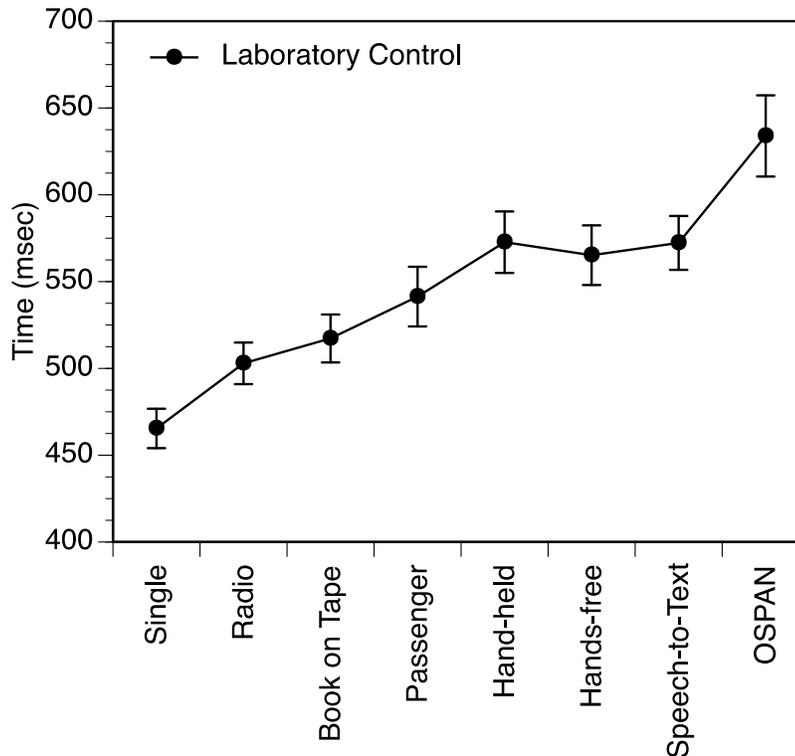
Experiment 1 Discussion

Experiment 1 was designed to provide a baseline assessment of several activities commonly engaged in while operating a motor vehicle (Stutts et al., 2003). In this assessment, participants did not drive but were seated in front of a computer monitor that displayed a static fixation cross. Identical to the more dynamic experiments reported below, participants were fitted with a head-mounted DRT device and an electrode cap for measuring EEG. Participants completed each of the

secondary tasks for 10 minutes while simultaneously responding to green lights from the DRT device. After completing each of the eight tasks, subjective workload ratings were taken.

Taken together, cognitive workload increased with condition, with single-task driving anchoring the low end and the OSPAN condition anchoring the high end. Clearly, not all in-vehicle activities have the same level of cognitive workload. Indeed, some of the more traditional in-vehicle activities, such as listening to the radio, were associated with negligible increases in cognitive workload. By contrast, some of the newer technologies, such as speech-to-text interactions with e-mail, were associated with some of the highest levels of workload. It is noteworthy that the in-vehicle activities that were evaluated were "pure" measures of cognitive workload in that the tasks did not require participants to divert their eyes from the road or their hands from the steering wheel.

The results from the different measures obtained in Experiment 1 had a good correspondence and together help lay the foundation for a metric of cognitive workload that increases across the conditions. As the cognitive workload associated with performing an activity increases, the cognitive distraction associated with performing that activity while operating a motor vehicle increases. Given the capacity limits of attention (Kahneman, 1973), performing an in-vehicle task that is associated with significant cognitive workload leaves less attention to be allocated to the task of driving (cf. Figure 1). That is, cognitive distraction is the consequence of performing an attention-demanding concurrent activity while driving (i.e., driving performance in Experiments 2 and 3 below should be adversely affected by in-vehicle cognitive workload).

Experiment 2: Driving Simulator

The goal of Experiment 2 was to extend the findings from Experiment 1 to operating a high-fidelity driving simulator. Given the increase in cognitive workload associated with performing the different in-vehicle activities, we expected that measures of driving performance would be adversely affected with their concurrent performance.[4] The driving simulator used a car following paradigm on a multilane highway with moderate traffic. Participants followed a lead vehicle that braked a-periodically throughout the scenario and, in addition to the measures collected in Experiment 1, we also collected brake reaction time and following distance, as these variables associated with the primary task of driving have been shown in earlier

---

[4] Despite the fact that the processing requirements of the DRT are minimal, it is possible that its inclusion could increase the cognitive workload of the driver compared to driving conditions without the DRT task. To test for this possibility, we ran another control experiment with 19 participants using the same protocol as Experiment 2, but without the DRT task. We focused on the subjective workload ratings from the NASA TLX because these measures were common to both experiments. A 2 (DRT task load: with vs. without the DRT task) X 8 (Condition) between subjects Multivariate Analysis of Variance (MANOVA) was performed to test for differences in subjective workload associated with the imposition of the DRT task. The MANOVA revealed a significant main effect of Condition, $F(35,15)=8.67$, $p<.01$, partial $\eta^2 = .95$; however, neither the main effect of DRT task load nor the DRT task load X Condition interaction were significant (all p's > .25). This establishes that the imposition of the DRT task did not increase the cognitive workload of the driver.

research to be sensitive to cognitive distraction (Caird et al., 2008; Horrey & Wickens, 2006).

Experiment 2 Methods

Participants: Thirty-two participants (22 men and 10 women) from the University of Utah participated in the experiment. Participants ranged in age from 19 to 36, with an average age of 23.5 years. All had normal neurological functioning, normal or corrected-to-normal visual acuity, normal color vision (Ishihara, 1993), a valid driver's license, and were fluent in English. All participants owned a cellular phone and reported that they used their phone regularly while driving. They were recruited via University-approved flyers posted on campus bulletin boards. Interested individuals contacted an e-mail address for further information and to schedule an appointment. Eligible participants reported a clean driving history (e.g., no at-fault crashes or history of traffic violations).

Equipment: In addition to the equipment used in Experiment 1, the present study used a fixed-base high fidelity driving simulator (made by L-3 Communications) with high-resolution displays providing a 180-degree field of view (see Figure 22). The dashboard instrumentation, steering wheel, gas, and brake pedals are from a Ford Crown Victoria sedan with an automatic transmission. The simulator incorporated vehicle dynamics, traffic-scenario, and road-surface software to provide realistic scenes and traffic conditions. In the driving simulator, the DRT task was implemented by mounting the red/green light on the vehicle dashboard directly in front of the participant. All other equipment was identical to that used in Experiment 1.

Figure 22. Image of the L3 Patrol-Sim driving simulator with a subject using a hand-held cell phone.

Procedure: The procedures used in Experiment 1 were also used in Experiment 2, with the following modifications. In Experiment 2, we used a car-following paradigm

in which participants drove on a simulated multilane freeway with moderate traffic (approximately 1500 vehicles/lane/hour). Participants followed a pace car that would apply its brakes a-periodically. Participants were not allowed to change lanes to pass the pace car, and were asked to maintain a two-second following distance behind the pace car. Participants were given a five-minute practice session to familiarize themselves with the driving simulator. In the practice session, participants were trained to follow a lead vehicle on the highway at a two-second following distance, braking whenever they saw the lead vehicle's brake lights illuminate. If they fell too far behind the lead vehicle, a horn sounded, cueing them to shorten their following distance. The horn was not used once the experimental testing commenced.

Experiment 2 Results

Driving Performance Measures: We examined two measures of driving performance in Experiment 2 that prior research has established are sensitive to cognitive distraction (Caird et al., 2008; Horrey & Wickens, 2006). Figure 23 presents the *Brake Reaction Time* (RT) measured as the time interval between the onset of the pace car's brake lights and the onset of the participant's braking response (i.e., a 1% depression of the brake pedal). Figure 24 presents the *Following Distance*, measured as the distance between the rear bumper of the pace car and the front bumper of the participant's car at the moment of brake onset. A repeated measures ANOVA found that both RT, $F(7, 217) = 10.11$, $p < .01$, partial $\eta^2 = .25$, and following distance increased across condition, $F(7, 217) = 6.26$, $p < .05$, partial $\eta^2 = .17$. A subsidiary linear mixed model analysis that held following distance constant found that brake RT increased as a function of condition over and above any compensatory effects associated with following distance, $F(7,3972)=12.77$, $p < .01$. These data establish that performing in-vehicle activities that differ in their attentional requirements have differential effects on driving performance (i.e., the greater the cognitive workload associated with a subsidiary in-vehicle activity, the greater the cognitive distraction).

DRT: The RT and accuracy data for the DRT task are plotted in Figures 4 and 5, respectively (Appendix B). A repeated measures ANOVA found that RT increased across condition, $F(7, 217) = 13.51$, $p < .01$, partial $\eta^2 = .30$, and that the area under the curve (A') decreased across condition, $F(7, 217) = 21.54$, $p < .01$, partial $\eta^2 = .41$.

NASA TLX: The data for the six NASA TLX subjective workload ratings are plotted in Figures 6-11 (Appendix B). The subjective workload ratings increased systematically across the conditions. A series of repeated measures ANOVAs found that NASA TLX ratings increased for mental workload, $F(7, 217) = 43.55$, $p < .01$, partial $\eta^2 = .58$, physical workload, $F(7, 217) = 5.03$, $p < .01$, partial $\eta^2 = .14$, temporal demand, $F(7, 217) = 26.92$, $p < .01$, partial $\eta^2 = .47$, performance, $F(7, 217) = 9.27$, $p < .01$, partial $\eta^2 = .23$, effort, $F(7, 217) = 35.48$, $p < .01$, partial $\eta^2 = .53$, and frustration, $F(7, 217) = 23.83$, $p < .01$, partial $\eta^2 = .44$.

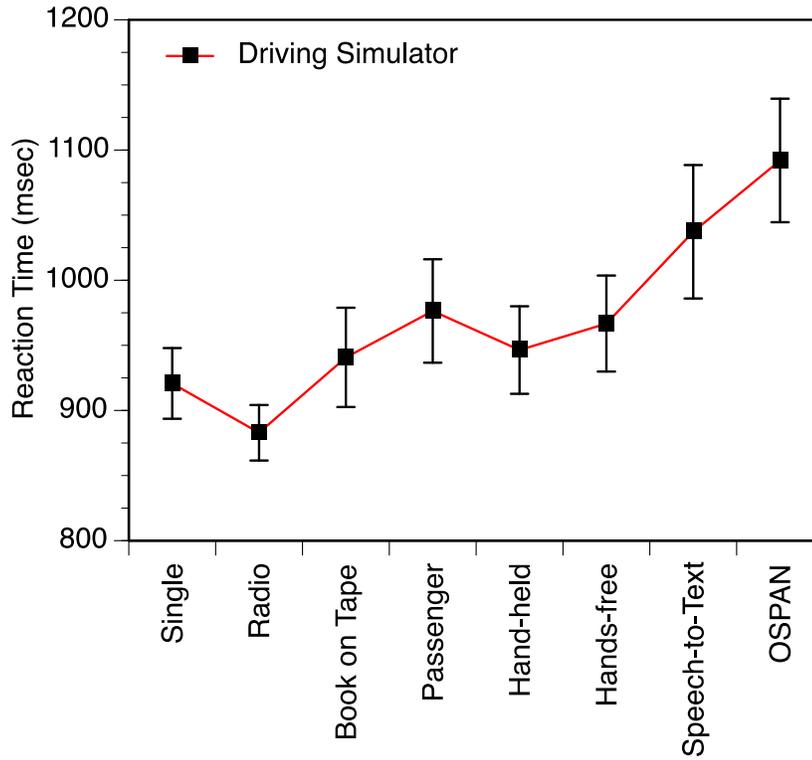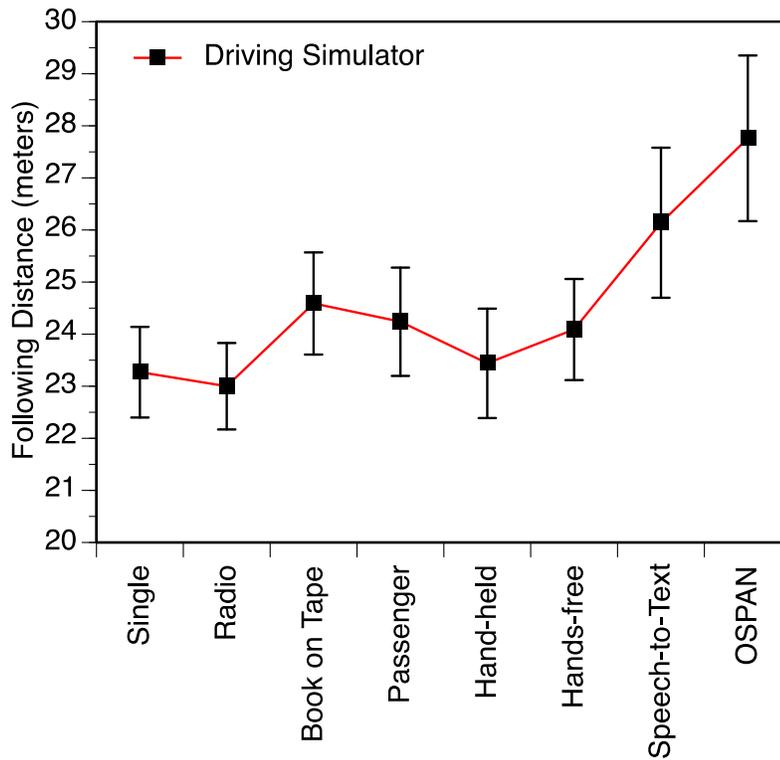**Figure 23. Brake Reaction Time from Experiment 2.**

Figure 24. Following Distance from Experiment 2.

Physiological Measures: EEG was recorded in Experiment 2 using the same protocol as that of Experiment 1. The resulting ERPs are plotted in Figures 12-19 (Appendix B) alongside the same conditions from Experiments 1 and 3. As discussed above, the ERP was degraded as we moved from the laboratory to the driving simulator due to the increased biological noise from eye/head/body movements and electronic noise from the driving simulator. Because of the added noise, we were not able to reliably measure P300 latency. As in Experiment 1, the P300 amplitude data, presented in Figure 21 (Appendix B), were quantified by computing the average area under the curve between 400 and 700 msec. An ANOVA of the P300 amplitude found a main effect of condition, $F(7, 217) = 4.38$, $p < .01$, partial $\eta^2 = .12$. Planned comparisons found that the single-task did not differ from the radio and book on tape, but was significantly different from the conversation and OSPAN tasks.

Experiment 2 Discussion

Experiment 2 replicated and extended the pattern obtained in Experiment 1. Most importantly, the increases in cognitive workload resulted in systematic changes in driving performance compared to non-distracted driving. In particular, brake reaction time to imperative events in the driving simulator systematically increased as a function of the cognitive workload associated with performing the different in-vehicle activities. Importantly, this pattern held even when controlling for the increased following distance drivers adopted in these conditions. The P300 data also replicate our earlier reports of suppressed P300 activity when comparing single-task and hands-free cell phone conditions (Strayer & Drews, 2007).

It is worth considering what the pattern of data would have looked like had participants protected the driving task at the expense of the other in-vehicle activities. In such a case, we would expect that the primary task measures would be insensitive to secondary-task workload (i.e., Figures 21 and 22, Appendix B (and Figure 24), would be flat and there would be no main effect of condition for these measures). Instead, we show that the mental resources available for driving are inversely related to the cognitive workload of the concurrent secondary task. Thus increasing the cognitive workload of the in-vehicle secondary tasks resulted in systematic increases in cognitive distraction.

Experiment 3: Instrumented Vehicle

The purpose of Experiment 3 was to establish that the patterns obtained in the laboratory and driving simulator generalize to the operation of an instrumented vehicle on residential roadways. This is important because the consequences of impaired driving in the city are different from that of a driving simulator (e.g., a crash in the real world can have life-or-death consequences, whereas this is not the case in the driving simulator). Participants drove an instrumented vehicle in a residential section of a city while concurrently performing the eight conditions used in Experiments 1 and 2. If the findings generalize, then there should be a good correspondence between the results of Experiment 3 and those of Experiments 1 and 2.

Experiment 3 Methods

Participants: Thirty-two participants (12 men and 20 women) from the University of Utah participated in the experiment. Participants ranged in age from 18 to 33, with an average age of 23.5 years. All had normal neurological functioning, normal or corrected-to-normal visual acuity, normal color vision (Ishihara, 1993), a valid driver's license, and were fluent in English. Participants' years of driving experience ranged from 2 to 17, with an average of 7.1 years. All participants owned a cellular phone and reported that they used their phone regularly while driving. They were recruited via University-approved flyers posted on campus bulletin boards. Interested individuals contacted an e-mail address for further information and to schedule an appointment. The Division of Risk Management Department at the University of Utah ran a Motor Vehicles Record (MVR) report on each prospective participant to ensure participation eligibility based on a clean driving history (e.g., no at-fault crashes or history of traffic violations). In addition, following University policy, each prospective participant was required to complete a University-devised 20-minute online defensive driving course and pass the certification test.

Equipment: In addition to the equipment used in Experiment 1, Experiment 3 used an instrumented 2010 Subaru Outback (see Figure 25). The vehicle was augmented with four 1080p LifeCam USB cameras that captured the driving environment and participants' facial features. An accelerometer, a high accuracy GPS unit, and steering, brake, acceleration, and rotation sensors recorded driving behaviors for later analysis. All other equipment was identical to Experiment 1.

Figure 25. Participant in the instrumented vehicle talking on a hand-held cell phone.

Procedure: The procedures used in Experiment 1 were also used in Experiment 3, with the following modifications. Prior to their appointment time, participants were sent the University of Utah IRB approved informed consent document, general demographic surveys, and instructions for completing the 20-minute online defensive driving course and the certification test. Prior to the experimental session, we obtained a Motor Vehicle Record report on the driver to ensure a clean driving record.

Before beginning the study, the driver was familiarized with the controls of the instrumented vehicle, adjusted the mirrors and seat, and was informed of the tasks to be completed while driving. The participant drove around a parking lot in order to become familiar with the handling of the vehicle. Next, participants drove one circuit on a 2.75-mile loop in the Avenues section of Salt Lake City, UT in order to become familiar with the route itself. The route provided a suburban driving environment containing nine all-way controlled stop signs, one two-way stop sign, and two stoplights (see Appendix D). A research assistant and an experimenter accompanied the participant in the vehicle at all times. The research assistant sat in the rear and the experimenter sat in the front passenger seat and had ready access to a redundant braking system and notified the driver of any potential roadway hazards.

The driver's task was to follow the route defined above while complying with all local traffic rules, including a 25 mph speed restriction. If drivers began to exceed 25 mph, they were reminded of this restriction by the research team. Throughout each condition, the driver completed the DRT. Each condition lasted approximately 10 minutes, which was the average time required to make one loop around the track. Safety directions were reiterated before each driving condition. At the conclusion of the study, participants returned to the Behavioral Sciences building where the EEG cap was removed and the participants were compensated for their time and debriefed.

Experiment 3 Results

Driving performance: Because participants did not follow a lead vehicle, as they did in Experiment 2, following distance or brake response data were not available for analysis. However, because high definition cameras were used to record the driving environment, manual coding of eye movement data was possible. Prior to analyzing the eye movement data, 24 critical locations were identified for analysis. These included all four-way and two-way stops, as well as pedestrian crosswalks. At each of these critical locations, eye movement data were coded frame-by-frame to record glances to the left and to the right of the forward roadway. Glances to the side mirrors, rearview mirror, and dashboard were also recorded. Scans were recorded as complete if drivers looked to both the left and right. Partially complete scans were recorded where the drivers looked to either the left or right, and incomplete scans were recorded where drivers failed to scan for hazards. Each drive was analyzed by at least two trained coders and any discrepancies in the coding were flagged and reviewed for consistency by a third coder. In general, coders were very accurate and only a small number of events needed to be double-checked.

The eye glance data for each condition are plotted in Figure 26. Repeated measures ANOVA indicated that at critical locations, drivers made progressively fewer scans to the right and left of the forward roadway as cognitive workload increased, $F(7, 168) = 5.92$, $p < .01$, partial $\eta^2 = .20$. These data replicate and extend the important work of Taylor et al., (2013) by establishing that the same failures to scan for roadway hazards observed in a driving simulator are found in an instrumented vehicle. Moreover, our research establishes that that there is a systematic decrease in scanning for hazards as cognitive workload increases.

Figure 26. Glances at Hazard Locations in Experiment 3.
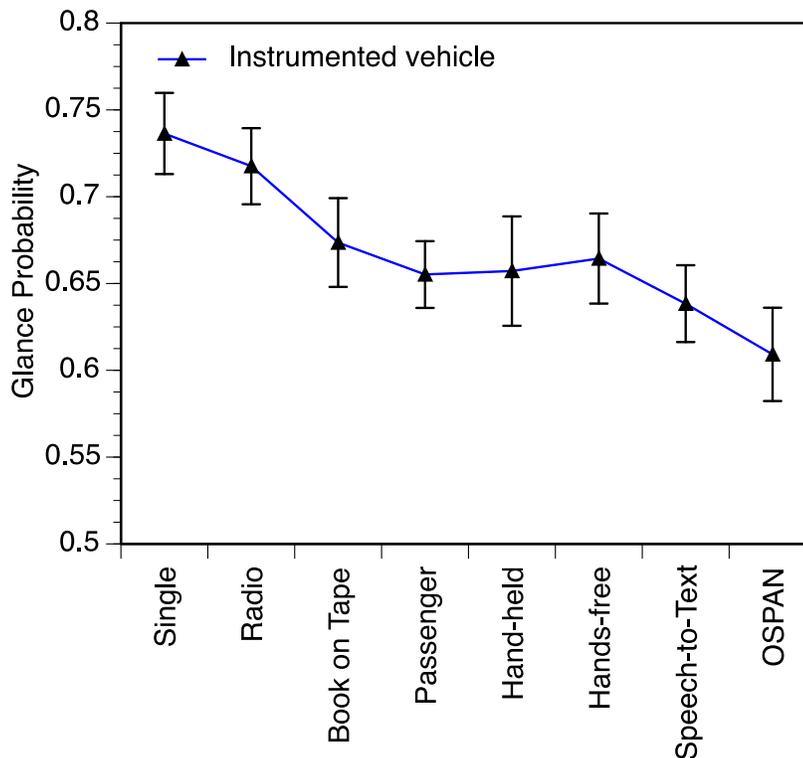
DRT: The RT and accuracy data for the DRT task are plotted in Figures 4 and 5 (Appendix B), respectively. A repeated measures ANOVA found that RT increased across condition, $F(7, 217) = 27.21$, $p < .01$, partial $\eta^2 = .47$, and that A' decreased across condition, $F(7, 217) = 19.17$, $p < .01$, partial $\eta^2 = .38$.

NASA TLX: The data for the six NASA TLX subjective workload ratings are plotted in Figures 6-11 (Appendix B). The subjective workload ratings increased systematically with condition. A series of repeated measures ANOVAs found that NASA TLX ratings increased for mental workload, $F(7, 217) = 52.46$, $p < .01$, partial $\eta^2 = .63$, physical workload, $F(7, 217) = 10.01$, $p < .01$, partial $\eta^2 = .24$, temporal demand, $F(7, 217) = 37.81$, $p < .01$, partial $\eta^2 = .55$, performance, $F(7, 217) = 19.66$, $p < .01$, partial $\eta^2 = .39$, effort, $F(7, 217) = 47.99$, $p < .01$, partial $\eta^2 = .61$, and frustration, $F(7, 217) = 26.06$, $p < .01$, partial $\eta^2 = .40$.

Physiological measures: EEG was recorded in Experiment 3 using the same protocol as that of the prior studies. The resulting ERPs are plotted in Figures 12-19 (Appendix B) alongside the same conditions from Experiments 1 and 2. The P300 component of the ERP was considerably degraded by the added noise of the instrumented vehicle, the added head and eye movements of the drivers as they scanned the driving environment, and the increased cognitive load of the driving task (i.e., driving complexity increased from the laboratory to the driving simulator to the instrumented vehicle). Moreover, the P300 became even less distinct at higher workloads (e.g., while concurrently performing the OSPAN task). As in Experiment 2, we were not able to reliably measure P300 latency. P300 amplitude, plotted in Figure 21 (Appendix B), was quantified by computing the average area under the curve between 400 and 700 msec. An ANOVA failed to find a main effect of condition, but planned comparisons found that the single-task did significantly differ from speech-to-text and OSPAN conditions ($p < .05$).

Experiment 3 Discussion

Experiment 3 replicated and extended the findings from the prior experiments in several important ways. Most importantly, they document that the patterns observed in the controlled laboratory setting of Experiment 1 and in the driving simulator setting of Experiment 2 generalize to what was observed in the instrumented vehicle. There was a systematic increase in cognitive workload across condition and, importantly, driving performance as measured by scanning for potential hazards decreased as a function of in-vehicle condition. This latter finding replicates prior studies (Harbluk & Noy, 2002; Harbluk et al., 2007; McCarley et al., 2004; Recarte & Nunes, 2000; Recarte & Nunes, 2003; Reimer, 2009; Reimer et al., 2012; Sodhi & Reimer, 2002) that have shown that visual scanning behavior is impaired with increases in cognitive workload. As such, it suggests that the diversion of attention from the task of driving results in a degraded representation of the driving environment (i.e., impaired situation awareness of the driving context, Gugerty, 1997; Kass, Cole, & Stanny, 2007). Taken together, the data provide clear evidence of the attentional demands of scanning the driving environment for potential hazards. In particular, scanning the driving environment is an active process that is disrupted by the diversion of attention to subsidiary in-vehicle activities.

General Discussion

The patterns observed in the three experiments reported here are strikingly consistent, establishing that lessons learned in the laboratory and driving simulator are in good agreement with studies of cognitive distraction on the roadway. In each case, they document a systematic increase in cognitive workload as participants performed different in-vehicle activities. The data for the three studies were entered into a MANOVA to determine how cognitive workload changed across condition for the three experiments. For the sake of clarity, we focused our analyses based upon secondary, subjective, and physiological assessments because these measures were identical across the three experiments. Obviously, there were no primary-task driving measures in Experiment 1, and the measures of brake reaction time and

following distance obtained in the simulator were not available in the instrumented vehicle nor were the visual scanning data from the instrumented vehicle available in the simulator.

The DRT task was based on ISO guidelines for measuring cognitive distraction (ISO, 2012). By using a head-mounted version of the DRT, the impact of head and eye movements on detection was negated. As drivers move their heads, the DRT device moved with them and remained in a constant location relative to the eyes. The resulting RT and accuracy data provide a much more finely calibrated metric than the more traditional measures of brake reaction time and following distance (which often co-vary, making unambiguous interpretation difficult). A MANOVA performed on the secondary-task DRT data revealed a significant effect of condition, $F(14,86)=19.58$, $p < .01$, partial $\eta^2 = .76$, experiment, $F(4,198)=26.84$, $p < .01$, partial $\eta^2 = .35$, and a condition X experiment interaction, $F(28,174)=45.01$, $p < .01$, partial $\eta^2 = .45$. Further analysis found a main effect of condition such that RT increased, $F(7,693)=65.02$, $p < .01$, partial $\eta^2 = .40$, and A' decreased, $F(7,693)=26.71$, $p < .01$, partial $\eta^2 = .42$ across condition. In addition, RT increased, $F(2,99)=85.14$ $p < .01$, partial $\eta^2 = .63$ and A' decreased, $F(2,99)=35.78$, $p < .01$, partial $\eta^2 = .42$, from Experiment 1 to 3. On the whole, there is good agreement with the DRT measures across experiments; however, the laboratory- and simulator-based studies would appear to provide a more conservative estimate of the impairments to driving associated with in-vehicle technology use.

A MANOVA performed on the subjective workload ratings revealed a significant effect of condition, $F(42,58)=26.48$, $p < .01$, partial $\eta^2 = .95$, and of experiment, $F(12,190)=2.86$, $p < .01$, partial $\eta^2 = .15$; however, the condition X experiment interaction was not significant. Across experiments, main effects of condition were obtained for mental workload, $F(7,693)=170.79$, $p < .01$, partial $\eta^2 = .63$, physical workload, $F(7,693)=16.08$, $p < .01$, partial $\eta^2 = .14$, temporal demand, $F(7,693)=90.04$, $p < .01$, partial $\eta^2 = .48$, performance, $F(7,693)=44.99$, $p < .01$, partial $\eta^2 = .31$ effort, $F(7,693)=140.92$, $p < .01$, partial $\eta^2 = .59$, and frustration, $F(7,693)=81.16$, $p < .01$, partial $\eta^2 = .45$. The NASA TLX measures also increased from Experiment 1 to 3 for mental workload, $F(2,99)=5.50$, $p < .01$, partial $\eta^2 = .10$, physical workload, $F(2,99)=8.34$, $p < .01$, partial $\eta^2 = .14$, temporal demand, $F(2,99)=8.38$, $p < .01$, partial $\eta^2 = .14$, effort, $F(2,99)=5.04$, $p < .01$, partial $\eta^2 = .09$, and frustration, $F(2,99)=7.98$, $p < .01$, partial $\eta^2 = .13$, but not for performance ($p > .14$). On the whole, the subjective workload measures were in agreement across six sub-scales, eight conditions, and three experiments. In particular, there was a consistent increase in subjective workload ratings from conditions 1-8 and also a systematic increase in subjective workload ratings from Experiments 1-3. The high degree of sensitivity, face validity, and the ease of collection added strengths to these measures.

A MANOVA performed on P300 amplitude revealed a main effect of condition, $F(7,93)=6.67$, $p < .01$, partial $\eta^2 = .33$, experiment, $F(2,99)=13.3$, $p < .01$, partial $\eta^2 = .21$, and a condition X experiment interaction, $F(14,88)=1.88$, $p < .05$, partial $\eta^2 = .12$. Overall, P300 amplitude was similar in magnitude for single-task, radio, and book on tape, smaller in magnitude for the conversation conditions (i.e., conditions 4-

7), and smallest for the OSPAN task. As was evident in Figures 12-19 (Appendix B), P300 amplitude was largest in Experiment 1, intermediate in magnitude for Experiment 2, and smallest in Experiment 3, and this undoubtedly reflects the degraded quality of the ERP signal as the experiments progressed from the laboratory to the driving simulator to the instrumented vehicle. In fact, the P300 amplitude was the noisiest of all the measures we recorded, with contamination from movements of the mouth, jaw, eyes, head, and body accompanied by environmental noise from the simulator and instrumented vehicle. Consequently, the P300 measures were the least sensitive of our measures to changes in cognitive workload and this limitation was most apparent in the instrumented vehicle.

In the main, moving from the laboratory to the driving simulator to the instrumented vehicle increased the intercept of the cognitive workload curves, and similar condition effects were obtained for the different dependent measures. This experimental cross-validation is important in its own right, establishing that the effects obtained in the simulator generalize to on-road driving. In fact, our measures in Experiment 1 were remarkably consistent with those obtained in Experiment 3, suggesting that there are occasions where the added complexity, expense, and risk of on-road study are unnecessary. Moreover, the similarity of the primary, secondary, subjective, and physiological measures provides convergence in our workload assessments. It is noteworthy that these tasks allowed the drivers to have their eyes on the road and their hands on the wheel (except when they were holding the cell phone). That is, these in-vehicle activities are cognitively distracting to different degrees. Whereas the procedural demands of the tasks themselves did not require the driver to divert their eyes from the roadway or to otherwise alter their scanning pattern, performing these cognitively demanding in-vehicle activities clearly altered the visual scanning behavior of the drivers in Experiment 3.

The data from our studies also speak to the fidelity of the driving simulator in studying cognitive distraction. There have been some suggestions that the patterns obtained in the driving simulator are not representative of the real world, perhaps because the consequences of a crash in the simulator are different from the consequences of a crash on the roadway. Although we made no attempt to match the driving scenarios in the simulator and instrumented vehicle, the patterns obtained in Experiments 2 and 3 were virtually identical. If anything, the data obtained in the simulator may underestimate the impairments associated with using different in-vehicle activities on the road (see also, Cooper et al., submitted). It is noteworthy that Experiment 1 provided a low cost alternative to the driving simulator and instrumented vehicle and the data provided in this study were very predictive of driving performance on the roadway (Lee, 2004).

Toward a Standardized Scale of Cognitive Distraction

The primary goal of the current research was to develop a metric of cognitive distraction associated with performing different activities while operating a motor vehicle. Because the different dependent measures are on different scales (e.g., msec, meters, amplitude, etc.), each was transformed to a standardized score. This involved Z-transforming each of the dependent measures to have a mean of 0 and a

standard deviation of 1 (across the experiments and conditions), and the average for each condition was then obtained. The standardized scores for each condition were then summed across the different dependent measures to provide an aggregate measure of cognitive distraction. Finally, the aggregated standardized scores were scaled such that the non-distracted single-task driving condition anchored the low-end (Category 1), and the OSPAN task anchored the high-end (Category 5) of the cognitive distraction scale. For each of the other tasks, the relative position compared to the low and high anchors provided an index of the cognitive workload for that activity when concurrently performed while operating a motor vehicle. The four-step protocol for developing the cognitive distraction scale is listed below.

Step 1: For each dependent measure, the standardized scores across experiments, conditions, and subjects were computed using $Z_i=(x_i-X)/SD$, where X refers to the overall mean and SD refers to the pooled standard deviation.

Step 2: For each dependent measure, the standardized condition averages were computed by collapsing across experiments and subjects (see Table 1, Appendix A, for the standardized condition averages for each dependent measure).

Step 3: The standardized condition averages across dependent measures were computed with an equal weighting for physical, secondary, subjective, and physiological metrics (see Figure 2). Table 1 (Appendix A) lists the 13 dependent measures that were used in the standardized condition averages, separated in grey by the metric of which they are subordinate. The measures within each metric were also equally weighted. For example, the secondary task workload metric comprised an equal weighting of the measures DRT-RT and DRT-A'. Note that eye glances, A', and P300 amplitude were inversely coded in the summed condition averages. Figure 27 (Appendix B) presents the average effect size of the difference between single-task and each of the remaining conditions using either a) the single-task SD (i.e., Glass's Delta), b) the pooled pair-wise SD (i.e., Cohen's D), or the pooled SD. There is obviously a good correspondence between the different effect size calculations.[5]

Step 4: The standardized mean differences were range-corrected so that the non-distracted single-task condition had a rating of 1.0 and the OSPAN task had a rating of 5.0.

$X_i = (((X_i-min)/(max-min))*4.0)+1$

---

[5]The three effect size estimates make different assumptions regarding the standardization process. Glass's Delta compares differences between conditions relative to the single-task standard deviation. Cohen's D compares the differences between conditions to the pair-wise pooled standard deviation. The pooled SD compares the differences between conditions to the pooled standard deviation of all conditions and is similar to the partial η^2 estimate for the main effect of condition in a MANOVA. This latter measure was the most conservative and is most closely related to the cognitive workload metric presented in Figure 28.

The cognitive distraction scale presented in Figure 28 ranges from 1.0 for the single-task condition to 5.0 for the OSPAN task. In-vehicle activities such as listening to the radio (1.21) or an audio book (1.75) were associated with a small increase in cognitive distraction, the conversation activities of talking to a passenger in the vehicle (2.33), or conversing with a friend on a hand-held (2.45) or hands-free cell phone (2.27) were associated with a moderate increase in cognitive distraction, and the speech-to-text condition (3.06) had a large cognitive distraction rating. A comparison of the rating system presented in Figure 28 with the effect size estimates in Figure 27 (Appendix B) provides a quick translation between the different metrics.

Figure 28. Cognitive Distraction Scale.
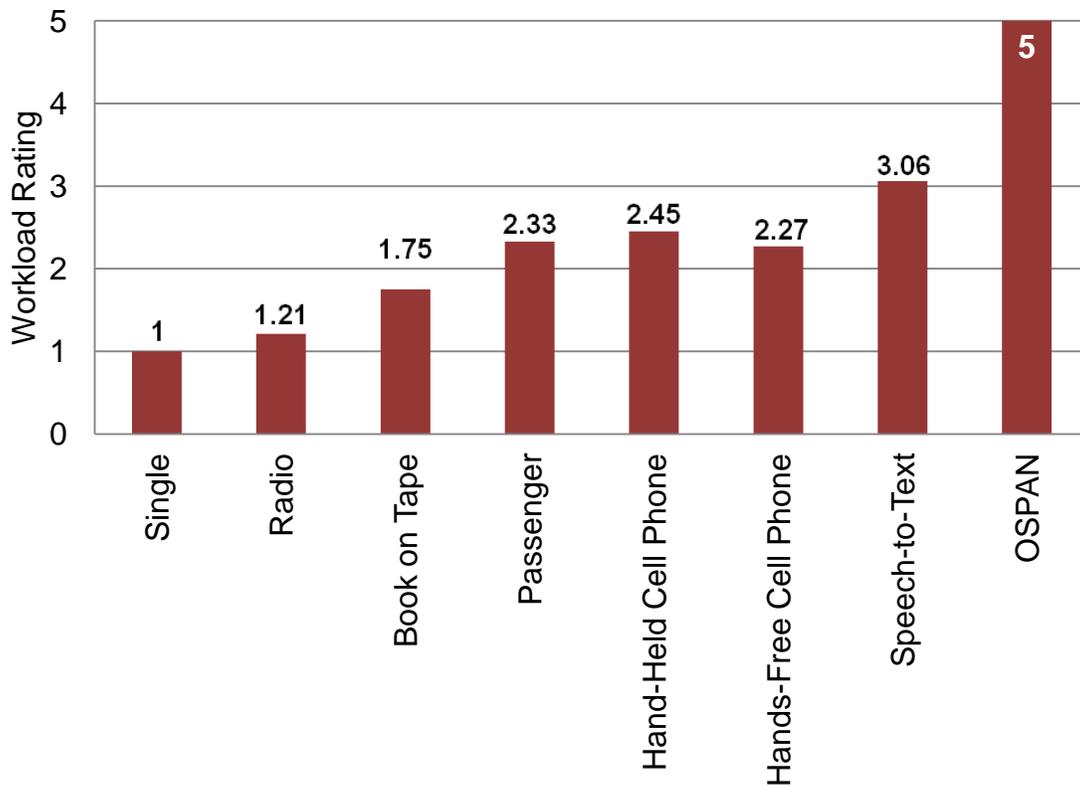
The current research establishes an experimentally validated cognitive distraction instrument that can be used to evaluate different in-vehicle activities. Measuring cognitive distraction has proven to be the most difficult of the three sources of distraction to assess because of the problems associated with observing what a driver's brain (as opposed to hands or eyes) is doing. The current research used a combination of primary, secondary, subjective, and physiological measures to assess cognitive distraction across a variety of in-vehicle activities and provides the most comprehensive analysis that has been performed to date.

These findings can be used to help craft scientifically-based policies on driver distraction, particularly as it relates to cognitive distraction stemming from the

diversion of attention to other concurrent activities in the vehicle. Some activities, such as listening to the radio or a book on tape, are not very distracting. Other activities, such as conversing with a passenger or talking on a hand-held or hands-free cell phone, are associated with moderate/significant increases in cognitive distraction. Finally, there are in-vehicle activities, such as using a speech-to-text system to send and receive text or e-mail messages, which produced a relatively high level of cognitive distraction.

The speech-to-text based system that we evaluated in the current research was associated with a Category-3 level of cognitive distraction. It is worth reiterating that we used a perfect fidelity speech-recognition system and there was no requirement to review, edit, or correct garbled speech-to-text translations. Such is not the case with current technology, but it is improving and may someday achieve perfect fidelity. Given the current trends toward more voice commands in the vehicle, this Category-3 level of cognitive distraction is troubling. The assumption that if the eyes were on the road and the hands were on the steering wheel then voice-based interactions would be safe appears to be unwarranted. Simply put, hands-free does not mean risk-free.

In the current research, conversation with a passenger in the vehicle or with a friend over a cell phone was associated with a Category-2 level of cognitive distraction. In an earlier study comparing passenger and cell-phone conversation (Drews, Pasupathi, & Strayer, 2008), the passenger was allowed to spontaneously help the driver with the task of driving (e.g., helping to navigate, pointing out hazards, or regulating conversation based upon the real-time demands of driving), and significant differences in route navigation were observed. Compared to single-task driving, when the participant was conversing with a friend in the vehicle there was no decline in navigation accuracy. By contrast, a conversation with a friend on a hands-free cell phone resulted in a 50 percent decline in navigation accuracy (i.e., half the participants missed their exit). What is the basis for the discrepancy between the current study and the study reported by Drews, Pasupathi, and Strayer (2008)? One important difference is that the DRT device that was used to measure RT and accuracy and served as a trigger for the ERP recordings was designed so that the *driver* could easily see the device; however, the *passenger* could not see the DRT lights and therefore could not adjust their conversation to aid the driver as they did when navigating to a roadway exit. When the passenger cannot help with the task of driving, as was the case in the current study, then any differences between conversation types should be minimal.

Increasingly, car manufacturers and third-party providers are presenting consumers with options to make movie or dinner reservations, send and receive text or e-mail messages, make postings on Facebook, interact with global position systems, and utilize voice commands for controlling functions of the vehicle. The lessons learned from the current research suggest that such voice-based interaction is not risk-free, and in some instances the impairments to driving may rise to the level associated with drunk driving (McEvoy et al., 2005; Redelmeier & Tibshirani, 1997; Strayer, Drews, & Crouch, 2006). Just because a new technology does not take the eyes off the road does not make it safe to be used while the vehicle is in motion.

Limitations

The cognitive distraction scale provides a comprehensive analysis of several of the cognitive sources of driver distraction. The scale does not directly measure visual/manual sources of distraction, although changes in visual scanning associated with cognitive workload are certainly included in the metric. As of yet, there is not a comprehensive mapping of cognitive distraction to on-road crash risk. It is reasonable to assume that there would be a monotonic relationship between cognitive distraction and crash risk. However, there are some points of contact between epidemiological studies and our cognitive distraction scale (e.g., cell phone conversations), but more reference points are clearly needed.

Finally, while driver distraction can theoretically be separated into visual, manual, and cognitive sources, this sort of balkanization may prove overly simplistic in the real world. In Experiment 3, we demonstrated that there is cross-talk between cognitive and visual processing of potential hazards on the roadway. Moreover, a task that has a high visual demand (e.g., text messaging) is also likely to require cognitive resources to read and process the message. Even when there are no demands for visual processing, interacting with cognitively demanding in-vehicle devices can alter where and how drivers look in the driving environment.

Summary and Conclusions

The goal of the current research was to establish a systematic instrument for measuring and understanding cognitive distraction in the vehicle, and this has been accomplished. Using that instrument, we established that there are significant impairments to driving that stem from the diversion of attention from the task of operating a motor vehicle, and that the impairments to driving are directly related to the cognitive workload of these in-vehicle activities. Moreover, compared to the other activities studied (e.g., listening to the radio, conversing with passengers, etc.) we found that interacting with the speech-to-text system was the most cognitively distracting. This clearly suggests that the adoption of voice-based systems in the vehicle may have unintended consequences that adversely affect traffic safety.

References

Angell, L., Auflick, J., Austria, P. A., Kochhar, D., Tijerina, L., Biever, W., Diptiman, T., Hogsett, J., & Kiger, S. (2006). *Driver workload metrics task 2 final report*. Washington, DC: DOT HS 810 635.

Brown, T. L., Lee, J. D., & McGehee, D. V. (2001). Human performance models and rear-end collision avoidance algorithms. Human Factors, *43*, 462–482.

Caird, J. K., Chisholm, S. L., Edwards, C. J., & Creaser, J. I. (2007). The effect of yellow light onset time on older and younger drivers' perception response time (PRT) and intersection behavior. *Transportation Research Part F: Traffic Psychology And Behaviour*, *10*, 383-396.

Caird, J. K., Willness, C. R., Steel, P., & Scialfa, C. (2008). A meta-analysis of the effects of cell phones on driver perfomance. *Accident Analysis and Prevention, 40*, 1282-1293.

Cooper, J. M., Medeiros-Ward, N., Strayer, D. L. (in press). The impact of eye movements and cognitive workload on lateral position variability.

Cooper, J. M., Yager, C. E., Ingebretsen, H., & Chrysler, S. T. (Under Review). An investigation of the effects of reading and writing text-based messages while driving.

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., Knipling, R. R. (2006). *The 100-car naturalistic driving study: phase II -- Results of the 100-car field experiment*. Washington, DC: DOT HS 810 593.

Drews, F. A., Pasupathi, M., & Strayer, D. L. (2008). Passenger and cell-phone conversation during simulated driving. *Journal of Experimental Psychology: Applied, 14*, 392-400.

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

Gopher, D., & Donchin, E. (1986). Workload: An experimentation of the concept. In K. Boff, L. Kaufmann, & J. Thomas, *Handbook of Perception and Performance (Vol. 2): Cognitive Processes and Performance* (pp. 1-49). New York: Wiley.

Gugerty, L. (1997). Situation awareness during driving: explicit and implicit knowledge in dynamic spatial memory. *Journal of Experimental Psychology: Applied, 3*, 42-66.

Harbluk, J. L., Noy, Y. I., Trbovich, P. L., & Eizenman, M. (2007). An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance. *Accident Analysis & Prevention*, 372-379.

Harbluk, L., & Noy, I. (2002). *The impact of cognitive distraction on driver visual behaviour and vehicle control*. Canada: Ergonomics Division, Road Safety Directorate and Motor Vehicle Regulation Directorate.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock, & N. Meshkati, *Human Mental Workload*. Amsterdam: North Holland Press.

Horrey, W. J., & Wickens, C. D. (2006). Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human Factors, 48*, 196-205.

Ishihara, S. (1993). *Ishihara's test for color-blindness*. Tokyo: Kanehara.

ISO. (2012). Road vehicles -- Transport information and control systems -- Detection-Response Task (DRT) for assessing selective attention in driving. ISO TC 22 SC 13 N17488 (Working Draft). *Under development by Working Group 8 of ISO TC22, SC 13.*

Jasper, H. A. (1958). The ten-twenty system of the International Federation. *Electroencepholography and Clinical Neurophysiology, 10*, 371-375.

Kahneman , D. (1973). Attention and effort. Englewood Cliffs, NJ: Prentice-Hall, 1973.

Kass, S. J., Cole, K. S., & Stanny, C. J. (2007). Effects of distraction and experience on situation awareness and simulated driving. *Transportation Research Part F, 10*, 321-329.

Kelley, J. F. (1983). An empirical methodology for writing user-friendly natural language computer applications. *Proceedings of ACM SIG-CHI '83 Human Factors in Computing Systems* (pp. 193-196). Boston: New York, ACM.

Kramer A. F., Sirevaag E. J., Braun R. *(*1987*)*. A psychophysiological assessment of operator workload during simulated flight missions. *Human Factors, 29*, 145–160.

Lee, J. D. (2004). Simulator Fidelity: How low can you go? *48th Annual Meeting of the Human Factors and Ergonomics Society*. Santa Monica, CA.

Lee, J. D., Caven, B., Haake, S., & Brown, T. L. (2001). Speech-based interactions with in-vehicle computers; The effect of speech-based e-mail on drivers' attention and roadway. *Human Factors, 43*, 631-640.

McCarley, J. S., Vais, M., Pringle, H., Kramer, A. F., Irwin, D. E., & Strayer, D. L. (2004). Conversation disrupts scanning and change detection in complex visual scenes. *Human Factors, 46*, 424-436.

McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. *Science, 211*, 77-80.

McEvoy, S. P., Stevenson, M. R., McCartt, A. T., Woodward, M., Haworth, C., Palamara, P., & Cercarelli, R. (2005). Role of mobile phones in motor vehicle crashes resulting in hospital attendance: A case-crossover study. *British Medical Journal*, 428-433.

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, 54,* 396-412.

NHTSA. (2012). Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehcile Electronic Devices. Department of Transportation. Docket No. NHTSA-2010-0053.

Parasuram, R., & Davies, D. R. (1984). Varieties of Attention. Academic Press.

Pollack, I.,& Norman, D. A. (1964). Non-parametric analysis of recognition experiments. Psychonomic Science,1 ,125-126.

Ranney, T., Mazzae, E., Garrott, R., & Goodman, M. (2000). NHTSA Driver Distraction Research: Past, Present and Future [online].

Recarte, M. A., & Nunes, L. M. (2000). Effects of verbal and spatial-imagery tasks on eye fixations while driving. *Journal of Experimental Psychology: Applied, 6*, 31-43.

Recarte, M. A., & Nunes, L. M. (2003). Mental workload while driving: effects on visual search, discrimination, and decision making. *Journal of Experimental Psychology: Applied*, 119-137.

Redelmeier, D. A., & Tibshirani, R. J. (1997). Association between cellular-telephone calls and motor vehicle collisions. *The New England Journal of Medicine, 336*, 428-458.

Reimer, B. (2009). Impact of cognitive task complexity on drivers' visual tunneling. *Transportation Research Record: Journal of the Transportation Research Board, 2138*, 13-19.

Reimer, B., Mehler, B., Wang, Y., & Coughlin, J.F. (2012). A field study on the impact of variations in short term memory demands on drivers' visual attention and driving performance across three age groups. *Human Factors, 54,* 454-468.

Reimer, B., Mehler, B. L., Pohlmeyer, A. E., & Coughlin, J. F. (2006). The use of heart rate in a driving simulator as an indicator of age-related differences in driver workload. *Advances in Transportation Studies*, 9-29.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological review*, *103*, 403-427.

Sanbonmatsu, D. M., Strayer, D. L., Medeiros-Ward, N., & Watson, J. M. (2013). Who multi-tasks and why? Multi-tasking ability, perceived multi-tasking ability, impulsivity, and sensation seeking. *PLOS ONE, 8*, 1-8.

Semlitsch, H. V., Anderer, P., Schuster, P., & Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology*, *23*, 695-703.

Sirevaag, E. J., Kramer, A. F., Coles, M. G., & Donchin, E. (1989). Resource reciprocity: An event-related brain potentials analysis. *Acta Psychologica*, *70*, 77-97.

Sirevaag, E. J., Kramer, A. F., Wickens, C. D., Reisweber, M., Strayer, D. L., & Grenell, J. H. (1993). Assessment of pilot performance and mental workload in rotary wing aircraft. *Ergonomics, 9*, 1121-1140.

Sodhi, M., & Reimer, B. (2002). Glance analysis of driver eye movements to evaluate distraction. *Behavior Research Methods, Instruments, & Computers, 34*, 529-538.

Strayer, D. L., & Drews, F. A. (2004). Profiles in Driver Distraction: Effects of Cell Phone Conversations on Younger and Older Drivers. *Human Factors*, *46*, 640-649.

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

Strayer, D. L., Drews, F. A., & Crouch, D. J. (2006). Comparing the cell-phone driver and the drunk driver. *Human Factors, 48*, 381-391.

Strayer, D. L., Drews, F. A., & Johnston, W. A. (2003). Cell phone induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied, 9*, 23-52.

Strayer, D. L., Watson, J. M., & Drews, F. A. (2011). Cognitive distraction while multitasking in the automobile. In B. Ross, *The Psychology of Learning and Motivation* (pp. 29-58).

Stutts, J., Feaganes, J., Rodman, E., Hamlet, C., Meadows, T., Rinfurt, D., Gish, K., Mercadante, M., & Staplin, L. (2003). Distractions in Everyday Driving. *AAA Foundation for Traffic Safety (on-line publication).*

Sussman, E. D., Bishop, H., Madnick, B., & Walker, R. (1985). Driver inattention and highway safety. *Transportation Research Record, 1047*, 40-48.

Taylor, T., Pradhan, A. K., Divekar, G., Romoser, M., Muttart, J., Gomes, R., Pollatsek, A., & Fisher, D. L. (2013). The view from the road; the contribution of on-road glance-monitoring technologies to understanding driver behavior. *Accident Analysis and Prevention.*

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory & Language, 28*, 127-154.

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.

Wang, J. -S., Knipling, R. R., & Goodman, M. J. (1996). The role of driver inattention in crashes: New statistics from the 1995 Crashworthiness Data System. *40th Annual Proceedings of the Association for the Advancement of Automotive Medicine*, (pp. 377-392). Vancouver Canada.

Watson, J. M., & Strayer, D. L. (2010). Profiles in Extraordinary multitasking ability. *Psychonomic Bulletin and Review, 17*, 479-485.

West, R.L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin*, *120*, 272–292.

Wickens, C., Kramer, A., Vanasse, L., & Donchin, E. (1983). Performance of concurrent tasks: A psychophysiological analysis of the reciprocity of information-processing resources. *Science*, *221*, 1080-1082.

Appendix A: Standardized Scores for Each Dependent Measure

Table 1. Standardized scores for each dependent measure. Note that the primary task measures of Brake RT and following distance (FD) were collected in Experiment 2, Glances at hazards were collected in Experiment 3, the secondary-task measures of DRT-RT and A' were collected in Experiments 1-3, the NASA TLX subjective workload measures of mental workload, physical workload, temporal demand, performance, effort, and frustration were collected in Experiments 1-3, and the physiological measures of P3 Latency (P3 Lat.) was collected in Experiment 1 and P3 Area measures were obtained in Experiments 1-3. Conditions 1-8 refer to single-task, radio, book on tape, passenger conversation, hand-held cell phone conversation, hands-free cell phone conversation, speech-to-text, and OSPAN, respectively.

	1	2	3	4	5	6	7	8
Brake RT	-.280	-.381	-.186	-.050	-.073	.015	.355	.600
FD	-.310	-.263	-.017	-.028	-.126	-.042	.243	.544
Glances	.527	.380	.036	-.108	-.092	-.036	-.239	-.468
DRT-RT	-.387	-.318	-.305	.161	.106	.071	.131	.541
DRT-A'	.248	.195	.223	-.004	-.005	.024	-.072	-.608
Mental	-.640	-.558	-.145	-.289	-.162	-.171	.410	1.554
Physical	-.321	-.289	-.136	-.138	.400	-.018	.158	.345
Temporal	-.495	-.486	-.220	-.277	-.070	-.172	.435	1.285
Performance	-.454	-.419	-.061	-.101	-.070	-.056	.113	1.049
Effort	-.599	-.488	-.184	-.326	-.112	-.184	.379	1.513
Frustration	-.413	-.428	-.218	-.386	-.129	-.040	.278	1.337
P3 Lat.	-.728	-.391	-.262	-.045	.237	.170	.233	.788
P3 Area	-.011	.130	.139	-.254	.078	.258	.092	-.431

Figure 4. DRT Reaction Time across Experiments 1-3.
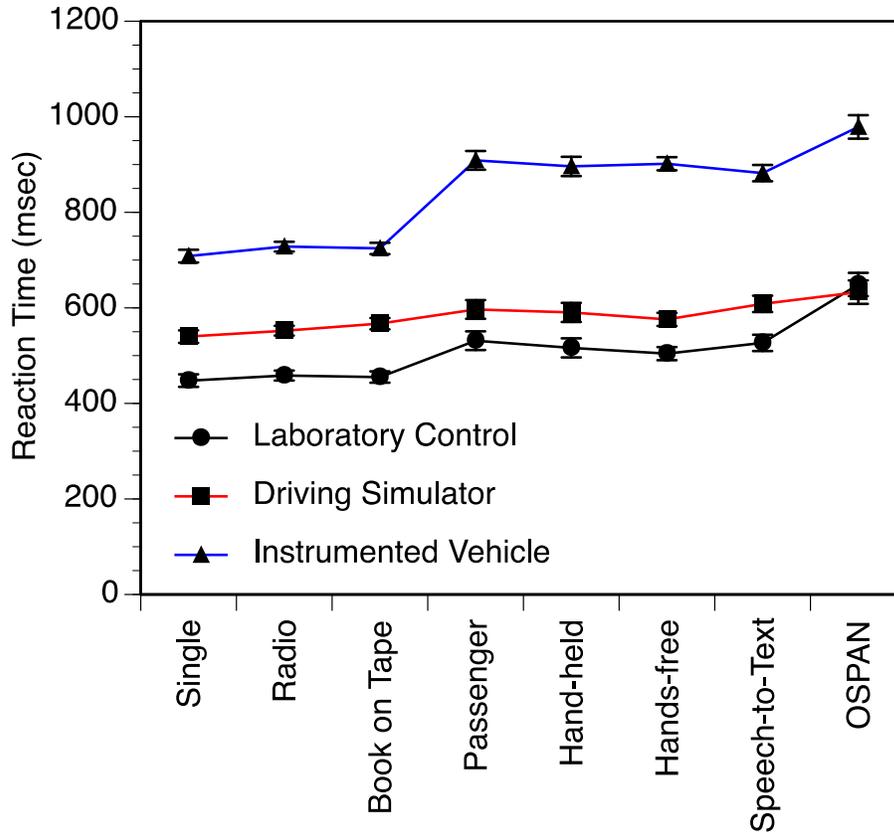
Figure 5. DRT A' across Experiments 1-3.
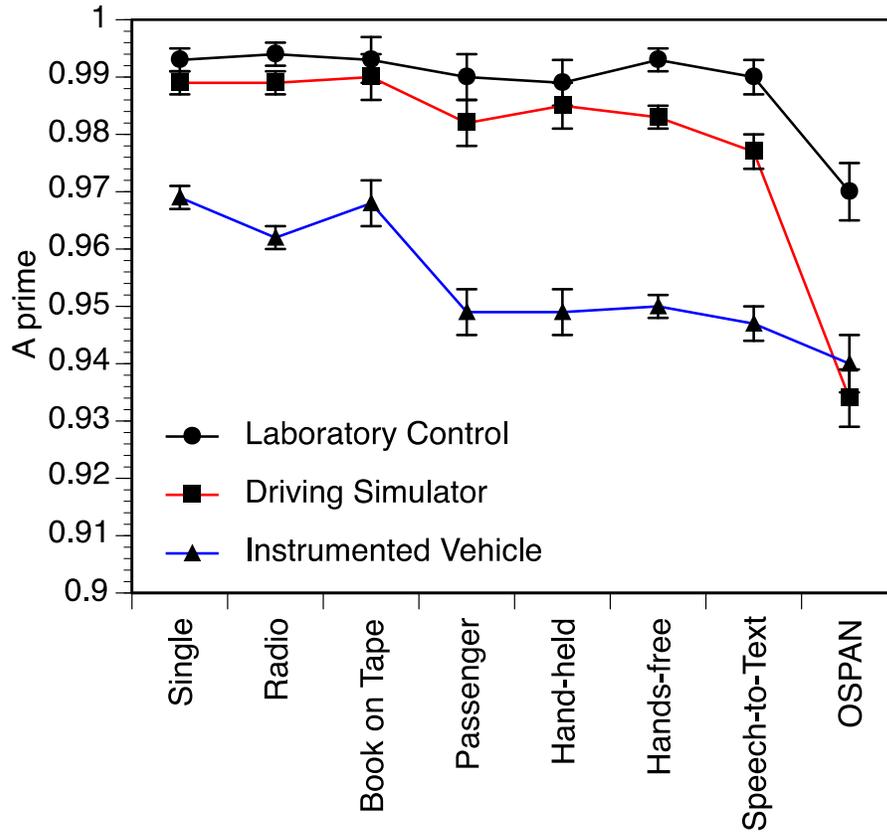
Figure 6. NASA-TLX – Mental Demand across Experiments 1-3.
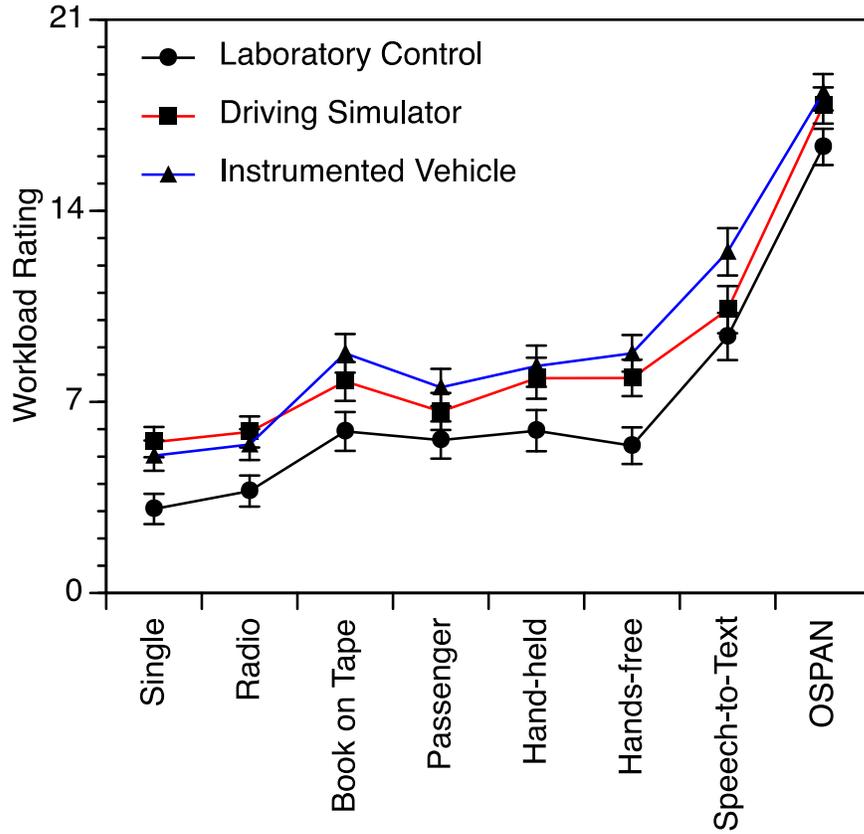
Figure 7. NASA-TLX – Physical Demand across Experiments 1-3.
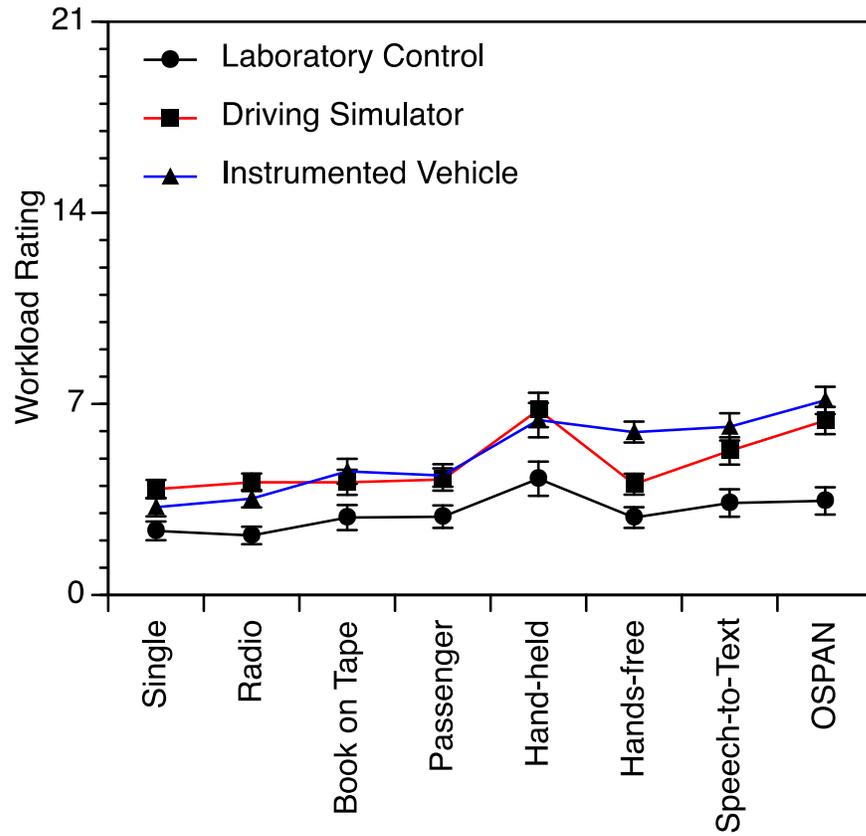
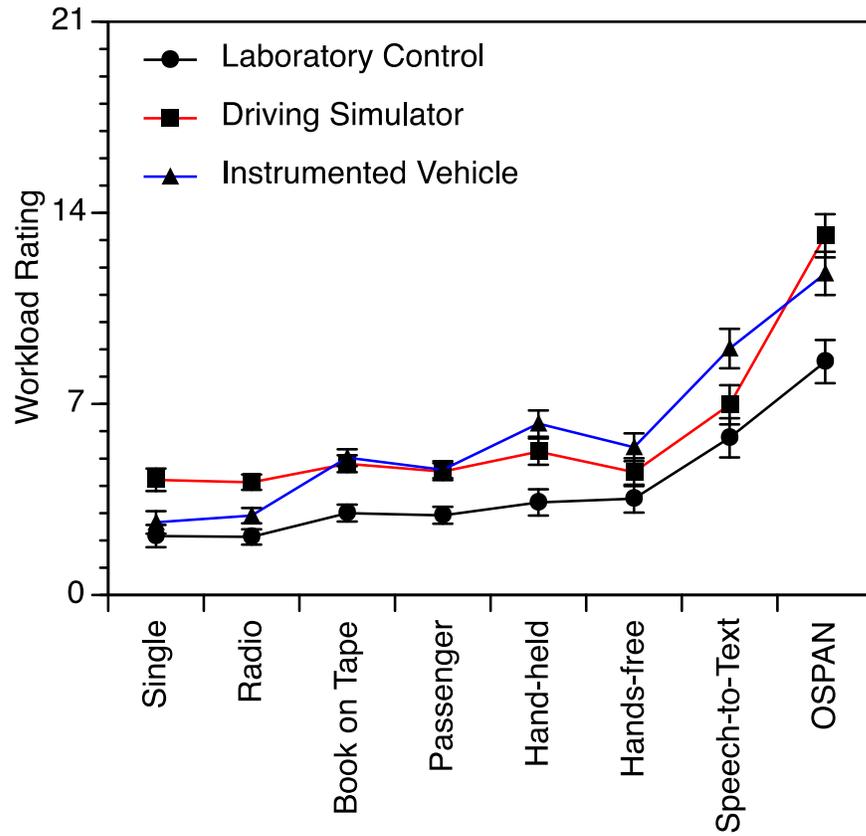**Figure 8. NASA-TLX – Temporal Demand across Experiments 1-3.**
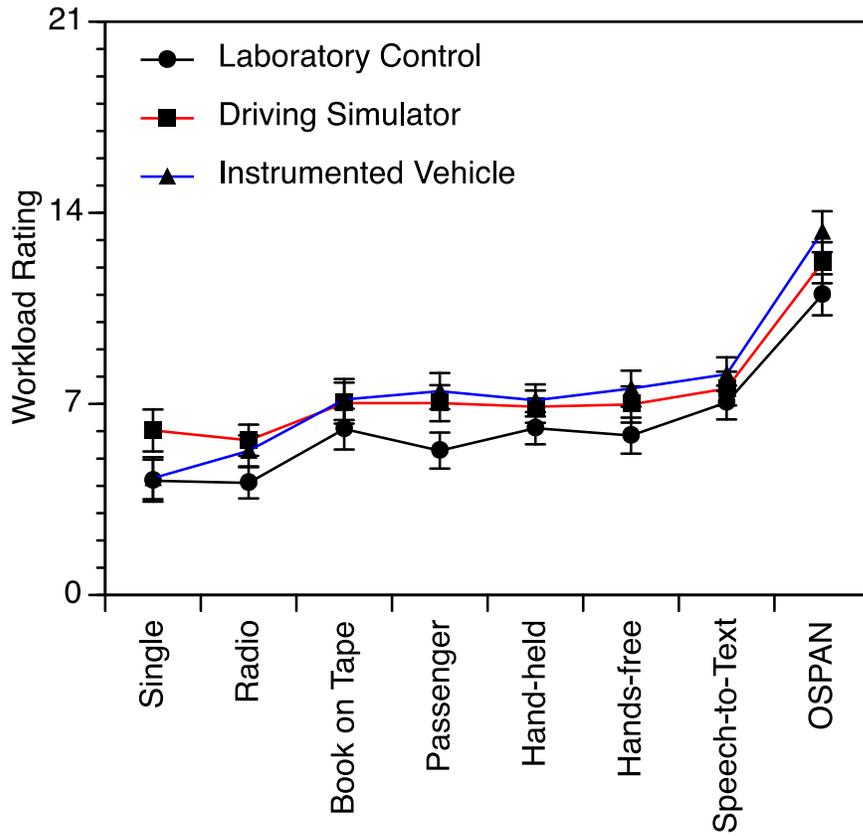
Figure 9. NASA-TLX – Performance across Experiments 1-3.
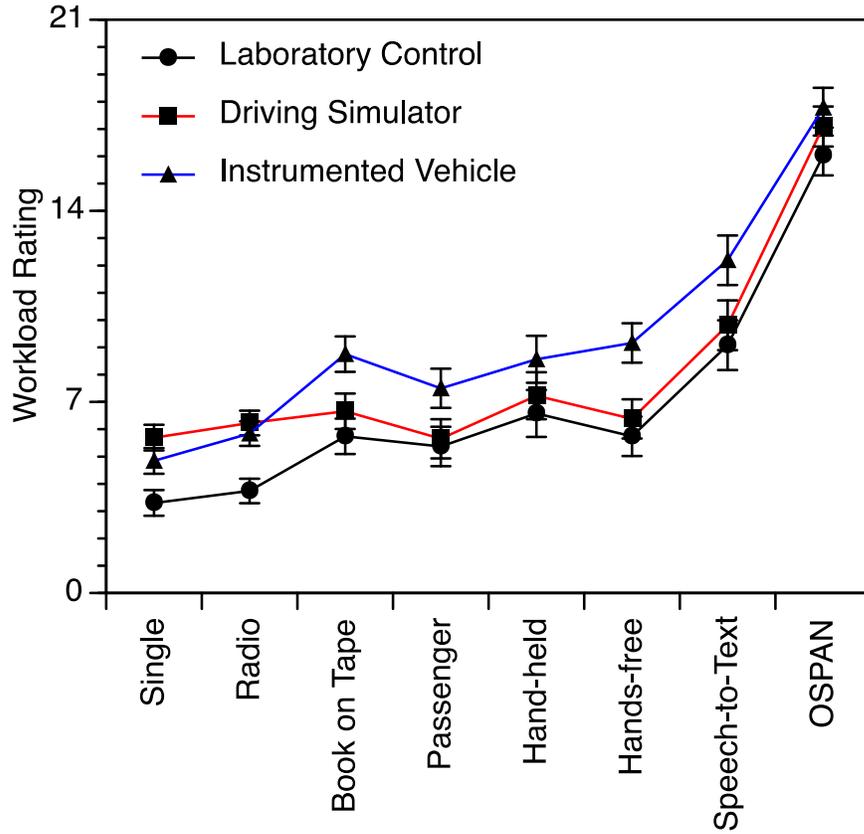
Figure 10. NASA-TLX – Effort across Experiments 1-3.
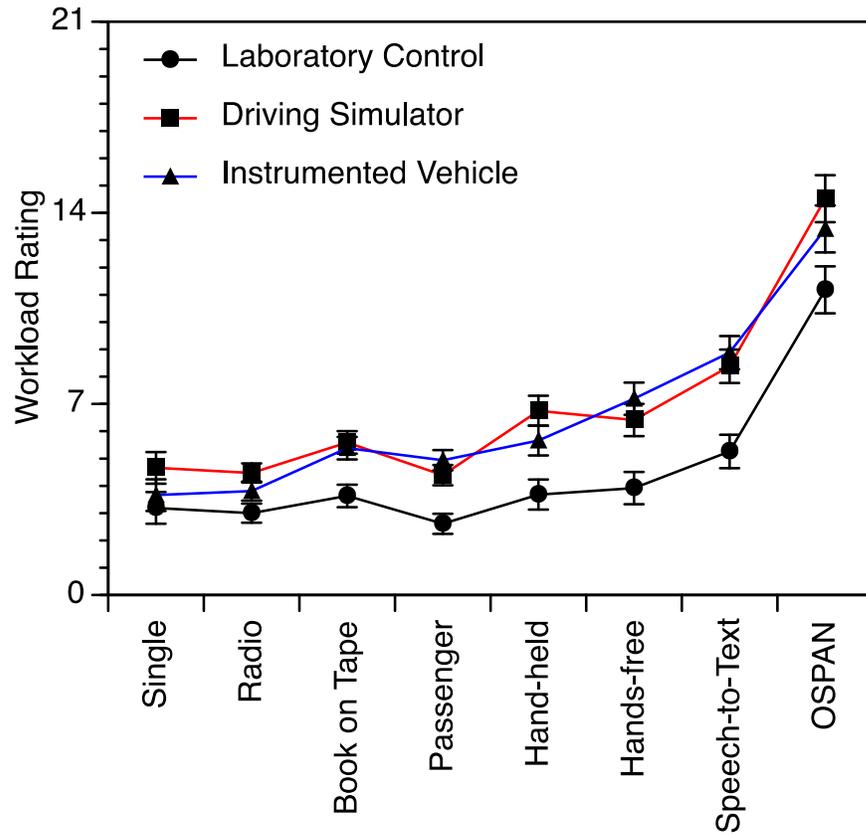
Figure 11. NASA-TLX – Frustration across Experiments 1-3.

Figure 12. Single-task ERPs at Pz across Experiments 1-3.
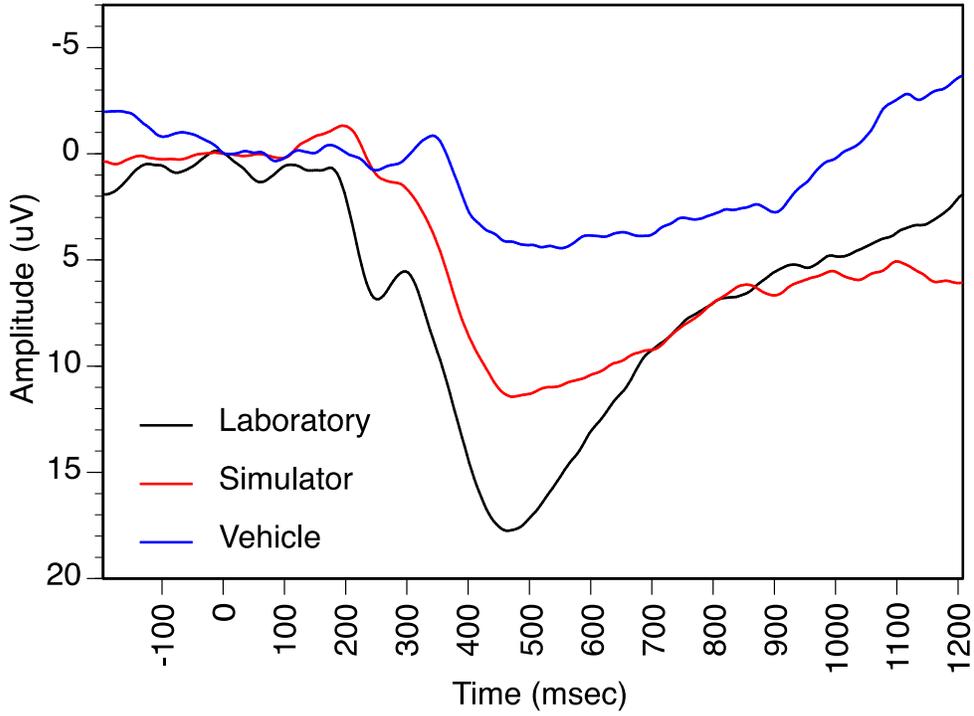
Figure 13. Radio ERPs at Pz across Experiments 1-3.
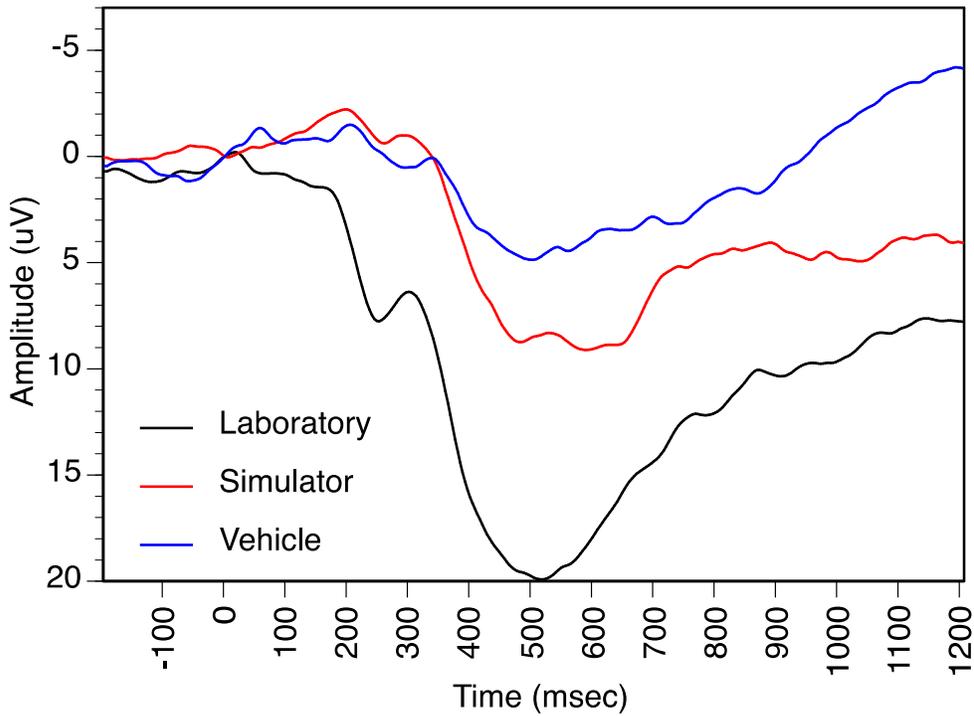
Figure 14. Book on Tape ERPs at Pz across Experiments 1-3.
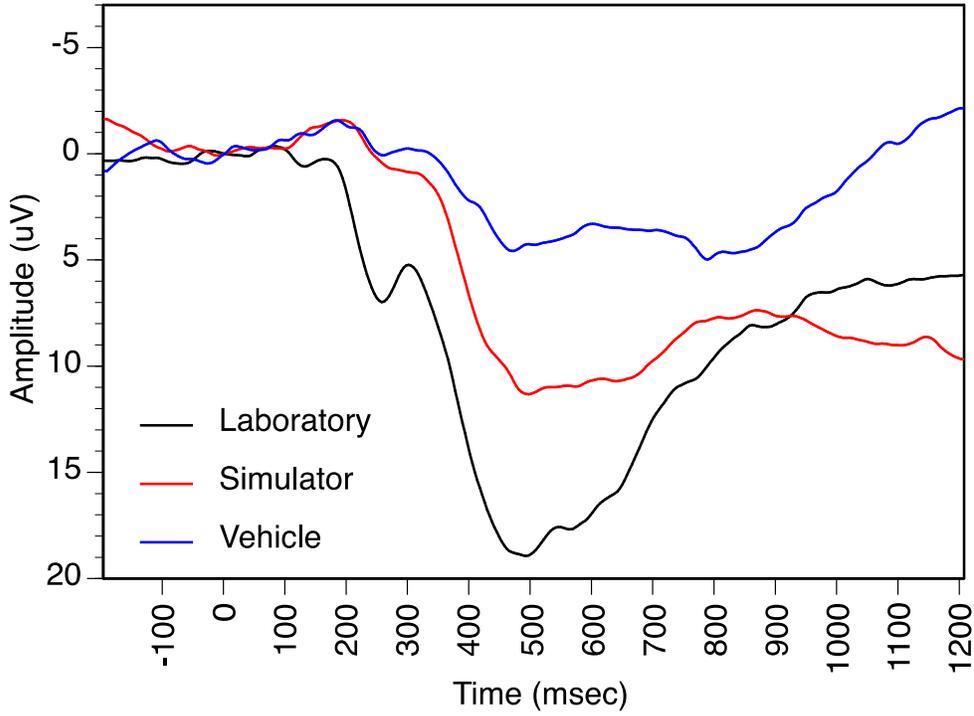
Figure 15. Passenger Conversation ERPs at Pz across Experiments 1-3.
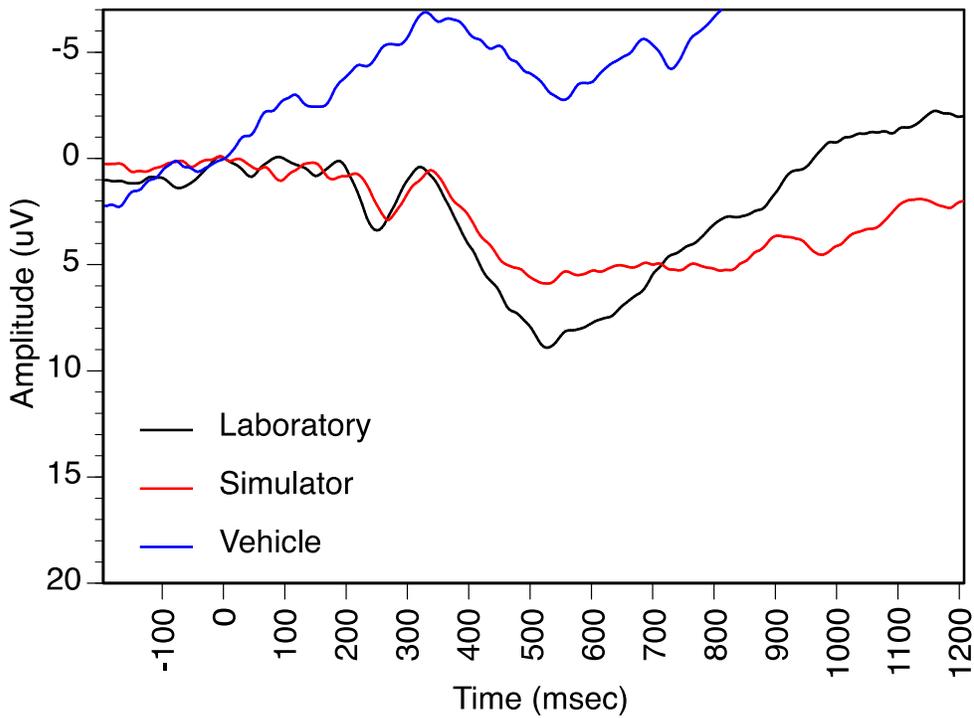
Figure 16. Hand-held Cell Phone ERPs at Pz across Experiments 1-3.
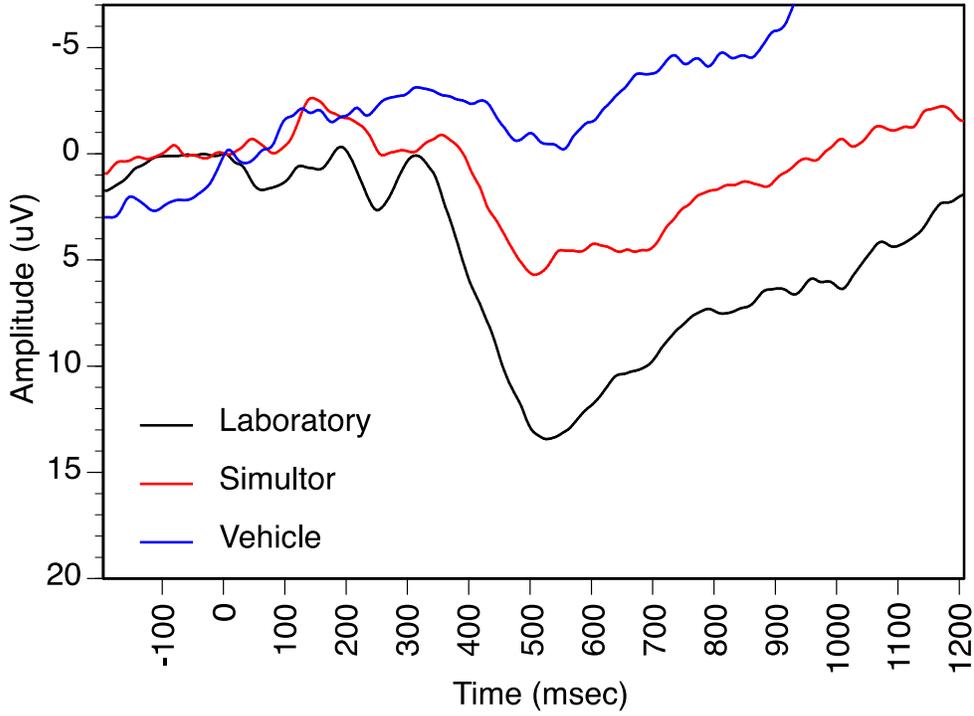
Figure 17. Hands-free Cell Phone ERPs at Pz across Experiments 1-3.
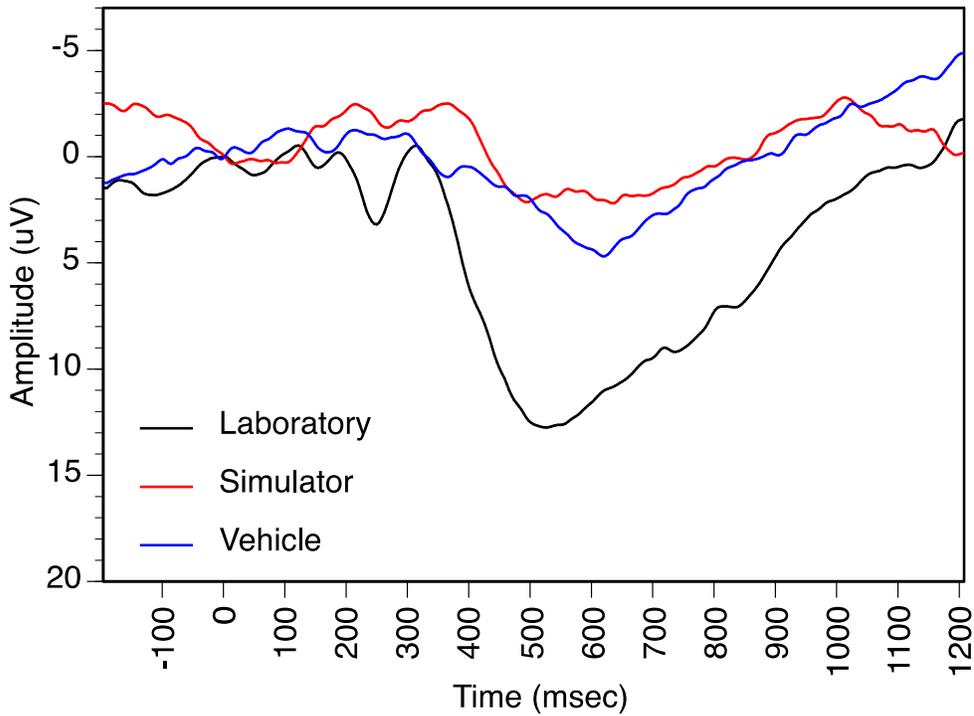
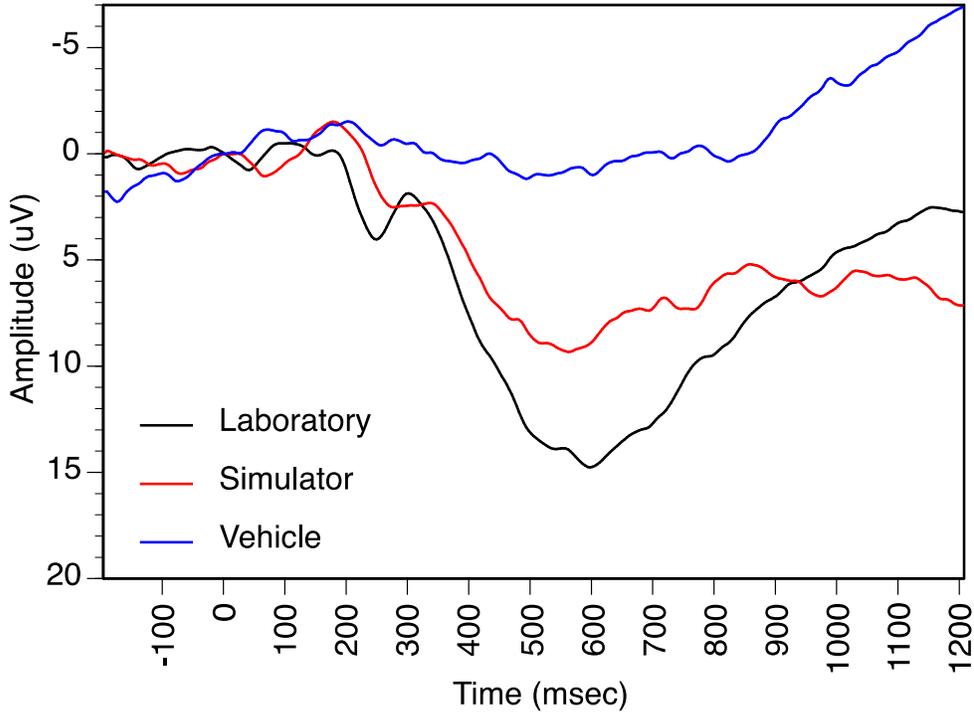**Figure 18. Speech-to-Text ERPs at Pz across Experiments 1-3.**
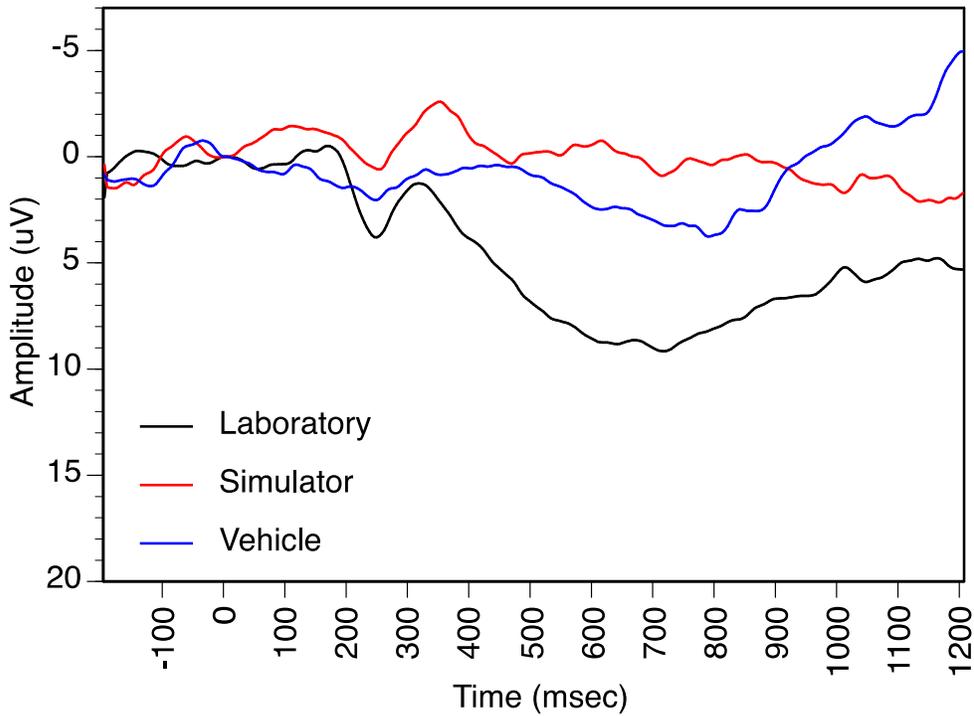

Figure 19. OSPAN ERPs at Pz across Experiments 1-3.

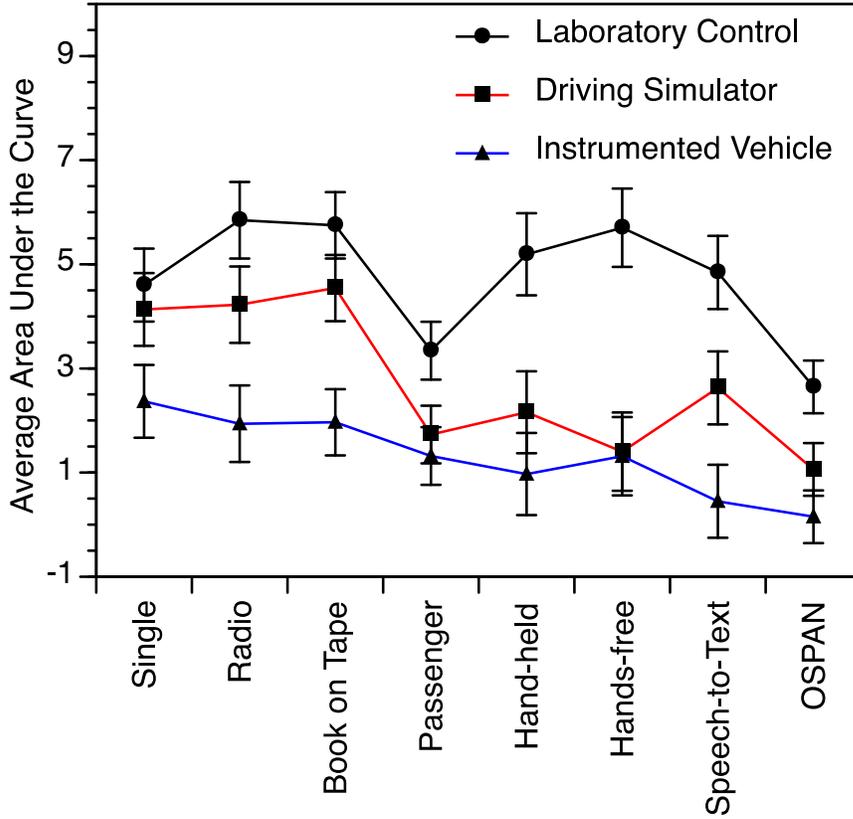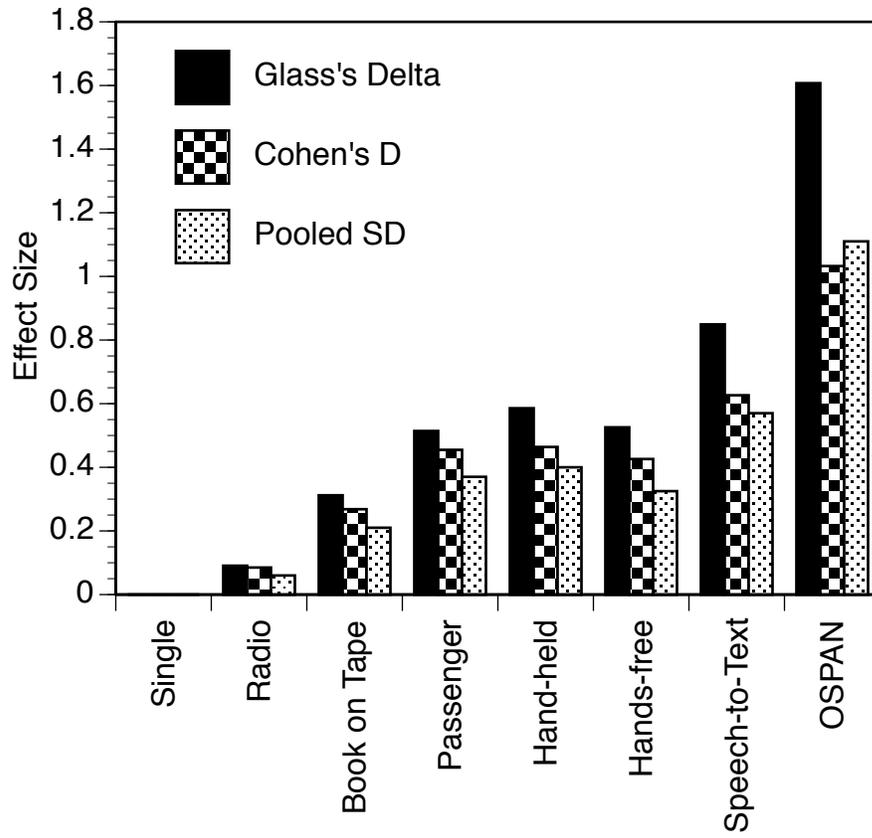**Figure 21. P300 Amplitude across Experiments 1-3.**

Figure 27. Effect Size Measures across All Experiments.

Appendix C: A Brief Overview of the ERP Methodology

The Event-Related Brain Potential (ERP) is composed of time-locked segments of Electroencephalographic (EEG) activity that have been averaged together. In our study, every time that a light from the DRT device was presented, the EEG recordings were marked for later signal processing. The average of the time-locked segments of EEG, the ERP, provides a metric of the brain activity associated with processing the light from the DRT task.[6]

The ERP represents a change in voltage (in microvolts) over time (in milliseconds). The ERP differs in morphology at the different electrode sites on the scalp and is made up of several "components" that reflect different mental activities performed by the participant. The ERP waveforms for each participant are averaged together to create a grand average ERP. Figure 1A presents the grand average ERP recorded at the midline parietal electrode site (Pz) in response to the onset of green light from the DRT device in the single-task condition of Experiment 1 (by convention, positive voltages are reflected by a downward deflection in the graph). The onset of the green light in the DRT task is at 0 msec.

The components of the ERP are labeled in relation to their peak latency and amplitude. In Figure 1A, the ERP at this electrode site is relatively flat for the first 175 msec and then a positive inflection occurs, peaking approximately 200 msec after stimulus onset (this is the "P2" component of the ERP). Following the P2, there is a small negative inflection peaking around 300 msec (this is the "N2" component of the ERP). Finally, there is a large positivity peaking around 475 msec (this is the "P300" component of the ERP). This P2-N2-P300 ERP complex reflects a snapshot of the brain's electrical activity associated with processing of the green light from the DRT device.

Our report focused on the P300 component of the ERP because of its sensitivity to cognitive workload (Kramer, Sirevaag, & Braun, 1987; Sirevaag et al., 1993). The P300 is maximal at the midline parietal electrode site (Pz) and its peak *latency* provides a measure of the mental timing and has been used in studies of mental chronometry. For example, a classic study by McCarthy & Donchin (1981) reported that increasing the perceptual/cognitive difficulty of a task increased the peak latency of the P300. In our study, we interpret longer peak latencies as reflecting a slowing of the perceptual and decision-making processes associated with the stimulus evaluation. Note that P300 latency is measured as the time from stimulus-onset to peak.

The peak *amplitude* of the P300 provides a measure of the attention allocated to a task (Sirevaag, et al., 1989; Wickens, et al., 1983). For example, in another classic

---

[6] The EEG comprises both the "signal" associated with the mental processing of the green light and the "noise" that is unrelated to the processing of the green light. With sufficient trials in the average, the noise in the EEG cancels and the resulting ERP reflects the time-locked electrical activity of the brain associated with processing of the green light (i.e., the noise decreases as a $\sqrt{N}$, where N is the number of trials that comprise the average). For each subject there were approximately 30 trials that went into the ERP condition averages.

ERP study, Sirevaag et al. (1989) varied the processing priority of two concurrent tasks and found a reciprocal relationship in the amplitude of the P300 associated with the two tasks. As participants allocated more attention to one of the tasks, the P300 amplitude associated with its processing increased and the amplitude associated with processing the other concurrent task decreased. In our study, we interpret a reduction in P300 amplitude as evidence that the concurrent tasks are placing increased demands on limited capacity attention. Note that P300 amplitude was quantified by computing the average area under the curve between 400 and 700 msec.

In our report, we use the P300 component of the ERP as a measure of the attentional demands of secondary in-vehicle activities. Given the capacity limitations of human attention (Kahneman, 1973), as the cognitive workload of an in-vehicle activity increases, the remaining attention that can be allocated to the processing of the DRT signals decreases, and this should be evident as both a lengthening of the P300 latency and a decrease in the P300 amplitude. As mentioned in the body of this report, a similar logic has been used in aviation psychology to determine the workload of pilots performing different activities (Kramer, Sirevaag, & Braun, 1987; Sirevaag et al., 1993).

Figure 1A. The grand average ERP recorded at the midline parietal electrode site (Pz) in the single-task condition of Experiment 1.
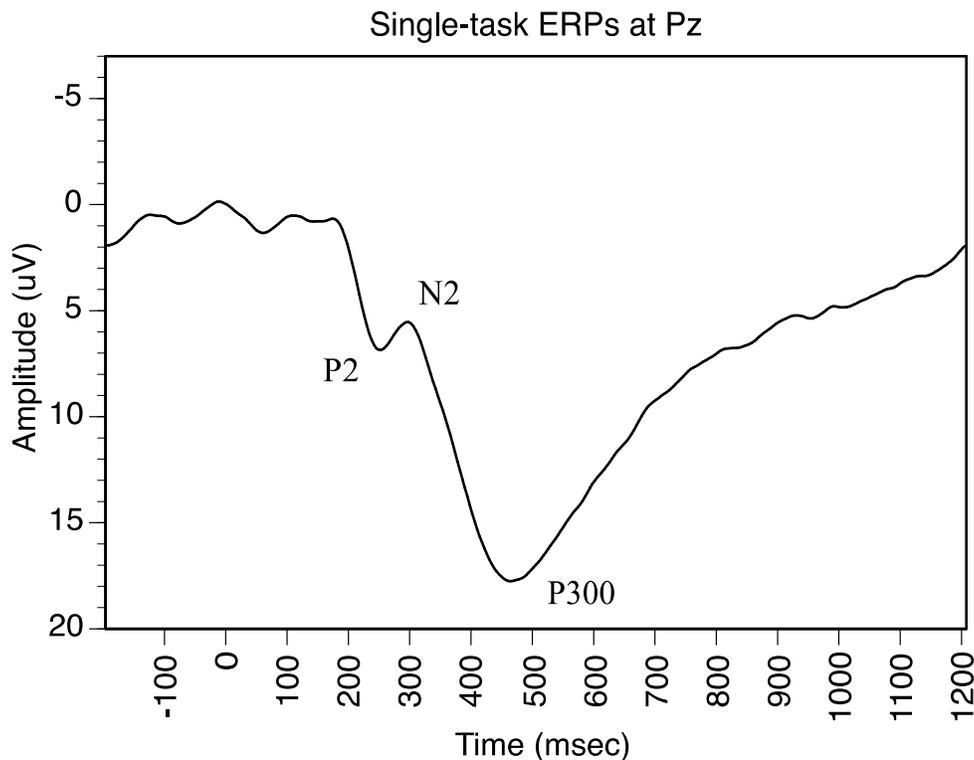
Appendix D: Route Description for Experiment 3

- 2nd and 3rd Avenues in the Avenues of Salt Lake City, UT between T and D Streets.

- The total distance of the route is 2.75 miles, with each side of the route being 1.3 miles. The route is a suburban driving environment.

- 3rd Avenue consists of two-way traffic with a bike lane on each side and street parking off to either side. There are five stop signs and one stoplight. Four of the five stop signs are all-way controlled stops, meaning that traffic coming from all directions must stop.

- 2nd Avenue consists of two-way traffic with street parking off to either side. There are five stop signs and one stoplight. All five stop signs are all-way controlled stops.

- Due to the wide nature of the streets, visibility is clear at all intersections.

- The average time to complete one loop is nine minutes.

- Stop Locations
  - o Stop Signs
    - 3rd Ave. and T, N, K, I, and D Streets (T St. was not all-way controlled)
    - 2nd Ave. and D, I, L, N, and R Streets
  - o Stop Lights
    - 3rd Ave. and E Street
    - 2nd Ave. and E Street