U.S. Department
of Transportation

**National Highway
Traffic Safety
Administration**

DOT HS 810 867 March 2007

Characteristics of Voice-Based Interfaces for In-Vehicle Systems and Their Effects on Driving Performance

DISCLAIMER

NOTE

REGARDING COMPLIANCE WITH

AMERICANS WITH DISABILITIES ACT SECTION 508

For the convenience of visually impaired readers of this report using text-to-speech software, additional descriptive text has been provided within the body of the report for graphical images to satisfy Section 508 of the Americans with Disabilities Act (ADA).

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT HS		

4. Title and Subtitle	5. Report Date
Characteristics of Voice-Based Interfaces for In-Vehicle Systems and Their Effects on Driving Performance	March 2007
	6. Performing Organization Code NHTSA/NVS-312
7. Author(s) Thomas A. Ranney, Ph.D., Transportation Research Center Inc., Elizabeth N. Mazzae, MSE, National Highway Traffic Safety Administration, G. H. Scott Baldwin, MSE, Transportation Research Center Inc., and M. Kamel Salaani, Ph.D., Transportation Research Center Inc.	8. Performing Organization Report No.
9. Performing Organization Name and Address National Highway Traffic Safety Administration Vehicle Research and Test Center P.O. Box 37 East Liberty, OH 43319	10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address National Highway Traffic Safety Administration 400 Seventh Street, S.W. Washington, D.C. 20590	13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code

15. Supplementary Notes

The authors wish to thank Mike Perel, Frank Barickman, Heath Albrecht, Adam Andrella, Ed Parmer, and Jodi Clark for their contributions to this work.

A test track experiment was conducted to measure the disruptive effects on driving of complex secondary tasks and to evaluate voice interface design characteristics that may help reduce distraction. Thirty-six drivers in three age groups (18-25, 30-45, 50-60) drove an instrumented vehicle while performing a combination of car following, peripheral target detection, and secondary tasks on a closed test track. Secondary tasks included a simulated navigation task, in which drivers were given specific questions to answer using information obtained from a traveler information system, and a simulated phone task, in which drivers performed the Baddeley short-term memory task. Four variations of the navigation task were used, including combinations of two factors: (1) mode of information acquisition (auditory vs. auditory + visual map), and (2) system reliability (no recognition errors vs. 20% errors). All secondary tasks were implemented with a hands-free voice interface. The results indicated significant degradation of driving performance associated with all secondary tasks. Specifically, drivers exhibited higher levels of steering entropy, which measures the number and magnitude of steering corrections relative to a baseline drive, and slower car-following responses. They also had higher levels of target detection errors, slower target detection response times, and higher levels of subjective workload while performing secondary tasks relative to a baseline condition with no secondary task. Secondary tasks that required drivers to look at a map while interacting with the traveler information system were consistently more disruptive than similar tasks that did not require the map for all performance measure categories. Decreased system reliability did not disrupt driving performance. The study results provide the following implications for voice interface design: (1) Unnecessary or redundant visual displays should be avoided; (2) Drivers appear to be able to tolerate a voice interface with less than perfect reliability. Additional research is recommended to explore the effects of higher system error rates and task pacing (self-paced vs. externally-paced) on the interference induced by secondary task performance.

17. Key Words	18. Distribution Statement
voice interface, driving performance, distraction, cognitive distraction, traveler information system, In-vehicle information systems, IVIS, hands-free phone conversation	Document is available to the public from the National Technical Information Service Springfield, VA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	110	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

Automakers and telematics developers are using voice interfaces to help solve potential distraction problems, thus allowing drivers to perform a variety of secondary tasks while driving. However, the question remains as to whether voice interfaces, which eliminate the need to look at displays and to manipulate controls, can help enough to allow safe operation of in-vehicle technologies or whether the remaining cognitive distraction will significantly degrade driving performance. This question was addressed in a previous test track study conducted at the National Highway Traffic Safety Administration's (NHTSA) Vehicle Research and Test Center (VRTC). The study compared driving performance decrements associated with secondary tasks that could be performed using both visual/manual and voice interfaces. The primary (driving) task combined car following and visual target detection. Secondary tasks involved a sequence of interactions with an in-vehicle computer and were representative of current in-vehicle technologies.

Performing the secondary tasks resulted in significant decrements to vehicle control, target detection, and car-following performance. The voice-based interface helped reduce the distracting effects of the secondary tasks. Modest improvements were observed for measures of vehicle control and target detection. However, the voice interface did not influence car-following performance, and because the car-following task was more cognitively demanding than the vehicle-control and target-detection tasks, it was concluded that the voice-based interface did not appreciably reduce the cognitive distraction associated with the secondary tasks. The present study looks more closely at the cognitive distraction associated with secondary tasks performed with voice interfaces.

Traveler information systems, accessible via telephone by dialing '511,' have been developed and implemented in parts of the United States since 2000. As of December 2006, 28 states had implemented 511 systems. Projections indicate that 511 service will be operational throughout the United States by 2010. Voice commands are used to navigate hierarchical menu structures. Users obtain current information about traffic conditions, including accident and construction delays, road conditions, and public transit. Some systems also include travel services. System users include commuters and travelers unfamiliar with an area. A significant percentage of 511-system use is expected to involve drivers calling from wireless phones. It is therefore of interest to determine the potential for distraction among users of these systems.

In the present study, thirty-six drivers in three age groups (18-25, 30-45, 50-60) drove an instrumented vehicle while performing a combination of car following, peripheral target detection, and secondary tasks of varying complexity on a closed test track with some traffic present. The first objective was to determine whether secondary tasks, all performed using a hands-free voice interface, interfered with driving performance. We also sought to determine how secondary task complexity was related to driving performance degradation. For this purpose we used a simulated phone conversation task, in which drivers listened to sentences, made judgments about them, and recalled targeted words. We also used a navigation task, in which drivers used information obtained from a simulated traveler information (511) system to answer specific questions. Pre-recorded messages containing traffic information about a hypothetical network of roads and cities were organized in a hierarchical menu structure that was accessed by voice commands.

The second objective was to evaluate the effects of two specific voice interface attributes, including: (1) whether or not the interface had a visual component (map); and (2) voice interface reliability. Four navigation task variations were used for this purpose, including combinations of two factors: (1) mode of information acquisition (auditory vs. auditory + visual map), and (2) system reliability (no recognition errors vs. 20% errors). To simulate different levels of system reliability, we developed the capability of introducing voice recognition errors into the simulated 511 system.

In addition to secondary task, the other experimental factors were driver age group and lead vehicle speed input signal. Participants' performance on car following, peripheral target detection, and other vehicle control measures while performing secondary tasks was compared with their performance while driving without secondary tasks. Target detection was implemented via the Peripheral Detection Task (PDT); drivers responded to an intermittent stream of simple visual targets reflected on the windshield of the instrumented vehicle.

The analyses addressed specific hypotheses. Our first hypothesis was that secondary task performance would be associated with degraded vehicle control, target detection, and car-following performance. The results indicated that secondary task performance was associated with significant degradation for all categories of driving performance. Specifically, drivers exhibited higher levels of steering entropy, which measures the number and magnitude of steering corrections relative to a baseline drive, and longer car-following delay, which is a measure of response speed. They also had higher levels of target detection errors, slower target detection response times, and higher levels of subjective workload while performing secondary tasks relative to a baseline condition with no secondary task.

Our second hypothesis was that the 511 tasks were more complex and thus potentially more distracting than the simulated phone task. We predicted that performance degradation associated with the 511 tasks would be greater than with the simulated phone task. The present results were only partially consistent with this prediction. In particular, we observed consistently lower car-following coherence for all categories of 511 tasks. Coherence reflects the degree to which a following driver is able to match the lead vehicle speed input signal. The lower coherence values indicate that the 511 tasks interfered more with car-following accuracy than did the simulated phone task. Drivers also drove at longer headways in three of the four 511 task conditions than in the phone task condition. The 511-Map tasks were associated with larger decrements in steering entropy and car-following delay, relative to the simulated phone task. Thus, the 511 tasks that required map use were consistently more disruptive than the simulated phone task, but those that did not require map use were not. Because the simulated phone task was considered to be more demanding than typical phone conversations, the differences observed between these conditions may understate the real-world differences between the disruptive effects of phone conversation and navigation tasks involving the 511 system.

The third hypothesis addressed differences in distraction potential associated with one specific task interface attribute, namely whether or not a map was required to perform the task. We predicted that the interference associated with the map tasks would be greater than for the no-map tasks and that this effect would be observed for both visual and cognitive aspects of driving. We found consistent evidence that the map trials were more disruptive than the no-map trials. Map trials were associated with greater steering entropy, longer car-following delay, lower PDT detection rates, and slower PDT response times. In addition, participants rated the map trials as more demanding than the no-map trials. Thus, while the problem-solving aspects of the map and no-map trials were similar, the additional requirements of looking at a map and integrating information from visual and auditory sources led to significantly worse performance in all performance measure categories.

The fourth hypothesis considered the effects of voice interface reliability on driving performance. We predicted that an increase in the occurrence of voice recognition errors (i.e. decreased system reliability) would be disruptive to drivers' concentration, leading to poorer driving performance. We found no evidence to support this prediction, suggesting that the 20% error rate used in the study did not disrupt drivers' concentration enough to influence their driving performance. Finally, we hypothesized that the disruption caused by an increase in simulated voice recognition errors would lead to longer secondary task completion times. We found no significant difference between the error and no-error conditions for this measure.

Performance differences among the age groups were evident only for car-following measures. Specifically, older drivers had lower coherence values, indicating more difficulty in car following. The older drivers also had longer delays, indicative of slower responses to lead vehicle speed changes, relative to the other two age groups. Modulus values for older drivers were consistently smaller, indicating more conservative responses to changes in lead vehicle speed. Finally, older drivers drove at longer headways than other drivers. None of the vehicle-control measures and none of the PDT measures revealed differences among the age groups.

In summary, the in-vehicle tasks performed using voice interfaces were associated with significant degradation of driving performance. This was true both for simulated 511 system tasks and for simulated hands-free phone tasks and leads to the conclusion that voice interfaces are not sufficient to eliminate the cognitive distraction associated with secondary tasks like those used in this study. Tasks that required drivers to look at a simulated map display were most disruptive, not only because of the requirement to look away from driving, but also because of increased cognitive demands associated with the requirement to interpret information obtained from the 511 system with the visual map display. The simulated phone task was only slightly less disruptive than the 511 tasks, however because the phone task is considered to be more demanding than typical phone calls, real-world use of 511 systems by drivers is likely to be more distracting than typical phone calls.

All secondary tasks were associated with significant cognitive distraction, which affected not only the cognitive aspects of driving but also visual target detection and vehicle control. Thus while voice interfaces allow drivers to keep their hands on the wheel and eyes on the road, the cognitive distraction associated with queries of a 511 traveler information system or moderately demanding hands-free phone conversation may impose a significant cognitive load that has the potential to degrade all components of driving performance.

The study results support the following implications for voice interface design: (1) Unnecessary or redundant visual displays should be avoided; (2) Drivers appear to be able to tolerate a voice interface with less than perfect reliability; (3) Designers of information acquisition tasks that require navigation of hierarchical menu structures should attempt to minimize distraction by simplifying information presentations as much as possible.

Additional research is recommended: First, we recommend using higher error rates to further explore the effects of system reliability on driving performance; and second, we recommend exploring the effects of task pacing (self-paced vs. externally-paced) on the interference induced by secondary task performance.

1.0 INTRODUCTION

1.1 Background

Automakers and telematics developers are using voice interfaces to help solve potential distraction problems, thus allowing drivers to perform a variety of secondary tasks while driving. However, the question remains as to whether voice interfaces, which eliminate the need to look at displays and to manipulate controls, can help enough to allow safe operation of in-vehicle technologies or whether the remaining cognitive distraction associated with secondary tasks of increasing complexity will significantly degrade driving performance.

Different types of distraction can be described within the context of a hierarchical model of driving behavior, developed by Michon (1985). According to this model, driving involves concurrent activity at three levels: operational or vehicle control; tactical or maneuvering; and strategic. The operational level includes the lowest level component behaviors such as lateral (steering) and longitudinal (speed) control. With experience, behavior at this level becomes automatic and requires very little of the driver's attention. Distraction effects at this level typically involve physical conflicts between vehicle control and secondary task demands. For example, secondary tasks that require manual inputs may be expected to interfere with steering behavior because both tasks require use of the hands. Similarly, secondary tasks that require drivers to look inside the vehicle may be expected to interfere with vehicle control if drivers are thus deprived of the visual input required to monitor vehicle position. Interference at this level has been referred to as peripheral (Strayer & Johnson, 2001) to underscore the absence of cognitive or attentional interference.

Driving behavior at the tactical level involves responses to common driving situations, such as deciding whether or not to stop for a changing traffic signal, merging, or passing. Drivers typically make complex speed/distance judgments and decisions about when to modulate steering or speed behavior in response to the expected actions of other vehicles. As such, behavior at this level involves considerably more cognitive activity than behavior at the operational level. Thus, secondary tasks that divert attentional or cognitive resources from driving are likely to degrade performance at the tactical level of control. For example, cognitively distracted drivers may be unable to devote sufficient attentional resources to assess the dynamics of the immediate traffic situation. This may lead them to choose inadequate gaps or safety margins for passing or entering the traffic stream. Finally, the strategic level is the highest level of Michon's hierarchy. Strategic components of driving behavior include such travel decisions as what time of day to make a trip and what route to take. Most decisions at this level are made before a trip begins. One exception is when a driver encounters a problem on a planned route and is required to revise the route while the trip is underway.

Based on this model, cognitive distraction will most likely be observed as degradation of driving behavior at the tactical level and occasionally at the strategic level. To assess the cognitive distraction associated with in-vehicle technologies, we sought to develop a methodology that included tactical driving situations. We decided that incorporating strategic elements into a test track experiment was infeasible due to our inability to alter the route or time associated with the experimental drive. We identified car following as a tactical situation that could readily be simulated on a closed test track. We chose the coherence car-following paradigm developed by Brookhuis and colleagues (Brookhuis, Waard, & Mulder, 1994), in which drivers maintain a constant headway behind a lead vehicle that is traveling according to a varying speed function that can be decomposed into one or more sine waves. The varying speed function is necessary to allow computation of car-following coherence, and the associated measures of phase shift and modulus. Coherence is a measure of squared correlation, which reflects the degree to which the following driver is able to match the lead vehicle speed signal. Brookhuis et al, (1994) demonstrated that wireless phone use while driving increased the phase shift of the two speed signals, reflecting an increase the car-following response time, which they referred to as delay. We therefore selected coherence and the associated measures, particularly delay, as measures that would likely be sensitive to cognitive distraction.

Implementing car following on a closed test track differs from real-world driving in that it removes some of the uncertainty associated with unexpected events in the real world. Reflecting the fundamental importance of visual target detection to driving (Rumar, 1990), we decided to add a visual target detection task to our data collection protocol. However, we wanted to use a relatively simple target-detection task to avoid imposing a significant processing load that might interfere with primary (car following) or secondary (in-vehicle technology) task performance. We chose the Peripheral Detection Task (PDT) for two reasons: First, the PDT is a simple visual task in which drivers respond to a sequence of simple targets that require no interpretation and thus do not impose a significant processing load; and second, the PDT provides a tool for measuring the interference in driving-relevant visual processing associated with the secondary tasks (Harms & Patten, 2003). For these reasons, we considered the PDT to represent a low-level operational component of the primary (driving) task.

Despite the minimal processing demands, we expected PDT performance to deteriorate with increasing secondary tasks demands, even for secondary tasks performed entirely with voice interfaces. Strayer and Johnston (2001) demonstrated that cognitively engaging secondary tasks have the potential to degrade target detection performance. In a laboratory study, they had drivers perform a combination of tracking and visual target detection while also performing simulated phone conversations. Their target detection task was designed to simulate traffic signals: Drivers had to respond to red lights and to ignore green lights. They found that when their participants were engaged in simulated phone conversations, they were more likely to miss designated targets and were slower to respond to those that were detected. This was true even though the phone conversation task did not require drivers to look away from the roadway. They interpreted their results to suggest that phone conversation disrupts visual target detection performance by diverting attention away from driving to the phone conversation. The PDT has also been shown to be sensitive to both driving workload and distraction from In-Vehicle Information Systems (IVIS), including both visual and cognitive components (Jahn, Oehme, Krems, & Gelau, 2005)

The car-following plus target-detection paradigm was used in a previous test track study in which drivers used both visual/manual and voice interfaces to perform secondary tasks that involved a sequence of interactions with an in-vehicle computer (Ranney et al., 2004a). We found that performing secondary tasks resulted in significant decrements to vehicle control, target-detection, and car-following performance, reflecting both operational and tactical-level interference (Ranney et al., 2004a). The voice-based interface helped reduce the distracting effects of the secondary tasks. Modest improvements were observed for measures of vehicle control and target detection, suggesting that the voice interface helped reduce the peripheral (visual and manual) impairment. However, the voice interface did not influence car-following performance, which suggests that it did not appreciably reduce the attentional (cognitive) distraction associated with the secondary tasks. We concluded that the methodology was sensitive to interference associated with distraction at both the operational and tactical levels of control. The present study utilized the same data collection protocol to explore the interference associated with a set of secondary tasks that varied in complexity and were performed with a hands-free voice interface.

Obtaining information from traveler information systems is one example of a complex task that may be performed by drivers in moving vehicles. Traveler information systems are accessible via telephone by dialing '511.' They provide current information about traffic conditions, including accidents and construction delays, road conditions, and information about public transit. Some systems also include travel services. 511 systems require voice-activated navigation of hierarchical menu structures and may be used by commuters and/or by travelers unfamiliar with an area. They have been developed and implemented in parts of the United States since 2000. As of December 2006, 28 states have implemented 511 services (Federal Highway Administration, 2007). Projections indicate that 511 service will be operational throughout the United States by 2010.

A significant percentage of 511 system use is expected to involve drivers calling from wireless phones, increasingly using voice interfaces. It is therefore of interest to determine the potential for distraction among users of these systems. In particular, there are several interface issues that have implications for driver distraction. For example, 511 deployment guidelines allow messages to consist either of synthesized speech, digitized speech, or recorded messages. Some existing systems include combinations of different types of messages. Differences in intelligibility of different types of speech, particularly in the context of driving background noises, are unknown. The complexity of 511 system menu structures also differs among various implementations. Densely populated areas with extensive freeway systems are likely to have more complex menu structures than less densely populated areas. Jacko and Salvendy (1996) argued that menu depth is related directly to task complexity because "increased depth involves additional visual search, decision-making, response selection, and greater uncertainty as to the location of the target" (p. 1195). The density of the area covered by each system node is a key determinant of the expected number of alerts due to traffic crashes or construction delays, which directly impacts the message length. Task complexity can also be defined in terms of the number of reporting areas that must be considered to address a specific travel question.

System reliability is an attribute of voice interface systems that can be expected to influence the ease of use and potential for distraction. In this context, system reliability refers to the performance of the Automatic Speech Recognition (ASR) system, as well as the communication signal strength. Casali, Williges and Dryden (1990) found significant performance differences between simulated speech recognition systems that varied recognition system accuracy between 91% and 99%. Specifically, they found that the accuracy level of the speech recognizer influenced task completion time and users' acceptability ratings, but not the number of errors left uncorrected. Their simulated task required each participant to make approximately 225 voice entries per trial. This resulted in differences of 10-20 errors per trial between accuracy conditions.

1.2 Objectives

The study objectives were based on the voice interface design characteristics described above. The first objective was to demonstrate that the secondary tasks, all performed using a hands-free voice interface, interfered with driving performance. We also sought to determine how secondary task complexity was related to driving performance degradation. For this purpose we used a simulated phone conversation task, in which drivers listened to sentences, made judgments about them, and recalled targeted words. We also used a navigation task, in which drivers used information obtained from a simulated traveler information (511) system to answered specific questions. Pre-recorded messages containing traffic information about a hypothetical network of roads and cities were organized in a hierarchical menu structure that was accessed by voice commands.

The second objective was to evaluate the effects of two specific voice interface attributes, including: (1) whether or not the interface had a visual component (map); and (2) voice interface reliability. Four navigation task variations were used for this purpose, including combinations of two factors: (1) mode of information acquisition (auditory vs. auditory + visual map), and (2) system reliability (no recognition errors vs. 20% errors). To simulate different levels of system reliability, we developed the capability of introducing voice recognition errors at a predefined rate into the simulated 511 system. In addition to secondary task, the other experimental factors were driver age group and lead vehicle speed signal.

Several specific hypotheses were tested in this experiment. First, the cognitive demands of the secondary tasks used in the present study were expected to be at least comparable to those used in the previous study (Ranney et al., 2004a), which required participants to navigate through automated phone systems. We therefore predicted that the performance of secondary tasks would be associated with degraded target detection, and car-following performance due to the diversion of both attentional resources away from the driving task. However unlike the secondary tasks used previously, none of our current tasks required significant physical manipulation of an interface. We therefore predicted that the absence of physical conflicts between driving and secondary task would lead to weaker degradation of vehicle control measures, relative to that predicted for car-following measures. Based on the results of Strayer and Johnston (2001), we predicted that the attentional demands of the secondary tasks would lead to degraded target-detection performance.

The 511 tasks used in this study were fundamentally different from the simulated phone task on two dimensions, pacing and cognitive demand. The 511 tasks were self-paced and thus allowed the driver to progress in accordance with the demands of the immediate driving situation. In contrast, the phone task, in which participants performed the Baddeley task (Baddeley, Logie, Nimmo-Smith, & Brereton, 1985), was externally paced; participants were required to listen and respond according to a schedule that they did not control. The second difference between the two tasks concerned the type of decision-making required and the associated cognitive demand. The simulated phone task required participants to make simple decisions about the meaning of sentences and to recall specified words. In contrast, the 511 tasks required participants to recall the structure of a command hierarchy, decide what type of information was required to answer specific questions, navigate the menu structure of the command hierarchy to obtain the required information, and interpret the information to answer the question posed.

While a forced-pace regimen like that used in the simulated phone task provides less flexibility for drivers to respond to unanticipated events, the problem solving and active negotiation of a hierarchical menu structure were considered more cognitively demanding than the simulated phone task. We therefore expected that the 511 tasks would be associated with more primary task interference than the simulated phone task, and that this difference would be most apparent on car-following performance, which as described above, involved more cognitive activity than the other categories of performance measures. This was our second hypothesis.

Our third and fourth hypotheses concerned differences associated with two specific attributes of the 511 tasks. Specifically, because of the increased visual demands associated with tasks requiring the use of a map, we predicted that these tasks would be associated with the highest levels of degraded performance. We expected this to be particularly evident for measures of target detection due to the direct conflict between target detection and use of the map. Similarly, we predicted that (simulated) increases in 511 system errors would increase the level of distraction associated with the 511 task, leading to further degradation of vehicle control, target detection and car-following performance. Finally, based on findings of Casali, et al., (1990), we expected secondary task completion times to increase with decreases in simulated speech recognition accuracy. This was our fifth specific hypothesis.

1.3 Approach

To address these issues, we conducted a closed-course experiment. We used the test protocol and measurement tools developed previously (Ranney et al., 2004a), which include a car-following paradigm, a Peripheral Detection Task (PDT), and an eye-tracker to measure eye glance behavior. Participants performed secondary tasks while performing the car-following and target-detection tasks on a closed test track. Secondary tasks included several navigation problem-solving tasks that required drivers to access and obtain information from a simulated 511 system, and the Baddeley task, which simulated a moderately demanding phone conversation. Both secondary tasks were accessed via a hands-free wireless phone with programmed speed dial functions. The experiment sought to quantify the distraction potential, defined as the extent to which driving performance is degraded by concurrent performance of the secondary task. The main independent variables were driver age group, the secondary task, and two attributes of the 511 system, including ASR system reliability (i.e., system error rate) and mode of information acquisition (auditory vs. auditory + visual map). We also varied the lead-vehicle speed signal in the car-following task. Participants' performance on car following, peripheral target detection, and other vehicle control parameters while performing secondary tasks were compared with their performance without secondary tasks.

2.0 METHOD

2.1 Test Track

The experiment was conducted on the Transportation Research Center's (TRC) 7.5-mile oval test track, located in East Liberty, Ohio. The track consists of three 12-foot wide concrete lanes plus a fourth inner blacktop lane. Two straight segments, each approximately 2.0 miles long are separated by curved and banked segments, which are approximately 1.75 miles in length. Other traffic, including a mix of passenger vehicles and trucks, was present during data collection. The two experimental vehicles used the rightmost concrete lane. Occasionally, slower moving vehicles necessitated a lane change into the middle concrete lane, however when this occurred, the lane change was completed before the secondary task was begun. Similarly, stopped traffic was occasionally present in the inner blacktop lane. This created a temporary visual distraction, but did not otherwise interfere with the trial. None of the trials was disrupted by slower or stopped traffic in the data collection lane. Data collection was suspended during inclement weather (e.g., when windshield wipers were required) and otherwise at the discretion of the experimenter, who monitored the speed and proximity of other traffic on the test track.

2.2 Experimental Design

The experiment used a 6 x 2 x 3 mixed design, in which secondary task (6) and speed input signal (2) were varied within participants, and participant age (3) was varied between participants. Secondary task conditions included a baseline condition, in which no secondary task was performed, a simulated phone task, in which participants performed the Baddeley task (Baddeley et al., 1985), and four 511 task conditions, in which participants used a simulated 511 (traveler information) system to solve hypothetical route-selection problems posed by the experimenter. The four 511 conditions comprised a 2 x 2 embedded design, which included all combinations of two additional factors, referred to as Map/No Map and Errors/No Errors. In the Map condition, the participants needed information presented on a map to complete the task. In the No Map condition, the information available from the 511 system was sufficient to answer the questions. In the Error conditions, the simulated 511 task was programmed to respond incorrectly on 20% of the replies to participants' commands. The No Error condition had no pre-programmed errors. Acronyms for these four conditions are shown below.

	No Error	Error
Map	511 MO	511 ME
No Map	511 OO	511 OE

Thus, the 511M*x* conditions combined auditory plus visual information processing with voice inputs, while the 511O*x* conditions combined auditory information processing with voice inputs. The Error/No Error manipulation simulated differences in interface reliability and was based on the work of Casali, et al., (1990). Two levels of speed input signal were included to vary the difficulty of the car following task. These included a simple sine wave and a complex signal composed of white noise banded over a wider frequency range.

2.3 Participants

Participants were 36 members of the general public, balanced amongst three age groups: 18-25, 30-45, and 50-60. Participants were recruited mainly through newspaper advertisements, which sought participants for a “driving research study.” Participants were required to be in good health, to have a valid driver's license, and to have at least 2 years of driving experience with 10,000 miles driven yearly. They were also required to have at least 6 months of experience using a wireless phone while driving and be comfortable interpreting a map while driving. To optimize eye tracking, drivers who required sunglasses for driving were excluded during phone screening.

2.4 Apparatus

Two vehicles, including a lead vehicle (LV) and a subject vehicle (SV) were used to implement the car-following paradigm. Both vehicles were equipped with automatic transmissions, Micro Data Acquisition System (MicroDAS) (Barickman & Goodman, 1999) and GPS receivers. GPS position readings were used to determine lane position and to derive vehicle speed. A Vorad radar device on each vehicle measured range (inter-vehicle spacing) and range rate to the other vehicle.

The SV had a secondary brake for emergency activation by the experimenter accompanying the subject. The SV also had event switches, which produce stepped voltages to mark the beginning and end of secondary task performance and test track straight segment boundaries in the continuous data stream. Data collection was started independently at the beginning of the trials by an experimenter in each vehicle. The two platforms collected data independently at a 30 Hz sampling rate.

2.4.1 Subject Vehicle

The SV MicroDAS was configured to collect vehicle speed, range, range-rate, lateral position, hand-wheel position, GPS timing signals, and subject responses to the PDT. In addition, we used GPS position information to provide better resolution for lane position measurement in portions of the test track with inadequate edge lines (e.g., test track entrance and exit). The primary SV data collection channels are displayed in Table 1.

Video cameras were used to collect data both inside and outside of the SV. Three interior cameras were used to record subject’s face, hands-on-wheel positions, and interactions with the in-vehicle technology. A camera mounted on the inside of the windshield recorded the forward road scene. These four cameras relayed images to a quad-video multiplexer, which combined the four video signals so they could be recorded on a single video frame for inclusion in the data stream. An additional forward-looking camera was used to record lane position and a rearward-facing camera was used by the experimenter to monitor traffic behind the SV.

The Peripheral Detection Task (PDT) consisted of an array (3 x 20 cm) of 23 LEDs positioned on the dashboard and shielded from direct view of the driver, as illustrated in Figure 1. High-intensity (12,000 mcd) LEDs were used to maximize visibility during bright daylight conditions. LED activation appeared as a reflection in the windshield located at positions with eccentricities ranging between approximately 5-25º to the left of the driver’s line of sight and 2-4º above the dashboard.

A map (5.25 x 9 in, discussed in Section 2.5), which was used for a subset of the 511 tasks, was fixed to the center console area of the vehicle.

Table 1. Subject vehicle data collection channels

Data Channel	Description	Units	Resolution
Vehicle Speed	Ground speed	km/h	1 km/h
Vorad Range	Distance to the LV	m	.5 m
Range-Rate	Relative velocity between the SV and the LV	m/s	.1 m/s
Lateral Position	Lateral position of the SV in reference to the center of the lane delineated by the painted edge markings	cm	2 cm
Lateral Velocity	SV Lateral velocity in reference to the painted edge markings	cm/s	2 cm/s
Road Curvature	Curvature of the upcoming roadway	1/m	6.3e-10 1/m
Offset Confidence	Reliability estimate of the lateral position	%	1 %
Road Curvature Confidence	Reliability estimate of the curvature data	%	1 %
Hand-Wheel Position	Angular position of the steering wheel (0 degrees = straight)	deg	.1 deg
UTC Time	Time of day	HH:MM:SS	1 s
Pulse Per Second	GPS pulse per second signal used to synchronize data from both platforms	0 or 1	+/- 1 μs
Event Task	PDT button press	0 or 1	1/30th s

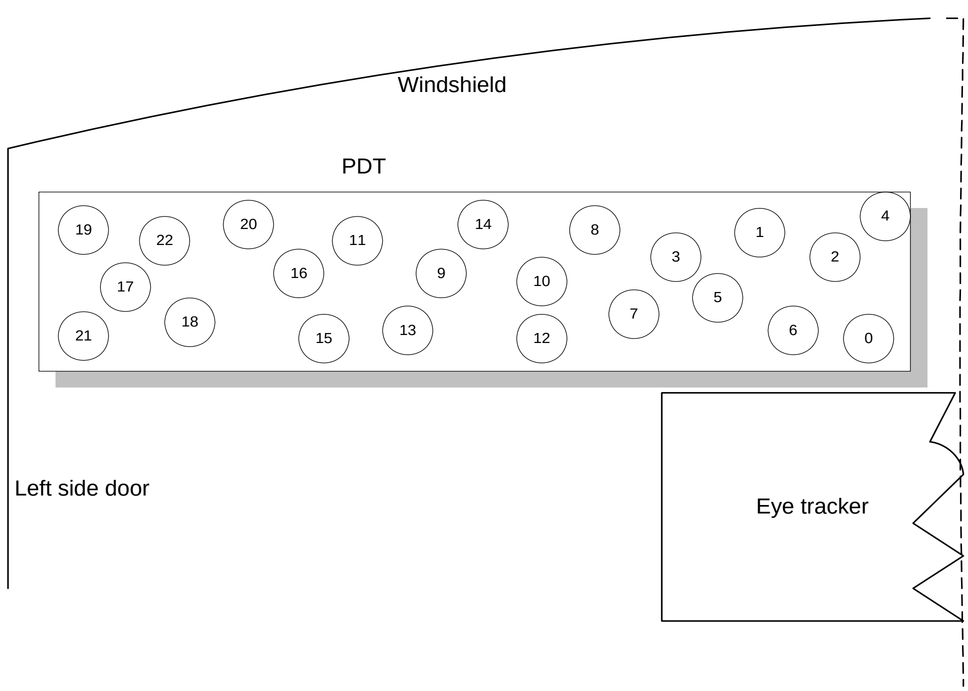


Figure 1. Layout of PDT LEDs

2.4.2 Lead Vehicle

The LV was equipped with a MicroDAS as well as a vehicle speed controller. The LV MicroDAS was configured to collect vehicle speed, tailway (distance to the SV), and GPS timing signals. The primary LV data collection channels are displayed in Table 2.

Table 2. Lead vehicle data collection channels

Data Channel	Description	Units	Resolution
Vehicle Speed	Ground speed	km/h	1 km/h
Range	Distance to SV	m	.5 m
Range-Rate	Relative velocity between the LV and the SV	m/s	.1 m/s
UTC Time	Time of day	HH:MM:SS	1 s
Pulse Per Second	GPS pulse per second signal used to synchronize data from both platforms	0 or 1	+/- 1 μSec

Interfacing a portable computer with a servo controller running a basic proportional integral derivative control loop in software created the LV speed controller. The PC consisted of a 486DX computer with a data acquisition board. The data acquisition board generated analog signals that were sent to the servo controller. Input/output (I/O) lines linked between the computer and the controller were tied into a user interface located near the driver. The user interface was responsible for enabling and disabling the controller and was also used to select the speed input signal. The vehicle speed input, as measured by the LV's transmission speed sensor, provided feedback for the system.

Sending signals from the computer through the D/A converter and sending the resulting analog voltage to the servo controlled LV speed. To make the vehicle accelerate, the servo would rotate clockwise wrapping a cable on a pulley, which was attached to the accelerator pedal. To make the vehicle decelerate, the servo would rotate counterclockwise relieving tension on the accelerator pedal and applying mechanical force to the brake.

To operate the automated speed controller, the experimenter selected one of two files loaded into the computer. One pattern was a trigonometric sine function with frequency of 0.03 Hz and extreme speed values of 50 and 65 mph. The associated acceleration and deceleration requirements were well within limits associated with normal driving (i.e. < .4 G). The second speed pattern was more complex (band filtered white noise) and intended to be less predictable, but had levels of deceleration and acceleration that were less severe than the simple sine wave. Before the speed controller could be engaged, the vehicle had to be traveling at least 55 mph. The experimenter then engaged the controller by pushing a button on the user interface. At this point, the controller modulated LV speed to represent the selected waveform.

2.4.3 Eye Tracking

A Seeing Machines faceLAB eye tracking system was used to record head and eye movements. The system used 2 stereo cameras mounted on the dashboard and was relatively unobtrusive. To assist the system in tracking facial features, participants applied 8 stickers to their faces during system calibration.

2.5 Simulation of a 511 Travel Information System

A "Wizard of Oz" (WOZ) approach was used to simulate the hypothetical traveler information system. This approach used a human-in-the-loop ('Wizard') to replace the speech recognition and automated menu navigation functions of a real 511 system. The system, which was designed to be consistent with 511 System Implementation Guidelines (511 Deployment Coalition, 2003), consisted of a Visual Basic program running on a laptop computer and operated by an experimenter in a remote location accessible via telephone. The program implemented the menu structure shown in Figure 2. It provided a graphical user interface that allowed the experimenter to select pre-recorded audio messages based on the participant's voice commands. Specific information pertaining to traffic and road conditions corresponded to hypothetical roadways shown on a map (Figure 3), which was mounted on the SV console. The map included Interstate highways and numbered segments that were used to answer questions. Each numbered Interstate highway shown on the map was included in the menu structure and messages were created containing information about traffic and road conditions for these roads. Separate audio (.wav) files were recorded for each node of the menu. In response to verbal inputs from the participant, the wizard activated the associated pre-recorded audio messages. These messages were transmitted from the laptop PC over the phone via a Plantronics MX10 headset switcher multimedia amplifier, which allowed the audio output from the PC to be directed through the phone. The control program retained a log of the identification and timing of each message activated.

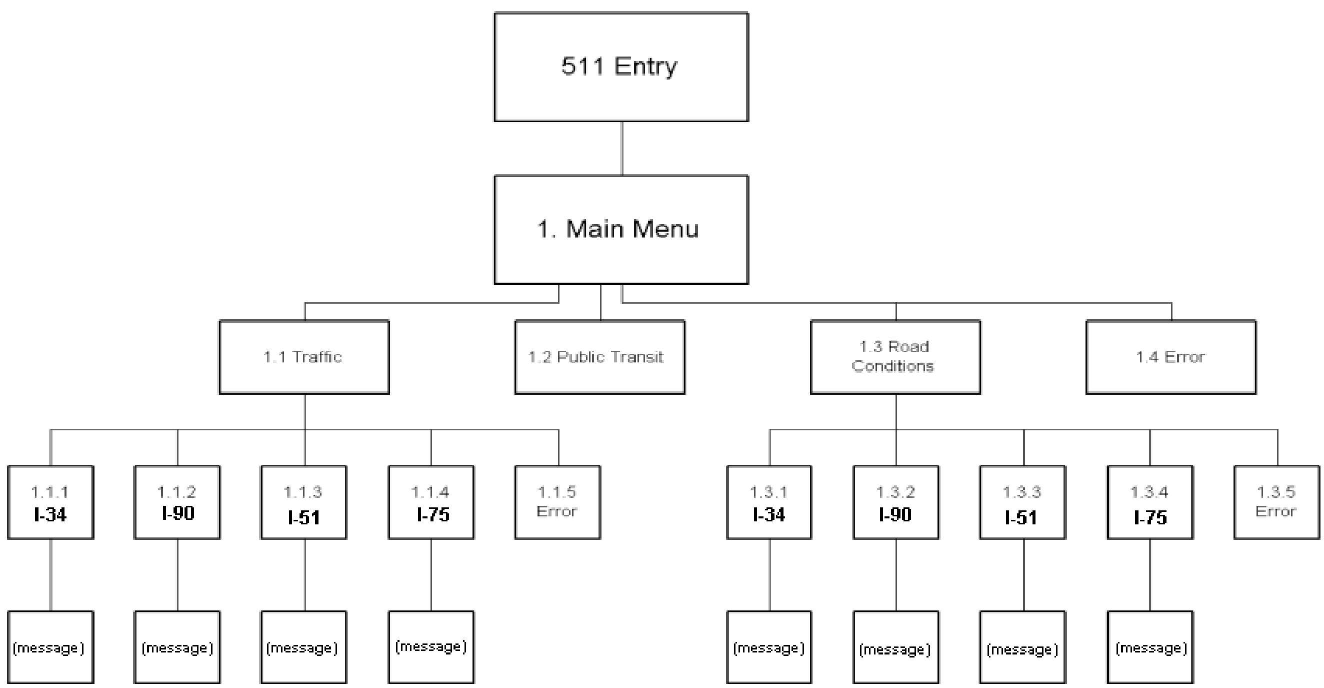


Figure 2. Simulated 511 system menu structure

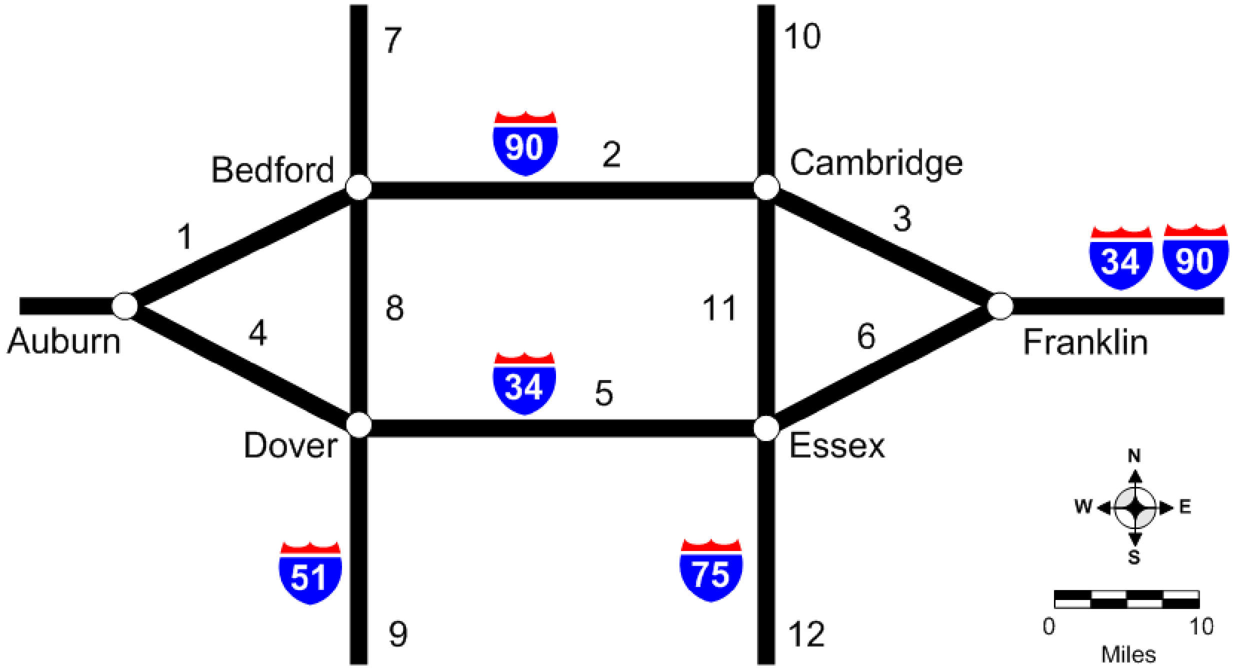


Figure 3. Map used for 511 task

2.5.1 Example of 511 System Interaction Sequence

When instructed to do so, the participant speed-dialed "511," which directed the call to the remote experimenter. The experimenter answered the call and activated the 511 system entry message by clicking on the topmost node of the interface (see Figure 2). The welcome message played at this time, followed by the instructions on the main menu (node 1.0 in Figure 2). Next, the remote experimenter listened to the participant's voice inputs and activated the corresponding pre-recorded audio file for each input. For example, the participant might say "traffic." In response, the remote experimenter activated the message represented by 1.1 in Figure 2, which asks for a route number. The participant would then say the route number of interest. The wizard then activated a message that detailed current conditions and estimated durations of any delays for that specific route.

In some instances, according to a pre-defined rate of failure, when the remote experimenter pressed a button on the interface in response to the participant's verbal request, the 511 system produced an error message stating, "I did not get that. Please (repeat)..." This approach allowed direct manipulation of the reliability of the simulated voice recognition system. All transactions were recorded to permit analysis of the associated timing.

2.6 Driving Tasks

2.6.1 Car-Following

A car-following paradigm modeled after that used by Brookhuis and colleagues (1994) was used. This task required participants to maintain a constant following distance behind a lead vehicle, which changed speed according to one of the two predefined waveforms. When implemented on the TRC test track, participants were required to follow lead vehicle speed changes on each of the (2-mile) straight road segments. In response to difficulties experienced by some pilot participants in maintaining a close following distance, the training protocol was modified to include additional training and feedback about the range of following distances considered acceptable. However, as in the previous study, because of documented individual differences in comfort associated with close following distances, a narrow range of following distances was not enforced. During the experiment, participants received feedback and monetary incentives based on their ability to maintain a consistent and relatively close following distance.

2.6.2 Peripheral Detection Task (PDT)

The Peripheral Detection Task (PDT) has been used in numerous studies to measure changes in drivers' ability to detect targets reflected on the windshield (Harms & Patten, 2003). This dashboard-mounted version with windshield reflections requires participants to detect targets at fixed locations. At intervals that varied randomly between 3 and 5 seconds, one of the 23 LEDs was illuminated (Figure 4). Each LED activation lasted 1.5 s, unless terminated by the driver's response. Drivers responded as quickly as possible by pressing a micro switch attached to their left index finger. During post-processing, valid responses were defined as responses recorded between 200 ms and 2000 ms following LED activation. Response times and proportion of targets detected were computed for each secondary task trial.

Figure 4. Illustration of PDT location with single LED activated

2.7 Secondary In-Vehicle Tasks

The protocol for this study involved two secondary tasks: 1) a navigation task using a simulated 511 traveler information system, and 2) a simulated phone task.

2.7.1 511 Task

Traveler information system tasks required participants to obtain and interpret information from a simulated 511 system to answer specific questions posed by the experimenter. Questions required participants to compare specified road segments with regard to a specific topic relating to traffic or road conditions. Traffic conditions included information about congestion and delay due to accidents. Road conditions included information about delays due to roadway construction or inclement weather. On half of the trials, the questions required use of a map (Figure 3).

Each question required the driver to initiate a phone call using speed dialing, navigate through a menu structure and obtain information relevant to the specific assigned problem. When the driver had solved the problem, he or she was instructed to recite the solution aloud so that the in-vehicle experimenter could record the answer and press the button to flag the end of the trial.

Each question required participants to compare information from two or more Interstate routes or numbered road segments. This required them to keep the route numbers in mind while navigating the menu structure and to remember the relevant information for each segment. For example, the driver may have been asked to determine which of two specified routes had no active construction or accidents. Alternately, the question might have asked the participant to compare two numbered road segments with respect to either traffic or road conditions. Use of the numbered segments required the use of a map to identify the route numbers that corresponded to each numbered segment.

Table 3 presents an example of the dialog involved in obtaining current traffic information about one of the routes shown in Figure 3. The generic aspects of the dialog were taken from an existing 511 system. User inputs are shown in italics.

Table 3. Example 511 system dialog

511 System	User Voice Input
Welcome to 511 Traffic and Travel Information. At any time you can say, “Go to menu” or “menu.”	*“Menu”* (voice input optional)
Menu. Here are all the categories you can choose from. When you hear the one you want, just say it. Traffic, Public Transit, Road Conditions. That’s all the categories. Just say the one you want.	*“Traffic”*
Say the name of a road to hear the detailed reports; for example say “I-64.”	*“I-75”*
Ok, I-75. There are three traffic incidents on I-75 . . .	

2.7.2 Simulated Phone Task

The Baddeley working memory span task was used to simulate a phone conversation of moderate intensity (Baddeley et al., 1985). The task required participants to listen to a sequence of sentences. After each sentence was presented, the participant was required to respond as to whether or not the sentence made sense. After each group of four sentences, the participant was prompted to recall either the subject or the object of each sentence. The targeted sentence component (subject or object) had been instructed previously. Thus, task performance required both decision-making, or judgment, and memory recall. Each group of four sentences comprised one task trial. Each call had four trials for a total of 16 sentences (8 meaningful + 8 nonsensical sentences). Participants were given 2.5 seconds for their sense/non-sense judgments and 6 seconds to recall the targeted sentence component.

To ensure that phone calls spanned the desired length of roadway, call audio files were constructed to last approximately 2.33 minutes.

All sentences were of the following construction: subject – action verb – object. Sentences were constructed so that the judgment decision could not be made until the object was heard. This required the participant to pay attention to the whole sentence before answering and also forced a consistent start to the response period (i.e., the completion of the last word of the sentence is the beginning of the response period.). It was necessary to balance the recall component of the task (subject/object) across calls. Thus, half of the calls requested a recall of the subjects of the sentence and half requested a recall of the sentence objects.

Sentences were constructed using common words having no more than three syllables. Sentences had a maximum of eight syllables and contained no adjectives. Sentences had been pilot tested to ensure consistent comprehension. Confusing or ambiguous sentences were eliminated following pilot testing.

The conversation task was presented by playing pre-recorded audio files of a human, male voice reciting the sentences and response prompts. Audio files were played over the phone using the same computer and telephone interface hardware described above for the simulated 511 system. Both the remote experimenter and the in-vehicle experimenter recorded responses to the task manually.

2.8 Monetary Incentives

Drivers were paid $20 per hour for their participation in the study. In addition, they had an opportunity to earn a modest amount of additional money during the experiment, based on their performance in the three tasks. Incentive amounts, shown in Table 4 by performance level, were designed to reinforce task priorities: the car-following task was most important, followed by the secondary task, and then the PDT.

Table 4. Monetary incentive amounts per lap

Task	Good Performance	Acceptable Performance	Poor Performance
Car Following	$1.50	$0.75	$0.0
Secondary	$1.00	$0.50	$0.0
PDT	$0.50	$0.25	$0.0
Total	$3.00	$1.50	$0.0

The participants had 6 opportunities to earn the monetary rewards shown in the table, one for each experimental condition. Evaluation criteria are shown in Table 5.

Table 5. Task performance incentive criteria

Task	Good Performance	Acceptable Performance	Poor Performance
Car Following	Maintains close following distance consistently with minor deviations	Maintains close following distance mostly with some noticeable deviations	Generally fails to maintain close following distance
511 Secondary Task	Uses proper voice commands to navigate hierarchical menu structure with minimal trouble and generally answers questions correctly	Has some difficulty navigating hierarchical menu structure and/or moderate number of incorrect answers	Has significant difficulty navigating menu structure and/or significant number of incorrect answers
Simulated Phone Secondary Task	Consistently makes correct judgments and recall	Moderate number of judgment/recall errors	Significant number of judgment/recall errors
PDT	Consistently attentive to target detection, detecting most targets	Moderate number of targets not detected	Fails to detect significant number of targets

3.0 PROCEDURES

Data were collected between June and August of 2004. The experimental protocol had four components: (1) introduction, general instructions and informed consent; (2) training and practice in a stationary vehicle; (3) test track data collection; and (4) participant debriefing. Each component is discussed in detail below.

3.1 Introduction, General Instructions and Informed Consent

Participants selected through phone screening were scheduled individually for a single session of approximately four hours. Upon arrival at the TRC proving grounds, the participant was escorted from the TRC guardhouse to Building 60 (NHTSA's Vehicle Research and Test Center, [VRTC]) by an experimenter. The participant was taken inside and given the Participant Information Summary, including informed consent statement and a confidentiality agreement, intended to protect other proprietary work ongoing at TRC. These documents are presented as Appendix A. The participant listened to audio recordings of the documents and was asked to follow along using the printed copy. After all questions had been answered, the participant signed the documents, thereby consenting to participate in the study and agreeing to maintain proving ground confidentiality

3.2 Training and Practice in a Stationary Vehicle

Task training and practice consisted of the following steps:

1) Instrumented Vehicle Orientation
2) Eye Tracker Setup (Initial Phase)
3) Test Track Guidelines
4) Driving Task (Car Following) Instructions
5) Phone Task Intro (with Phone Use Instructions)
6) Conversation (Baddeley) Task Instructions and Practice
7) 511 Traveler Information System Training and Practice
8) PDT Instructions
9) Monetary Rewards
10) Rating Scale Mental Effort
11) Eye Tracker Setup (Final Phase)

The participant was escorted to the SV. When seated in the vehicle, the participant was given an overview of the vehicle controls and displays, including adjusting the seat, steering wheel, and mirrors (see Appendix B). The participant was then asked to affix latex markers to his or her face for eye tracker calibration. During this procedure, the experimenter instructed the participant concerning head position and point of gaze.

Next, the participant was given test track guidelines (Appendix C), driving task instructions (Appendix D), and phone task introduction (with phone use instructions) (Appendix E). This was followed by practice dialing the phone. The simulated phone-conversation task was described (Appendix F) and the participant was given practice. The participant was then given a reference map for an imaginary roadway system and asked to read 511 traveler information system training instructions (Appendix G), which included practice. The participant was then given practice using the 511 system to answer specific questions and feedback about performance accuracy.

The PDT was described to the participant (Appendix H), after which practice was given. Next, the participant read about the monetary performance incentive system (Appendix I). Then, the Rating Scale Mental Effort (RSME) was described (Appendix J).

Following the completion of all training, eye tracker calibration was completed. The participant was then given an opportunity to ask questions about any aspect of the protocol. Data collection began following a break.

3.3 Test Track Data Collection

Accompanied by the experimenter, the participant was instructed to drive the SV to the test track, following the LV at all times. The experimenter was able to communicate directly with the LV driver (an experimental confederate) during all driving tasks via two-way radio. Similarly, both the LV driver and the experimenter accompanying the participant were able to communicate directly with the test track control tower at any time during the driving tasks.

The experiment consisted of 9 or more complete laps of the 7.5-mile test track. The LV operated in the first lane of traffic and collected data continuously throughout each lap, for a total of 18 trials, including practice. Additional trials were included if planned trials were aborted due either to equipment malfunction or other traffic in the travel lane. Each trial required approximately 2.5 minutes, during which the participant drove approximately 2–2.25 miles. Trials were run only on the two straight segments of the test track.

The participant was instructed to perform both primary tasks (car following and PDT target detection) on each trial. In addition, the participant was given instructions concerning one of three secondary task conditions, which include: (1) no secondary task, (2) simulated phone task, and (3) 511 task. Secondary task conditions changed after two trials had been completed in a given condition. Experimenters recorded participants' responses to the simulated phone task trials and 511 task trials. The lead and following vehicles stopped after completion of two trials in a given condition. The lead-vehicle driver initiated stops in the rightmost blacktop lane, which is one lane to the right of the lane in which the experiment was conducted. During the stops, the experimenter asked the participant to complete the RSME for the previous lap, provided feedback on performance, and gave new secondary task instructions to the participant. When the participant understood the new secondary task instructions, the experimenter notified the LV driver that the next lap could begin. The lead-vehicle driver accelerated in the blacktop lane and moved from the blacktop lane to the rightmost concrete lane when traffic was sufficiently clear to allow both vehicles to change lanes. The LV gradually increased speed to 55 mph.

The experimenter cued the participant to begin and, if necessary, to end each trial. The simulated phone task was paced automatically. The 511 tasks were self-paced and designed so that most participants could complete them within the 2-mile data collection interval. However, if a participant had excessive difficulty, the experimenter ended the data collection to allow sufficient time to prepare for the subsequent trial. Between trials, the experimenter instructed the participant to discontinue close car following and PDT performance, while reminding the participant not to get too far behind the LV.

After approximately 1–1.5 minutes, the experimenter instructed the participant to resume close car following and PDT performance in anticipation of the beginning of the next trial.

Practice consisted of three laps, each involving a different combination of primary and secondary tasks. On the first practice lap, the participant was instructed to establish a safe but close following distance that was to be used during all data collection trials. The participants were encouraged to maintain a close following distance appropriate for suburban freeways. Participants selected their own following distances; however, those who failed to follow closely enough to respond to LV speed during practice were encouraged to maintain a closer following distance. When the participant had established a consistent following distance, the experimenter instructed him or her to begin the PDT simultaneously for the remainder of the first practice lap. After this and each subsequent lap, the participant was instructed to stop the vehicle and complete the RSME. The experimenter also provided performance feedback.

The second practice lap included car following, the PDT, and the simulated phone task. At the beginning of each trial, the participant was asked to initiate a phone call, which initiated the simulated phone task. The third practice lap utilized the 511 traveler information system secondary task. Before each trial, the participant was given a specific question to answer concerning the imaginary roadway system. One trial allowed use of the map and one trial did not allow use of the map. This completed practice and the participant was given an opportunity to ask questions and to take a short break before the main trials began.

During each trial, the LV speed varied between approximately 50 and 65 mph. Prior to the onset of each trial, the LV driver activated an automated speed control mechanism, which controlled the LV speed according to one of two predefined patterns. The experimenter accompanying the participant had a secondary brake switch, which would activate braking in the event of a situation in which the SV became dangerously close to the rear of the LV. Occasionally, there was a vehicle stopped in the designated travel lane and the driver of the LV changed lanes. The experimental participant was instructed to change lanes whenever the LV changed lanes. In this way, the LV driver ensured a safe path ahead on the test track.

3.4 Participant Debriefing

At the completion of data collection, the LV and SV exited the track and returned to the VRTC. The participant exited the vehicle and proceeded to the conference room. The experimenter paid the participant a total of two amounts: (1) Base pay for participation, and (2) Performance incentive pay. The experimenter answered questions and returned the participant to his or her personal vehicle.

3.5 Data Reduction

The data were reduced to create performance measures in four categories:

1. Driving performance measures
2. PDT measures
3. Secondary task measures
4. Eye glance measures

Driving performance measures included measures of car-following performance, which included coherence, phase shift (delay) and modulus (gain). Other measures included lane position, primarily standard deviation of lane position (SDLP), and steering measures, including steering reversal rate, hold rate, and steering entropy (Boer, 2000). PDT measures included the percentage of correct responses during each trial and mean and standard deviation of response time during each trial. Secondary task measures included time to complete each task, the number of interactions and number of incorrect interactions. Eye glance measures included measures of glance frequency, glance duration, and the percentage of time spent looking at designated areas, which included the LV, PDT display, phone, and map.

Detailed specifications for data reduction are presented in Appendix K.

4.0 RESULTS

Analyses were conducted for five categories of performance measures: (1) vehicle control, including measures of steering and lane position; (2) decision making in car following, including headway, coherence and associated measures; (3) target detection, including the percentage of targets correctly detected and response time; (4) miscellaneous, including measures of secondary task performance and workload ratings; and (5) eye glance measures. An analysis of variance (ANOVA) was computed for each dependent measure using Proc MIXED in SAS, version 8.02. Age group (3 levels) was the single between-subjects factor. Secondary task (6 levels) and speed input signal (2 levels) were varied within subjects. We selected a set of planned comparisons to address the main questions of interest in the study. Specifically, in addition to the baseline (secondary task = None) and simulated phone task (Baddeley) conditions, we defined the Map, No Map, Error and No Error conditions as shown below:

	No Error	Error
Map	511M0 (1)	511ME (2)
No Map	51100 (3)	5110E (4)

Thus, the planned comparison of Map vs. No Map conditions involved the combination of cells (1+2) versus cells (3+4). Similarly the planned comparison of Error vs. No Error involved the comparison of cells (1+3) versus cells (2+4). In addition to the Map vs. No Map and Error vs. No Error comparisons, we planned to test all combinations of these (combined) factors with the Baddeley and None (no secondary task) conditions, plus Baddeley vs. None, for a total of 11 planned comparisons. Separate *F* tests were computed for each planned comparison. Probability values were adjusted for familywise error by using Hochberg's step-up method (Westfall, Tobias, Rom, Wolfinger, & Hochberg, 2003). Adjusted p values of less than .05 are considered to be statistically significant. Because the comparisons were planned in advance, there is no formal requirement that the omnibus F test be statistically significant before undertaking the planned comparisons (Myers & Well, 1991). However, we always conducted the omnibus *F* test for the secondary task main effects.

4.1 Vehicle Control Measures

Measures of vehicle control included several different measures of steering performance. Steering entropy (Boer, 2000) is an information measure of uncertainty, which characterizes the degree to which a steering signal differs from a smooth predictable signal. It is dimensionless and varies between 0.0 and 1.0. Higher values indicate more erratic steering characterized by high frequency corrective actions, assumed to follow periods of steering inactivity, due to distraction. More traditional measures, including steering reversal rate and steering hold rate (MacDonald & Hoffman, 1980) are defined as the number of occurrences divided by the time and therefore do not include information about the magnitude of the steering corrections. We also included a measure of the proportion of time during each trial associated with steering inactivity. Computational details are presented in Appendix K.

4.1.1 Steering Entropy

ANOVA results indicated a significant main effect of secondary task, $F\ (5,363) = 22.54$, $p < .0001$. The main effect of the speed signal was not significant, $F\ (1,363) = 0.25$, $p = .6194$, nor was the main effect of age group, $F\ (2,33) = 0.43$, $p = .6523$. None of the interactions was statistically significant.

The results of planned comparisons associated with the secondary task conditions are presented in Table 6. Means of individual task conditions are presented in Figure 5.

Table 6. Planned comparisons for effects of secondary task conditions on steering entropy

Comparison	DF	*F* value	Adjusted p-value	Interpretation
Map vs. No Map	1, 363	13.90	.0011	Map > No Map
Error vs. No Error	1, 363	0.25	.6194	
Baddeley vs. None	1, 363	34.59	< .0001	Baddeley > None
Baddeley vs. Map	1, 363	14.18	.0011	Map > Baddeley
None vs. Map	1, 363	111.46	< .0001	Map > None
Baddeley vs. Error	1, 363	5.99	.0595	
None vs. Error	1, 363	85.35	< .0001	Error > None
Baddeley vs. No Map	1, 363	0.52	.6194	
None vs. No Map	1, 363	56.46	< .0001	No Map > None
Baddeley vs. No Error	1, 363	4.17	.1258	
None vs. No Error	1, 363	78.02	< .0001	No Error > None

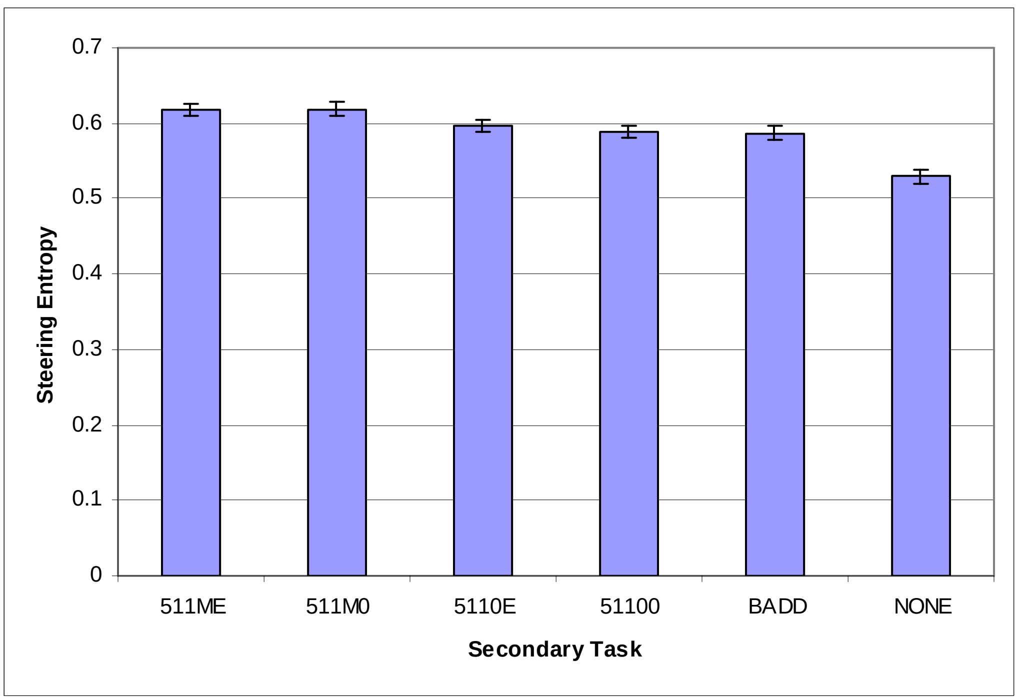


Figure 5. Mean steering entropy by secondary task condition (± standard error)

Steering entropy was significantly greater in every secondary task condition than in the baseline condition. In addition, the secondary task conditions that required use of a map had significantly greater steering entropy than both the no-map conditions and the simulated phone task condition. Use of the map to perform the 511 tasks was therefore most disruptive of steering entropy.

4.1.2 Steering Reversal Rate

We computed the average number of steering reversals per second for use as a dependent measure in the analysis of variance. We found that the steering reversal rate was significantly affected by the secondary task, $F\ (1, 165) = 11.41$, $p < .0001$. The results of planned comparisons are presented in Table 7. The mean values for each level of secondary task are presented in Figure 6.

Table 7. Planned comparisons for effects of secondary task conditions on steering reversal rate

Comparison	DF	*F* value	Adjusted p-value	Interpretation
Map vs. No Map	1, 165	3.87	.3056	
Error vs. No Error	1, 165	0.35	.8216	
Baddeley vs. None	1, 165	31.43	< .0001	Baddeley > None
Baddeley vs. Map	1, 165	0.67	.8216	
None vs. Map	1, 165	53.17	< .0001	Map > None
Baddeley vs. Error	1, 165	0.07	.8216	
None vs. Error	1, 165	45.29	< .0001	Error > None
Baddeley vs. No Map	1, 165	0.62	.8216	
None vs. No Map	1, 165	32.33	< .0001	No Map > None
Baddeley vs. No Error	1, 165	0.05	.8216	
None vs. No Error	1, 165	39.03	< .0001	No Error > None

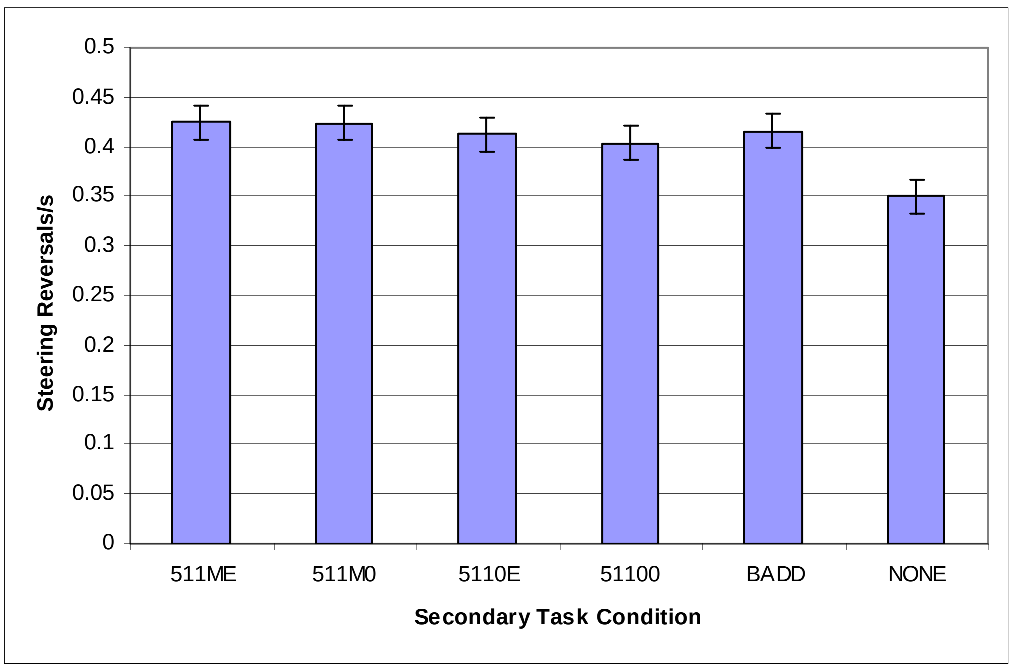


Figure 6. Mean steering reversals/second by secondary task condition (± standard error)

The steering reversal rate was lowest in the baseline condition. All secondary task conditions had significantly greater steering reversal rates; however, there were no differences among the various secondary task conditions for this measure.

The speed input signal did not affect steering reversal rate, *F* (1, 198) = 0.82, p = .3675, nor were there differences among the three age groups, *F* (2, 33) = 0.45, p = .6446. None of the interactions was statistically significant.

4.1.3 Steering Hold Rate

We computed the number of steering holds per second for each trial and used this as the dependent measure in our analysis of variance. The main effect of secondary task was statistically significant, *F* (5, 165) = 4.14, p = .0014. Results of planned comparisons are presented in Table 8. Mean values for each level of secondary task are presented in Figure 7.

Table 8. Planned comparisons for effects of secondary task conditions on steering hold rate

Comparison	DF	F value	Adjusted p-value	Interpretation
Map vs. No Map	1, 165	2.12	.5361	
Error vs. No Error	1, 165	1.41	.5361	
Baddeley vs. None	1, 165	4.84	.2047	
Baddeley vs. Map	1, 165	3.27	.4344	
None vs. Map	1, 165	18.90	.0003	None > Map
Baddeley vs. Error	1, 165	2.89	.4559	
None vs. Error	1, 165	17.97	.0004	None > Error
Baddeley vs. No Map	1, 165	0.38	.5361	
None vs. No Map	1, 165	9.98	.0150	None > No Map
Baddeley vs. No Error	1, 165	0.53	.5361	
None vs. No Error	1, 165	10.68	.0118	None > No Error

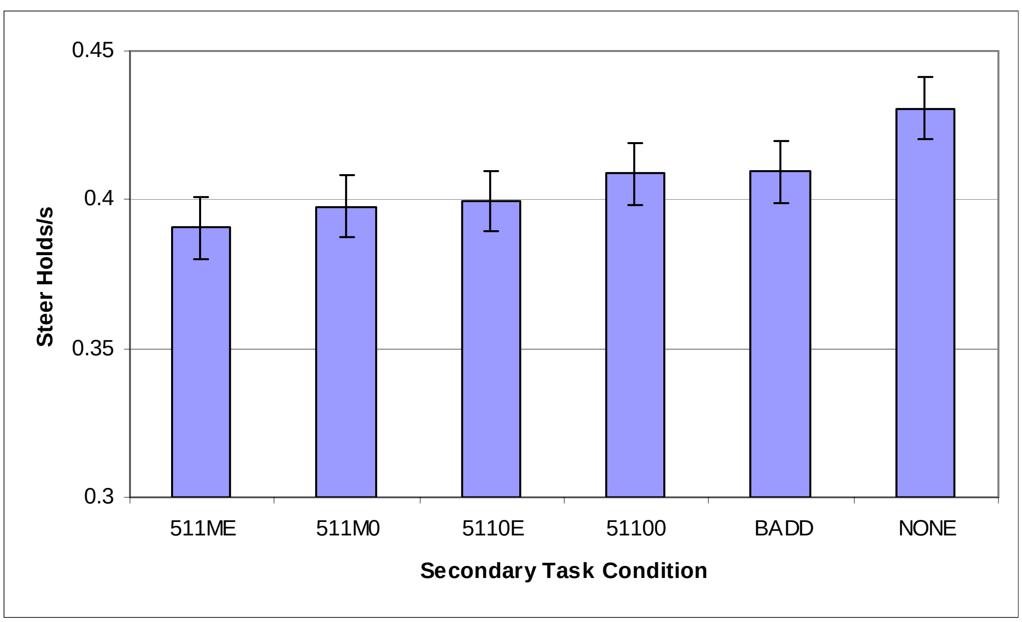


Figure 7. Mean steering hold rate by secondary task condition (± standard error)

The baseline condition was associated with the highest steering hold rates. Based on the results of planned comparisons, the 511 task conditions all had significantly lower steering hold rates, while the simulated phone task had an intermediate value that was not statistically different from either the baseline or the 511 task conditions. Neither the main effect of speed, F (1, 198) = 0.06, p = .8110, nor the main effect of age group, F (2, 33) = 0.20, p = .8229, nor any of the interaction effects was statistically significant.

4.1.4 Proportion of Time without Steering Activity

To get a measure of the proportion of time without steering activity, we summed the duration of all steering holds of any duration within a trial and divided this by the total duration of the trial. The resulting measure, referred to as p(hold), appeared to be more sensitive to the experimental manipulations than the hold frequency, analyzed above. The ANOVA results indicated that the main effect of secondary task was statistically significant, F (5, 165) = 11.58, p < .0001. However, neither, the main effect of speed input signal, F (1, 198) = 0.53, p = .4684, nor the main effect of age group, F (2, 33) = 0.36, p = .7012, nor any of the interactions among these factors was statistically significant. Results of the planned comparisons are presented in Table 9, while mean values for the secondary task conditions are presented in Figure 8.

Table 9. Planned comparisons for effects of secondary task conditions on proportion of trial without steering activity

Comparison	DF	F value	Adjusted p-value	Interpretation
Map vs. No Map	1, 165	4.76	.1835	
Error vs. No Error	1, 165	1.55	.8948	
Baddeley vs. None	1, 165	23.43	< .0001	None > Baddeley
Baddeley vs. Map	1, 165	3.66	.2871	
None vs. Map	1, 165	56.30	< .0001	None > Map
Baddeley vs. Error	1, 165	1.74	.7539	
None vs. Error	1, 165	47.75	< .0001	None > Error
Baddeley vs. No Map	1, 165	0.02	.8948	
None vs. No Map	1, 165	32.74	< .0001	None > No Map
Baddeley vs. No Error	1, 165	0.53	.8948	
None vs. No Error	1, 165	39.88	< .0001	None > No Error

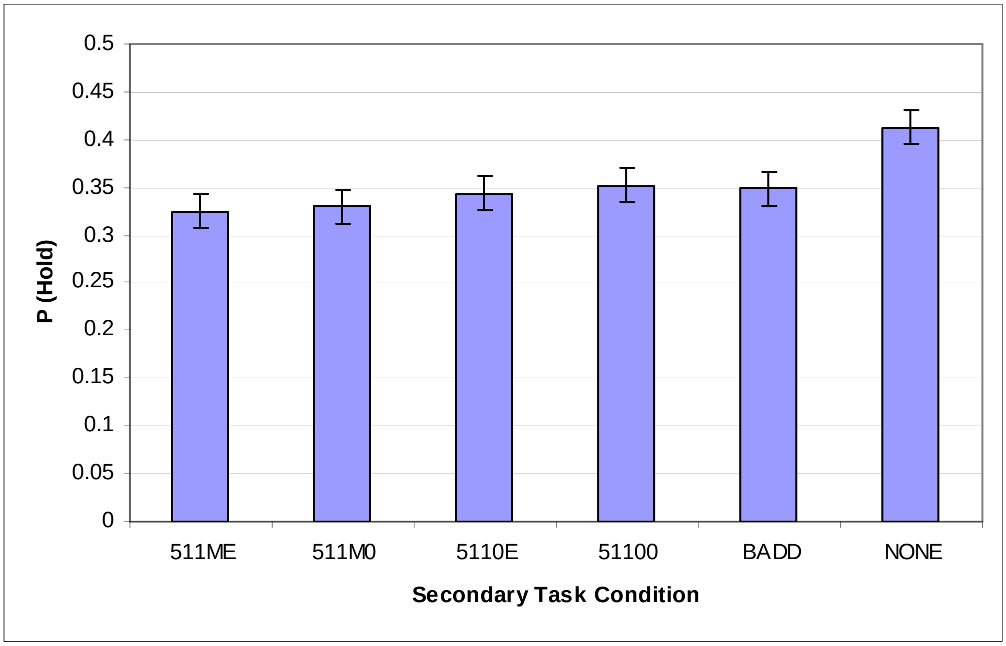


Figure 8. Mean proportion of time of steering inactivity by secondary task condition (± standard error)

The results indicate strong differences between the baseline condition and all other secondary task conditions, with no differences between any of the secondary task conditions.

We examined the correlations between the various measures of steering performance. The results are presented in Table 10.

Table 10. Correlations between pairs of steering measures

	Hold rate	Steer reversal rate	P (hold)
Steer reversal rate	-.73		
P (hold)	.83	-.93	
Steering entropy	-.25	.26	-.28

It is evident, based on the relatively low correlations, that steering entropy reflects behavior that is largely different from the other steering measures. In contrast, the probability of a steering hold is highly correlated with both the steering reversal and hold rates. The negative correlation between steering reversal rate and the probability of a hold indicates that as the proportion of steering inactivity increases, the steering reversal rate decreases. This follows from the definitions of steering reversal rates and suggests that these measures are not independent.

4.1.5 Standard Deviation of Lane Position (SDLP)

We computed the standard deviation of lane position for each trial. We eliminated trials that involved lane changes and eliminated segments of each trial that indicated GPS dropout errors. The resulting trials were all the same length. The ANOVA results indicated significant main effects of secondary task, F (5, 289) = 3.14, p = .0089, and input speed signal, F (1, 291) = 11.71, p = .0007. The main effect of age group was not significant, F (2, 29.7) = 0.69, p = .5085. Means for secondary task conditions are presented in Figure 9 and results of planned comparisons are presented in Table 11.

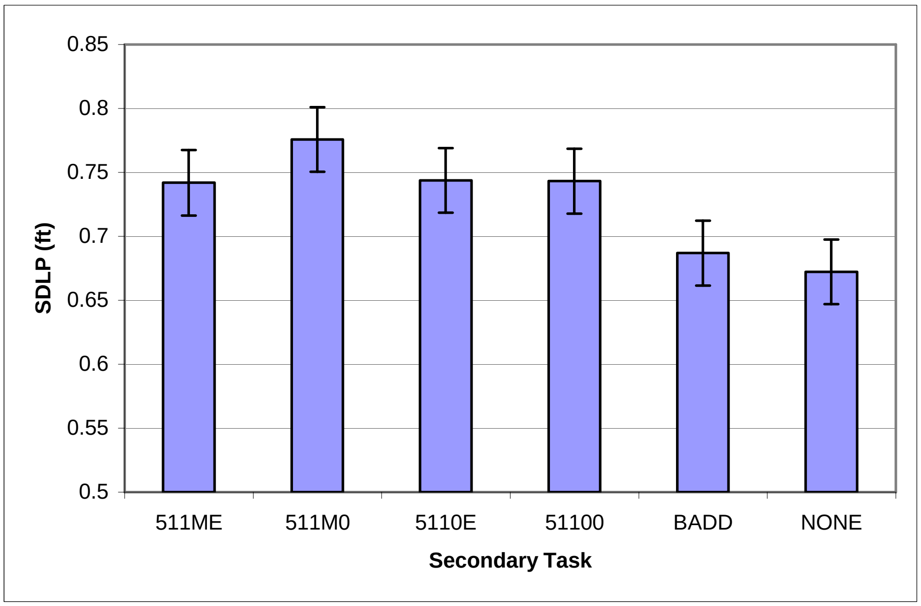


Figure 9. Mean SDLP by secondary task condition (± standard error)

Table 11. Planned comparisons for effects of secondary task conditions on SDLP

Comparison	DF	F value	Adjusted p-value	Interpretation
Map vs. No Map	1, 289	0.82	.6347	
Error vs. No Error	1, 289	0.92	.6347	
Baddeley vs. None	1, 289	0.23	.6347	
Baddeley vs. Map	1, 289	7.08	.0659	
None vs. Map	1, 289	10.44	.0137	None < Map
Baddeley vs. Error	1, 289	3.61	.2331	
None vs. Error	1, 289	6.08	.0854	
Baddeley vs. No Map	1, 289	3.72	.2331	
None vs. No Map	1, 289	6.24	.0854	
Baddeley vs. No Error	1, 289	7.23	.0659	
None vs. No Error	1, 289	10.66	.0135	None < No Error

The trend shown by the means in Figure 9 suggests strong differences among the various secondary task conditions. However, the relatively large standard error estimates, shown by the error bars, reveal a large amount of variation in this measure. Therefore, while the trend is consistent with the hypothesis that the 511 tasks were associated with higher levels of lane position error than the simulated phone task or baseline, the results of statistical tests show that several of these tests did not reach statistical significance. Thus, the only differences to attain statistical significance were between two of the 511 conditions (map, no error) and baseline. As suggested by the figure, both findings may be a reflection of the map effect, since the 511MO condition had the largest mean value.

SDLP values associated with trials that used the complex speed input signal were greater ($M = 0.76$) than those associated with trials that used the sine wave ($M = 0.69$).

4.2 Driver Decision Making in Car Following

Brookhuis and colleagues (1994) developed the car-following coherence paradigm and have argued that car-following delay, formally defined as the phase shift between the two speed signals, is a measure that incorporates decision making and judgment. This measure is thus a measure of mid-level tactical performance, incorporating more cognitive components than the vehicle control measures considered above.

4.2.1 Coherence

Coherence is a measure of squared correlation, which reflects the degree to which the following vehicle is able to match the periodicity of the LEAD VEHICLEspeed signal. Coherence is used both as a measure of car-following performance and as a test of whether the associated measures of phase shift (car-following delay) and modulus (car-following gain) are interpretable. Computational details are presented in Appendix K.

Coherence values were computed for 408 (94.4%) of the 432 trials run. Approximately 89% of these trials had coherence values of .8 or greater, indicating highly accurate car following. It is worth noting that all but one of the trials with coherence values less than .8 were associated with the complex speed signal. This indicates that the drivers had greater difficulty adapting to the speed changes of the complex speed signal. The ANOVA results indicated a significant main effect of secondary task, $F(5,341) = 5.89$, $p < .0001$. The results of planned comparisons are presented in Table 12. The means are shown in Figure 10.

Table 12. Planned comparisons for effects of secondary task conditions on car-following coherence

Comparison	DF	F value	Adjusted p-value	Interpretation
Map vs. No Map	1, 341	2.92	0.2653	
Error vs. No Error	1, 341	1.55	0.4284	
Baddeley vs. None	1, 341	0.00	0.9941	
Baddeley vs. Map	1, 341	17.41	0.0004	Baddeley > Map
None vs. Map	1, 341	17.34	0.0004	None > Map
Baddeley vs. Error	1, 341	8.80	0.0199	Baddeley > Error
None vs. Error	1, 341	8.75	0.0199	None > Error
Baddeley vs. No Map	1, 341	7.71	0.0238	Baddeley > No Map
None vs. No Map	1, 341	7.67	0.0238	None > No Map
Baddeley vs. No Error	1, 341	15.86	0.0007	Baddeley > No Error
None vs. No Error	1, 341	15.79	0.0007	None > No Error

As shown by the test results, the baseline and simulated phone task conditions had generally higher coherence values than the other four secondary task conditions. We considered the possibility that this pattern may have been due in part to the fact that these two conditions had longer durations than the other task conditions. Specifically, we looked at the correlation between task duration and coherence. Since this correlation was essentially non-existent (r < .01), we concluded that the difference in task durations did not influence the differences in coherence due to secondary task condition. Note that the correlations between task duration and phase and modulus were no stronger.

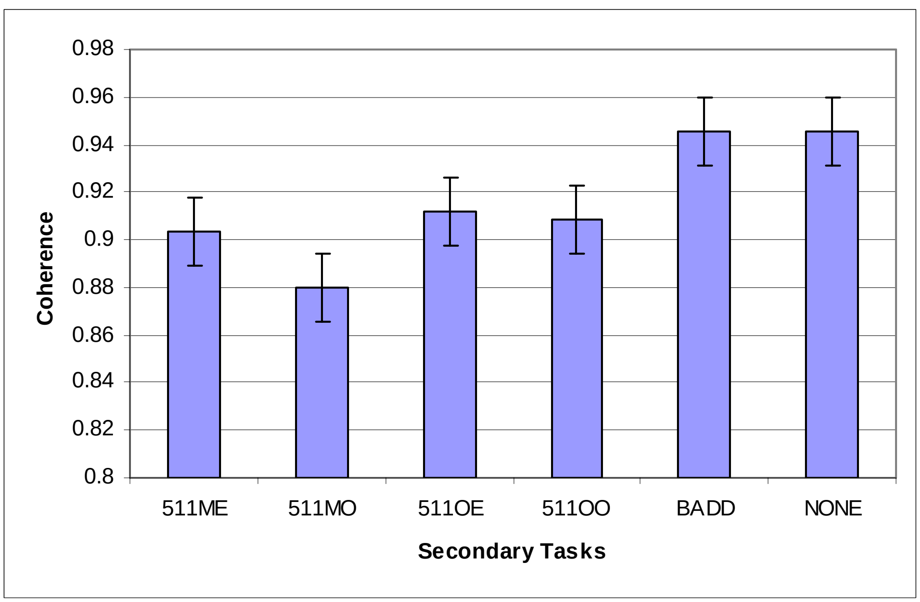


Figure 10. Mean coherence by secondary task condition (± standard error)

Coherence differed significantly by the type of input signal, F (1, 341) = 163.6, p < .0001 and by age group, F (2,31) = 4.05, p = .0273. However, this effect differed across age groups, as evidenced by the significant Age Group x Speed Signal interaction, F (2, 341) = 5.84, p = .0032, which is shown in Figure 11.

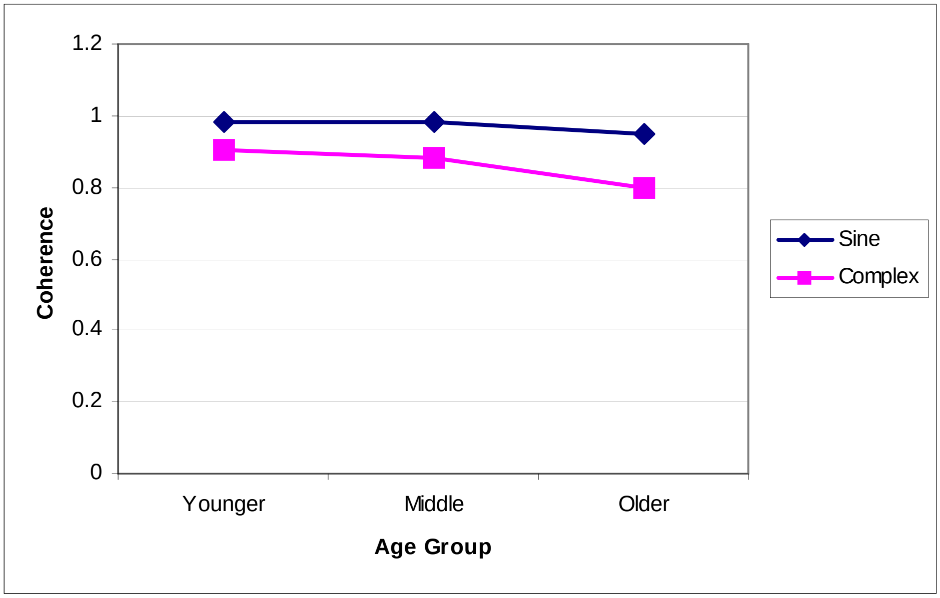


Figure 11. Effects of speed signal on coherence by age group

Post hoc comparisons, summarized in Table 13, revealed that the older age group had significantly lower coherence scores than the other two age groups on trials in which the complex speed signal was used.

Table 13. Post hoc comparisons for effects of age group x speed signal on car-following coherence

Comparison	DF	F value	Adjusted p-value	Interpretation
Old vs. Middle for Complex	1, 42.2	8.98	.0182	Middle > Old for Complex
Young vs. Middle for Complex	1, 42.2	0.64	.8563	
Old vs. Middle for Sine	1, 42.2	1.33	.7678	
Young vs. Middle for Sine	1, 42.2	0.01	.9394	

4.2.2 <u>Phase Shift (Delay in Car Following)</u>

When coherence is relatively high (e.g., ≥ 0.80), the driver is adequately following the lead vehicle's speed changes, which implies that the associated measures are meaningful. We therefore selected only trials for which coherence was greater than .8 for further analysis. We found that 44 (11%) of 408 trials had coherence less than .8. To assess the possibility of bias due to different sample sizes, we compared the percentage of trials with coherence values less than .8 by interface condition and age group. There were no differences across interface conditions; however, there were differences among the age groups. Specifically, the older group had a greater proportion of trials with coherence less than .8. This pattern serves to underscore the generally lower coherence values among the older drivers, shown above. The potential effect of this difference on subsequent analyses is likely to be minimal, due to the relatively high proportion of trials among all groups with coherence > .8. However, any differences between older drivers and the other age groups could be slightly understated due to the reduced sample size for this group.

The phase shift represents the driver's delay in responding to the changes in lead vehicle speed. It is a response time that reflects performance over the entire data collection interval, in this case, during the performance of the secondary task. The ANOVA revealed significant main effects of secondary task, $F(5, 140) = 11.15$, $p < .0001$, and age group, $F(2, 30.8) = 8.00$, $p = .0016$. The planned comparisons for the main effect of secondary task are presented in Table 14. Mean values for the different secondary task conditions are presented in Figure 12.

Table 14. Planned comparisons for effects of secondary task conditions on car-following delay

Comparison	DF	F value	Adjusted p-value	Interpretation
Map vs. No Map	1, 148	22.15	< .0001	No Map < Map
Error vs. No Error	1, 148	0.26	0.6845	
Baddeley vs. None	1, 127	4.89	0.0888	
Baddeley vs. Map	1, 134	19.64	0.0001	Baddeley < Map
None vs. Map	1, 133	47.96	< .0001	None < Map
Baddeley vs. Error	1, 134	7.00	0.0457	Baddeley < Error
None vs. Error	1, 133	26.47	< .0001	None < Error
Baddeley vs. No Map	1, 135	0.17	0.6845	
None vs. No Map	1, 133	8.41	0.0262	None < No Map
Baddeley vs. No Error	1, 135	4.83	0.0888	
None vs. No Error	1, 134	21.95	< .0001	None < No Error

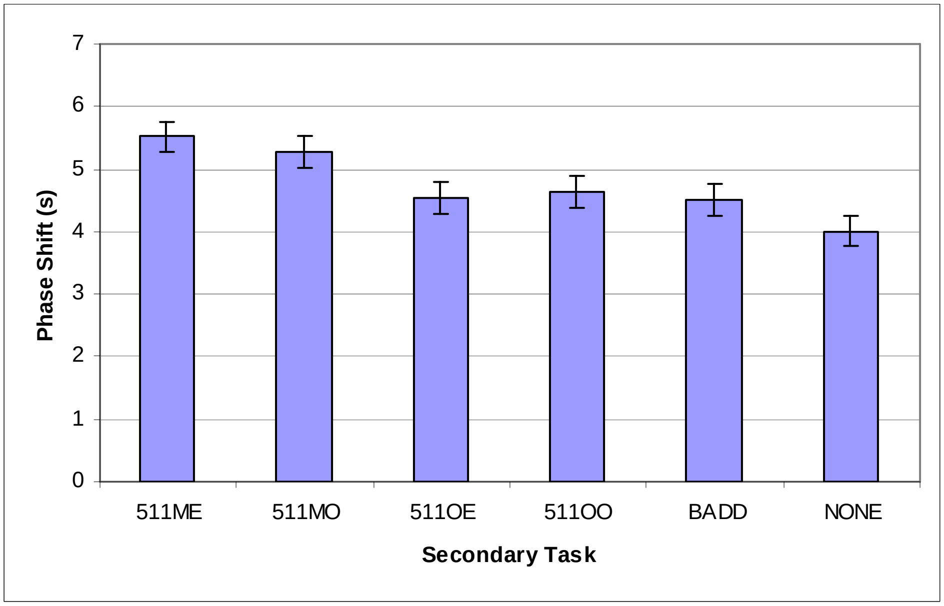


Figure 12. Mean car-following delay by secondary task condition (± standard error)

The results indicate that car-following delay was shortest on trials with no secondary task and longest on the trials with a 511 task requiring the use of the map. As shown in Table 14, all conditions were significantly greater than the baseline condition, reflecting the increased load associated with any secondary task. For this measure, the performance degradation associated with the simulated phone task was no different from that associated with the 511 task that did not require use of the map. The 511 tasks requiring map use were associated with significantly longer delays than all other task conditions.

Means for the three age groups are shown in Figure 13.

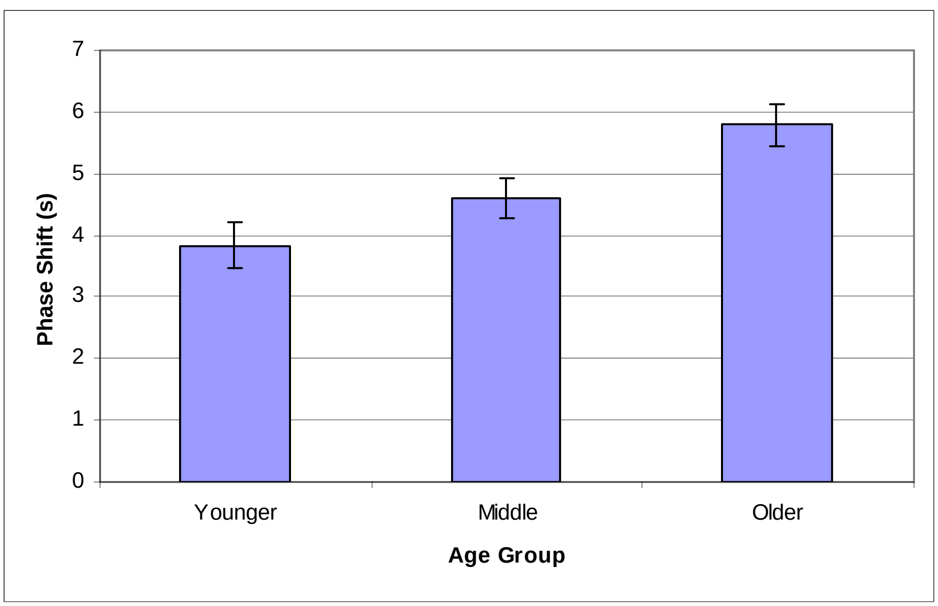


Figure 13. Mean car-following delay by age group (± standard error)

Clearly, car-following delay increased with increasing age. The mean delay for drivers in the young age group was 3.8 s, versus 4.6 s for the middle group and 5.8 s for the older group. Statistical tests, summarized in Table 15, show that the older group had significantly longer delay on average than the middle age group, but that the difference between the middle age group and the younger group was not statistically significant.

Table 15. Pairwise comparisons for effects of age group on car-following delay

Comparison	DF	F value	Adjusted p-value	Interpretation
Old vs. Middle	1, 31	15.49	.0009	Old > Middle
Middle vs. Young	1, 30.3	2.35	.1360	

4.2.3 Modulus

Modulus (gain) reflects the following driver's responses at the extreme values of the lead vehicle speed. Modulus values near 1.0 indicate that the following driver is closely matching the extreme speed values of the lead vehicle. Modulus values significantly greater than 1.0 indicate potentially aggressive overcorrection and values smaller than 1.0 indicate undercorrection. Values at either extreme reflect increased potential for safety problems resulting from inadequate following distances.

As with phase, modulus values are only meaningful if the coherence is relatively high. Therefore, we selected only trials for which coherence was greater than .8 for further analysis. ANOVA results revealed significant main effects of secondary task, $F\ (5,151) = 10.35$, $p < .0001$ and age group, $F\ (2, 31.6) = 5.89$, $p = .0067$. Results of planned comparisons are presented in Table 16. Means for the two significant main effects are presented in Figures 14 and 15.

Table 16. Planned comparisons for effects of secondary task conditions on car-following modulus (gain)

Comparison	DF	F value	Adjusted p-value	Interpretation
Map vs. No Map	1, 158	0.19	.7944	
Error vs. No Error	1, 158	5.66	.1115	
Baddeley vs. None	1, 137	29.45	< .0001	None > Baddeley
Baddeley vs. Map	1, 145	0.40	.7944	
None vs. Map	1, 144	29.91	< .0001	None > Map
Baddeley vs. Error	1, 145	0.33	.7944	
None vs. Error	1, 144	44.89	< .0001	None > Error
Baddeley vs. No Map	1, 145	0.07	.7944	
None vs. No Map	1, 144	34.28	< .0001	None > No Map
Baddeley vs. No Error	1, 145	2.14	.7280	
None vs. No Error	1, 144	21.43	< .0001	None > No Error

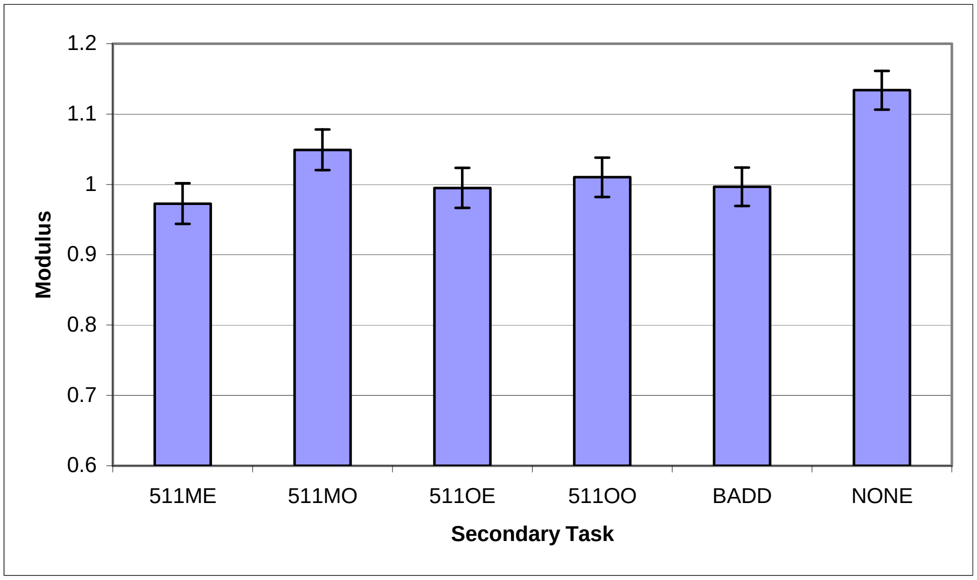


Figure 14. Mean car-following modulus by secondary task condition (± standard error)

Modulus was significantly higher in the baseline condition than in all other conditions. None of the other secondary conditions were significantly different in terms of modulus. The fact that modulus values for trials involving secondary tasks were closer to one than those associated with baseline trials is puzzling. It suggests that drivers were doing a better job of matching the extreme speed values while performing a secondary task and that they were more aggressively over-correcting in the baseline condition.

Part of the explanation for this finding is related to the effects of driver age group on modulus. As shown in Table 17 and Figure 15, older drivers had significantly lower modulus values than either middle aged or younger drivers, reflecting more conservative behavior. The difference between middle and young drivers was not significant; however, the trend is consistent with a general age effect of increasingly conservative responding with increasing age.

Table 17. Pairwise comparisons for effects of age group on car-following modulus

Comparison	DF	F value	Adjusted p-value	Interpretation
Old vs. Middle	1, 31.8	9.98	.0069	Old < Middle
Middle vs. Young	1, 31	0.34	.5624	

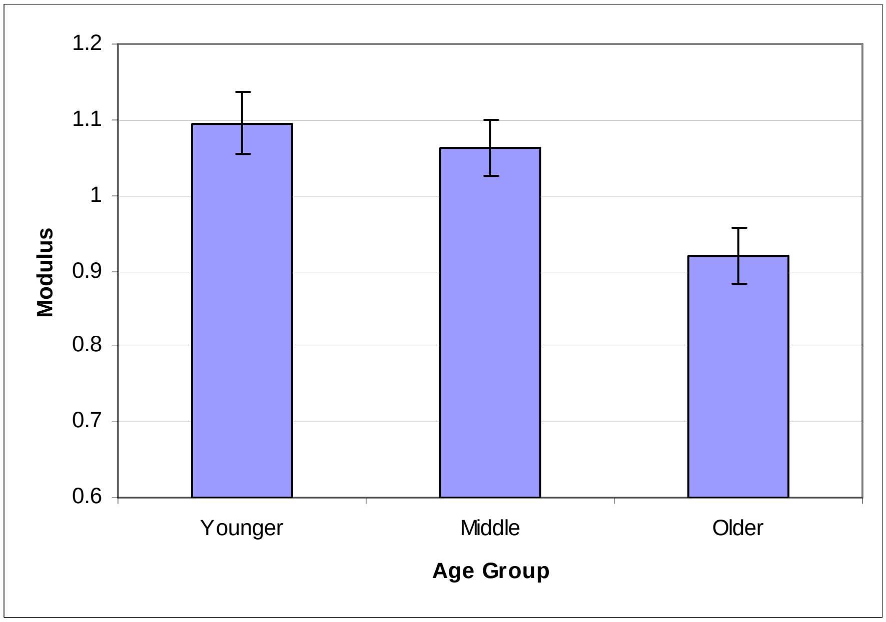


Figure 15. Mean car-following modulus by age group (± standard error)

4.2.4 Headway

During the experiment, participants were instructed to maintain a constant following distance during all trials. Our previous work (Ranney et al., 2004a) as well as that of Brookhuis (Brookhuis, De Vries, & De Waard, 1991) has shown that drivers have considerable difficulty maintaining a fixed following distance. We therefore allowed drivers to select their own following distance and encouraged them to maintain that distance. However, we have seen that despite instructions, some drivers increased their following distances while performing secondary tasks.

The present ANOVA results indicated significant main effects of secondary task, F (5, 145) = 5.03, p = .0003, and of age group, F (2,29) =3.63, p = .0392. The means and results of the planned comparisons for the former effect are presented in Figure 16 and Table 18, respectively, while those for the latter effect are presented in Figure 17 and Table 19, respectively.

Table 18. Planned comparisons for effect of secondary task on mean headway

Comparison	DF	F value	Adjusted p-value	Interpretation
Map vs. No Map	1, 145	6.34	.0774	
Error vs. No Error	1, 145	0.83	.3961	
Baddeley vs. None	1, 145	0.72	.3961	
Baddeley vs. Map	1, 145	19.72	.0002	Baddeley < Map
None vs. Map	1, 145	11.96	.0064	None < Map
Baddeley vs. Error	1, 145	9.25	.0224	Baddeley < Error
None vs. Error	1, 145	4.24	.1654	
Baddeley vs. No Map	1, 145	5.69	.0919	
None vs. No Map	1, 145	1.97	.3961	
Baddeley vs. No Error	1, 145	14.32	.0022	Baddeley < No Error
None vs. No Error	1, 145	7.85	.0404	None < No Error

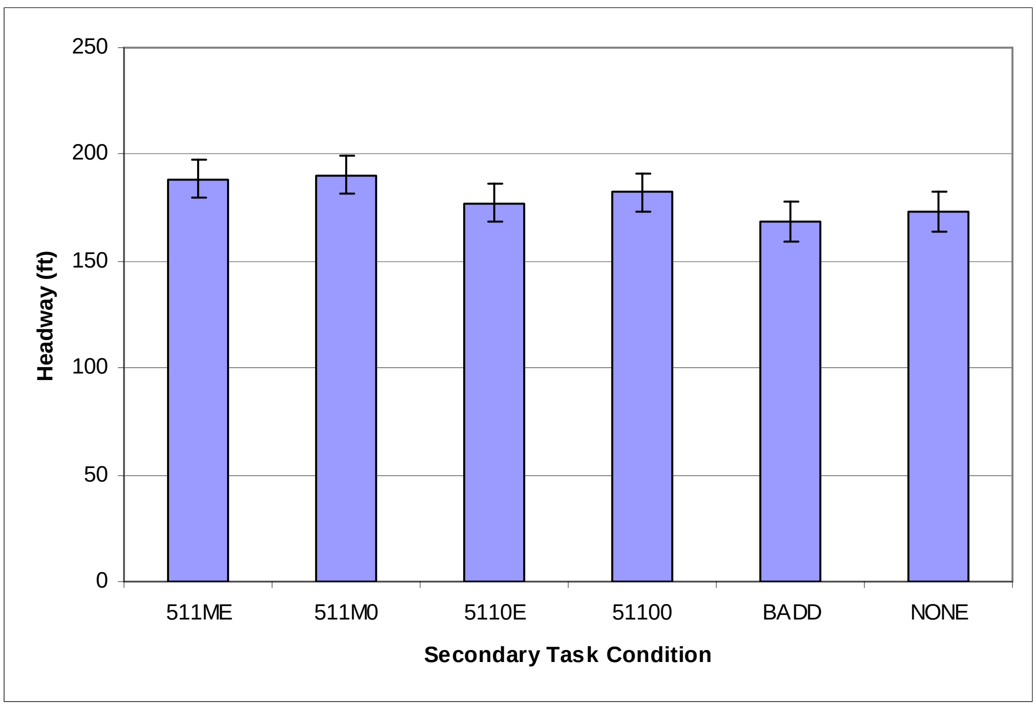


Figure 16. Mean headway by secondary task condition (± standard error)

Table 19. Pairwise comparisons for effects of age group on mean headway

Comparison	DF	F value	Adjusted p-value	Interpretation
Old vs. Middle	1, 29	6.49	.0328	Old > Middle
Middle vs. Young	1, 39	0.69	.4138	

The results indicate that drivers in the older age group drove at consistently longer following distances than drivers in the other two age groups.

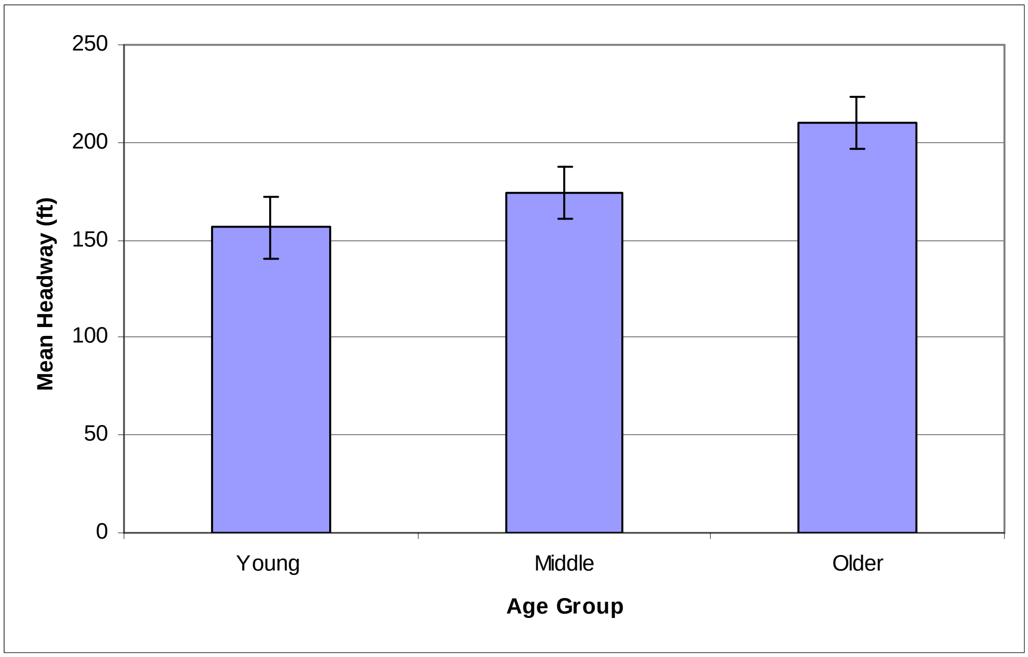


Figure 17. Mean headway by age group (± standard error)

Previously (Ranney et al., 2004a), we noted a relatively strong correlation between headway and coherence. We therefore examined the correlation in the present data. The relatively low correlation (r = -.35) indicates that coherence was not strongly influenced by the headway.

4.3 <u>Target Detection Performance</u>

During each driving trial, drivers responded to approximately 20 targets in the Peripheral Detection Task (PDT). Responses recorded between 0.2 and 3.0 seconds following the target activation were considered correct responses. We computed mean response time and the percent correct for each driving trial. ANOVAs were computed using the model described above.

4.3.1 PDT Percent Correct

Secondary task had a significant effect on the percent of targets detected, F (5,165) = 14.18, $p < .0001$. Planned comparisons are presented in Table 20 and the means are presented in Figure 18.

Table 20. Planned comparisons for effect of secondary task on proportion of PDT targets detected

Comparison	DF	F value	Adjusted p-value	Interpretation
Map vs. No Map	1, 165	28.86	< 0.0001	No Map > Map
Error vs. No Error	1, 165	0.13	0.9400	
Baddeley vs. None	1, 165	26.12	< 0.0001	None > Baddeley
Baddeley vs. Map	1, 165	3.89	0.2010	
None vs. Map	1, 165	61.99	< 0.0001	None > Map
Baddeley vs. Error	1, 165	0.01	0.9400	
None vs. Error	1, 165	33.94	< 0.0001	None > Error
Baddeley vs. No Map	1, 165	5.83	0.0843	
None vs. No Map	1, 165	12.16	0.0038	None > No Map
Baddeley vs. No Error	1, 165	0.13	0.9400	
None vs. No Error	1, 165	30.63	< 0.0001	None > No Error

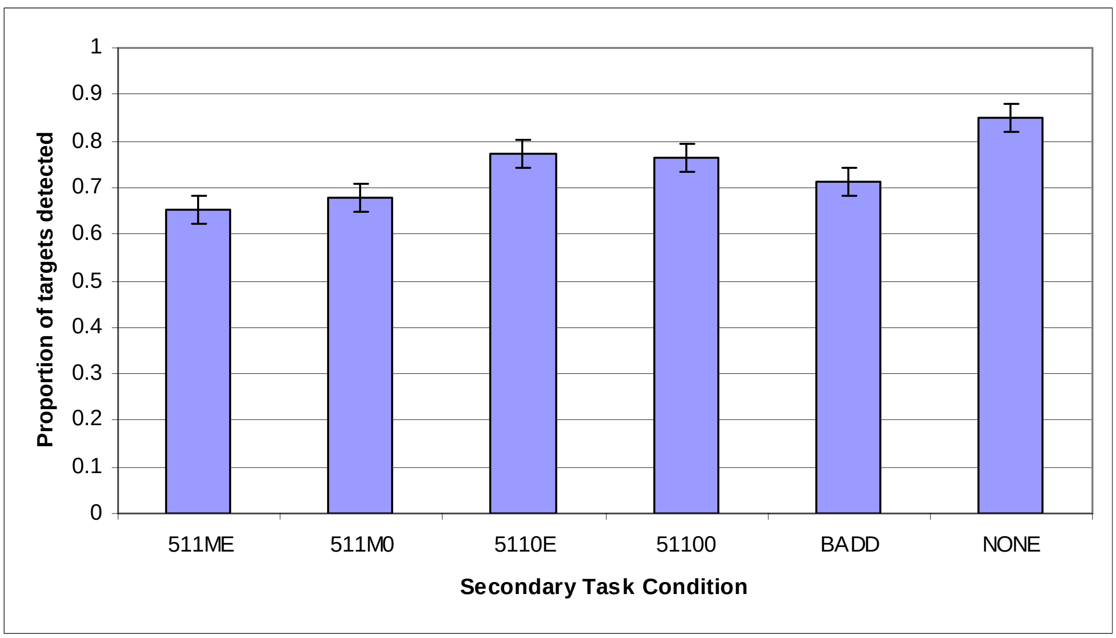


Figure 18. Mean proportion of PDT targets detected by secondary task condition (± standard error)

As predicted, the proportion of PDT targets detected was highest when there was no secondary task. All comparisons with the baseline condition revealed significantly fewer targets detected. Furthermore, the performance in the 511 tasks that required the use of a map was most degraded, relative to the baseline performance and as shown in Tab1e 20, 511 map tasks were associated with significantly fewer targets detected than 511 tasks that did not require the use of a map. There was no main effect of age group, F (2, 33) = 1.24, p = .3015, however the Agegroup x Speed Input Signal interaction was significant, F (2, 198) = 3.54, p = .0307. This interaction effect is shown in Figure 19. Adjusted post hoc comparisons of mean pairs revealed no significant differences between age groups separated by speed or by age group, suggesting a relatively weak effect.

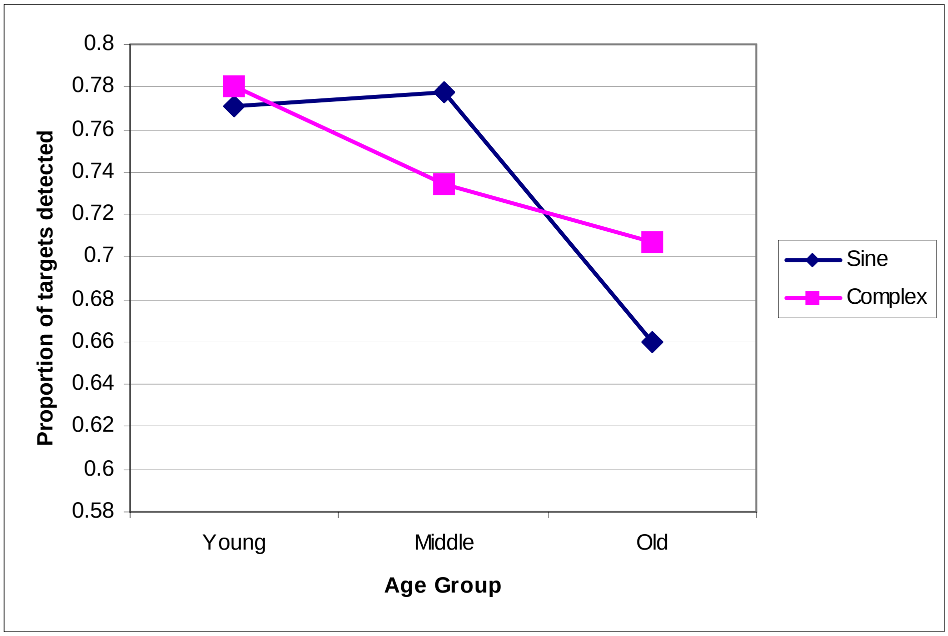


Figure 19. Effects of Age Group and Speed Signal on proportion of PDT targets detected

4.3.2 PDT Response Time

Secondary task had a significant effect on the PDT Response Time, F (5,161) = 24.77, p < .0001. Planned comparisons are presented in Table 21 and the means are presented in Figure 20.

Table 21. Planned comparisons for effect of secondary task on PDT response time

Comparison	DF	*F* value	Adjusted p-value	Interpretation
Map vs. No Map	1, 162	16.48	0.0005	No Map < Map
Error vs. No Error	1, 162	0.05	0.9440	
Baddeley vs. None	1, 161	81.93	< 0.0001	None < Baddeley
Baddeley vs. Map	1, 163	0.00	0.9440	
None vs. Map	1, 161	107.78	< 0.0001	None < Map
Baddeley vs. Error	1, 163	3.29	0.2857	
None vs. Error	1, 161	74.44	< 0.0001	None < Error
Baddeley vs. No Map	1, 161	11.48	0.0044	No Map < Baddeley
None vs. No Map	1, 160	50.24	< 0.0001	None < No Map
Baddeley vs. No Error	1, 161	2.68	0.3110	
None vs. No Error	1, 160	78.30	< 0.0001	None < No Error

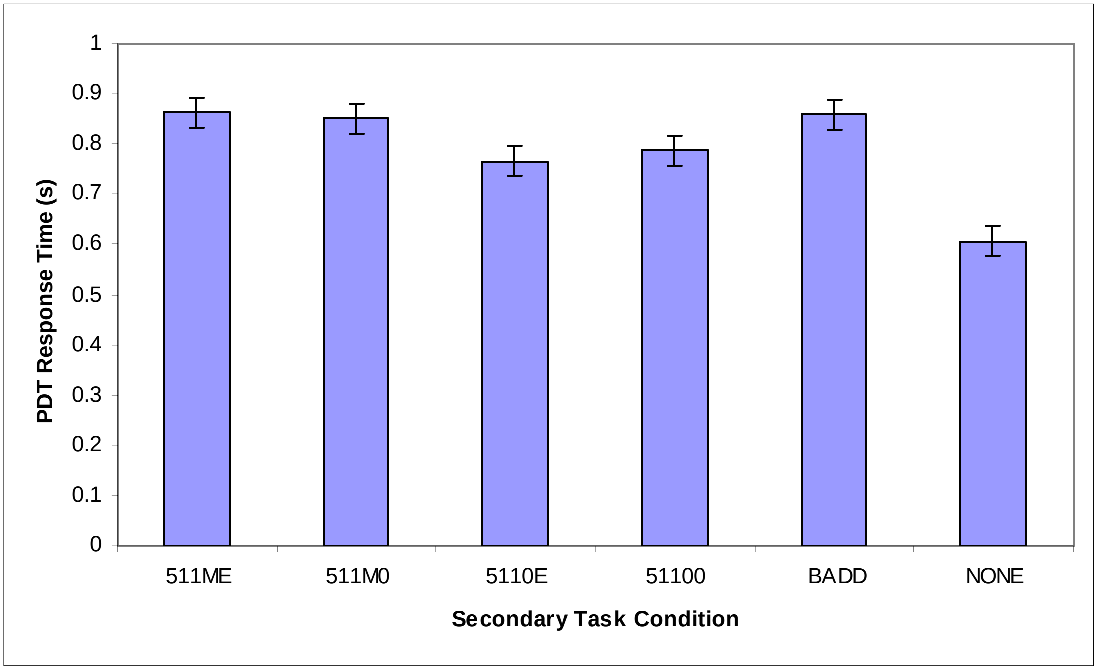


Figure 20. Mean PDT response time by secondary task condition (± standard error)

Neither the main effect of age group, F (2, 33) = 1.50, p = .2378, nor the main effect of speed, F (1, 192) = 0.02, p = .8842, nor any of the interactions involving secondary task, age group, and speed input signal was statistically significant.

4.4 Other Performance Measures

4.4.1 RSME Workload Ratings

Workload ratings were analyzed with the same statistical model, except speed input signal was not a factor because ratings were not taken for each individual driving trial. The main effect of secondary task was significant, F (5, 163) = 58.07, p < .0001. The results of planned comparisons are presented in Table 22 and the means for each condition are presented in Figure 21.

Table 22. Planned comparisons for effects of secondary task conditions on RSME workload ratings

Comparison	DF	F value	Adjusted p-value	Interpretation
Map vs. No Map	1, 163	18.22	0.0002	Map > No Map
Error vs. No Error	1, 163	1.92	0.5036	
Baddeley vs. None	1, 163	165.41	< 0.0001	Baddeley > None
Baddeley vs. Map	1, 163	1.92	0.5036	
None vs. Map	1, 163	268.59	< 0.0001	Map> None
Baddeley vs. Error	1, 163	0.05	0.8281	
None vs. Error	1, 163	229.85	< 0.0001	Error > None
Baddeley vs. No Map	1, 163	4.29	0.1993	
None vs. No Map	1, 163	165.09	< 0.0001	No Map > None
Baddeley vs. No Error	1, 163	0.82	0.7325	
None vs. No Error	1, 341	15.79	< 0.0001	No Error > None

Workload ratings were significantly higher than baseline for all categories of secondary task. Otherwise, drivers rated secondary tasks requiring use of the map as requiring more effort than those that did not require the map. The simulated phone task was essentially midway between the two 511 conditions (Map and No-Map) and was thus not significantly different from either condition. Despite a suggestive trend in that direction, the 511 trials involving system errors were not rated as significantly more difficult than those without errors.

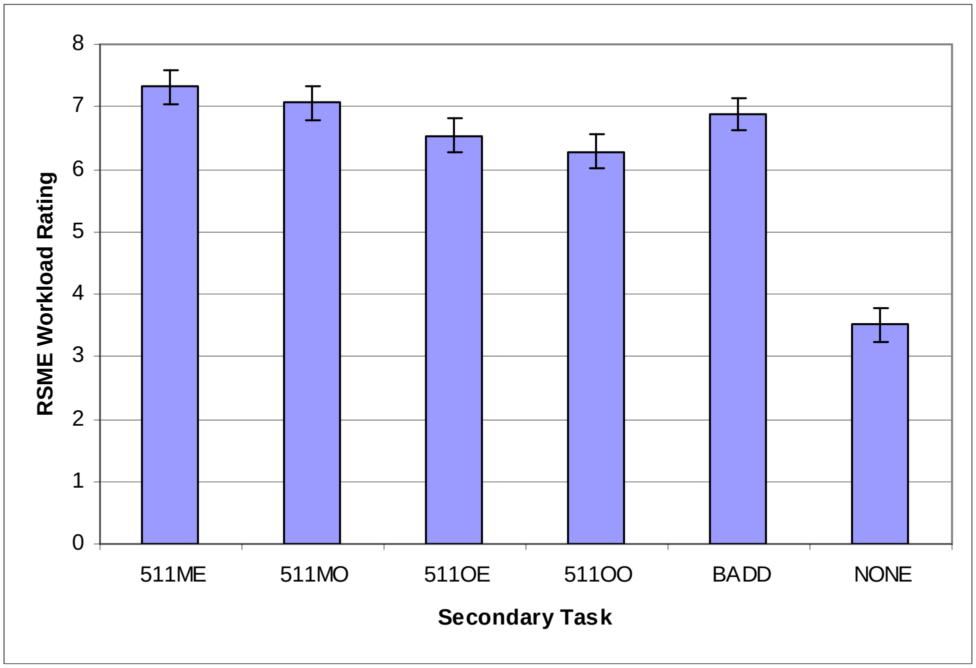


Figure 21. Mean RSME workload ratings by secondary task condition (± standard error)

Neither the main effect of age group, *F* (2, 33) = 0.49, p = .6150, nor the interaction between secondary task and age group, *F* (10, 163) = 1.21, p = .2900 was statistically significant.

4.4.2 <u>Task Durations</u>

Means and standard deviations of task durations are presented for the main three secondary task categories in Table 23.

Table 23. Mean and SD secondary task durations

Task	Mean (s)	SD
511	111.4	25.7
Baddeley	141.5	2.8
None	134.5	9.4

Differences shown in Table 23 reflect inherent differences between the secondary tasks. Specifically, the simulated phone task is a paced task over which the driver has no control. This is reflected in the small standard deviation. Similarly, the data collection interval associated with the baseline condition was determined by the length of the straight segments and the paced speed of the lead vehicle. It is evident that the phone task was performed not just during the straight segment, but also over the entire straight segment, plus approximately 7 seconds on average of curved road.

Figure 22 presents the mean and standard errors for duration of the four types of 511 tasks. Table 24 presents the results of statistical tests of differences between these means for two specific tests, namely whether trials involving a map had longer completion times than those with no map, and whether trials with 511 system errors had longer completion times than those without errors.

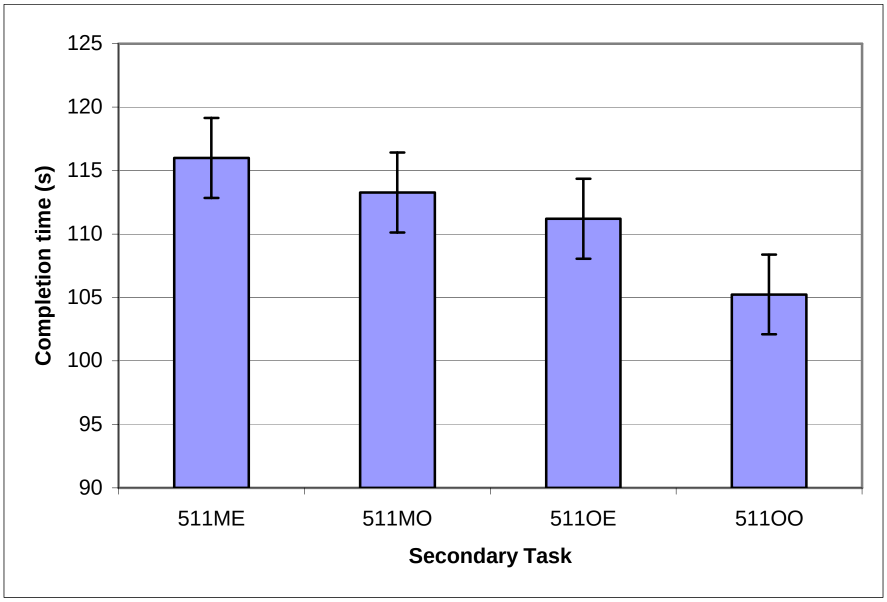


Figure 22. Mean completion times by secondary task condition (± standard error)

Table 24. Paired comparisons of completion times by category of 511 task

Comparison	DF	F value	Adjusted p-value	Interpretation
Map vs. No Map	1, 243	4.94	0.0542	Map >= No Map
Error vs. No Error	1, 243	2.26	0.1337	

The results indicate a marginally non-significant difference between the map and no-map conditions. The direction of the effect is consistent with our predictions that tasks requiring use of the map take longer to complete than those that do not require the map.

4.5 Eye Glance Analysis

We examined the confidence values associated with the eye position measures. The mean value of eye tracking system confidence for the eye gaze data was less than 40%, which means that the system was generally not confident about its conclusion about where the driver was looking. Only 10% of the trials had average confidence levels of between 60 and 80%. We therefore concluded that the analysis of eye glance data would not provide useful information. We next examined the confidence values for the head position data, since the position of the head is generally an easier inference to make. For this measure, the mean value was close to 60%, which is still below the acceptable level recommended by the system manufacturer. Additional analysis is required to identify the cause of our problems with the eye tracker.

5.0 DISCUSSION

At the outset, we presented five hypotheses that formed the basis for this experiment. In the first section below, we present an overview of the main results applicable to each hypothesis. We then summarize the effects of age group and speed input signal. In the final section, we present a broader integrated discussion incorporating the results for all measures, including their implications for the design and evaluation of voice interface systems.

5.1 Summary of Results Related to Specific Predictions

The first prediction was that the secondary tasks used in the present study would impair all aspects of driving performance, including vehicle control, car-following, and target detection. Generally, the results were consistent with this prediction. Performing any secondary task was associated with increased steering entropy, which is a composite measure of steering correction frequency and magnitude. Similarly, trials involving secondary tasks were associated with longer car-following delay, which is a measure of the driver's response time for adapting to lead vehicle speed changes. Drivers also detected proportionately fewer PDT targets and had longer response times when engaged in secondary tasks, relative to baseline trials. Finally, the subjective workload ratings were consistently higher for trials involving secondary tasks. Together, these findings indicate significant degradation of all aspects of driving performance associated with the secondary tasks.

The second prediction was based on the hypothesis that the 511 tasks were potentially more distracting than the simulated phone task. Specifically, we predicted that the performance degradation associated with the 511 tasks would be greater than with the simulated phone task. The present results were only partially consistent with this prediction. In particular, the consistently lower coherence values observed for all categories of 511 tasks indicated that these tasks interfered more with car-following accuracy than did the simulated phone task. Drivers also exhibited longer headways in three of the four 511 task conditions than in the simulated phone task condition. The relatively low correlation between coherence and headway indicates that lower coherence values were not directly related to the longer headways. However, headway was less a performance measure than an adjustment made by drivers, presumably either to compensate for the increased demands of the added secondary tasks or as an unintended result of neglecting the car-following task while performing the secondary tasks.

Two measures provided additional partial support for our second prediction. Specifically, the 511-Map trials were associated with significantly greater steering entropy than the simulated phone task. The 511-Map trials also had longer car-following delays than did the simulated phone task. Thus, there is consistent evidence that the 511-Map trials were more disruptive to driving than the simulated phone task, but only limited support for the hypothesis that the 511-No Map tasks were more detrimental to driving performance than the simulated phone task.

The third hypothesis addressed differences in distraction potential associated with one specific 511 task-interface attribute, namely whether or not a map was required to perform the task. We predicted that the interference associated with the 511 Map tasks would be greater than for the 511 No-Map tasks and that this effect would be observed for both the visual and cognitive aspects of driving. We found consistent evidence that the 511-Map trials were more disruptive than the 511-No Map trials and that the differences were apparent in all categories of driving performance. Map trials were associated with greater steering entropy, longer car-following delay, lower PDT detection rates, and slower PDT response times. In addition, participants rated the map trials as more demanding than the no-map trials. Thus, while the problem-solving aspects of the map and no-map trials were similar, the additional requirement of looking at a map and integrating information from visual and auditory sources led to significantly worse performance in all categories of performance measures.

The fourth hypotheses considered the effects of system reliability on driving performance. We predicted that low system reliability would be disruptive to concentration, leading to poorer driving performance. We found no effects of our error manipulation, suggesting that the 20% error rate used in the study was not sufficient to disrupt drivers' concentration enough to influence their driving behavior.

Finally, we predicted that system reliability would influence the time required to complete the secondary tasks. We found that the difference between the error and no-error conditions was not significant for this measure.

5.2 Effects of Age Group

Only car-following measures revealed differences attributable to age groups. Specifically, older drivers had lower coherence values, indicating more difficulty in car following. The older drivers also had longer delays, indicative of slower responses to lead vehicle speed changes, relative to the other two age groups. Modulus values for older drivers were consistently smaller, indicating more conservative responses to changes in lead vehicle speed. Finally, older drivers drove at longer headways than other drivers. As suggested above, this reflects either an attempt to compensate for the increased demands of secondary tasks, or an unintended result of neglecting the car-following task while performing the secondary tasks. None of the steering measures and none of the PDT measures revealed differences among the age groups.

5.3 Effects of Speed Input Signal

We used two speed input signals for car following, including a simple sine wave and a more complex signal composed of white noise banded over a wider frequency range. We expected the complex signal to be less predictable and thus more difficult to follow. We found some evidence to support this hypothesis. Coherence values were consistently lower for the complex signal than for the sine wave. However, we also noticed that this phenomenon appeared to reflect performance near the beginning of the trial that did not extend through the entire trial.

The standard deviation of lane position (SDLP), which measured drivers' ability to maintain lateral position, was strongly affected by the speed input signal. The complex speed signal was associated with higher levels of lane position variability than the sine wave. One possible explanation is that following the lead vehicle programmed with the complex signal involved increased visual demands, which necessitated diverting attention away from maintaining lateral control. However, sufficient visual resources remained to maintain consistent target detection performance. Together, the results suggest that the complex signal was more demanding than the simple sine wave signal. However, the absence of interactions between the speed input signal and secondary task conditions indicates that the basic effects of secondary task performance on car following were independent of the speed input signal.

5.4 Expanded Discussion of Results

Performance measures were selected in three basic categories: vehicle control, car following, and target detection. According to the proposed underlying hierarchy (Michon, 1985), driving consists of concurrent activity at three levels: (1) strategic, (2) tactical/maneuvering, and (3) operational/vehicle control. In the present study, car following represents a mid-level, tactical, or maneuvering behavior because maintaining a constant following distance requires active interpretation of visual cues and dynamic decision making to modulate vehicle speed (Brookhuis et al., 1994). In contrast, measures of steering performance represent low-level, operational, or vehicle-control behaviors, which are generally executed automatically and require minimal conscious attention. The question of where target detection fits in the control hierarchy depends on the specific task requirements, including the amount of interpretation necessary to identify targets and the complexity of the response. In this study, the targets were simple and did not require active interpretation, as there were no distractors. The required response was also simple. Thus the PDT was assumed to involve primarily visual rather than visual plus cognitive factors and would thus be more consistent with behavior at the lowest level of the control hierarchy. Accordingly, performance degradation observed in car-following measures can be interpreted as direct evidence of cognitive distraction, while degradation of vehicle control and target-detection measures represent lower-level peripheral distraction. We now consider the results for measures in each category.

5.4.1 <u>Vehicle Control</u>

Measures in this category, which included steering entropy and the standard deviation of lane position (SDLP), characterized lateral vehicle control. Steering entropy was the most sensitive performance measure in this category. It revealed consistent degradation associated with all secondary tasks. In addition, it was one of only several measures to reveal differentiation among the secondary task conditions, most notably between the 511-Map task and other secondary tasks. In contrast, SDLP was considerably less sensitive to detecting degradation due to the secondary tasks. Only two secondary task conditions revealed differences from the baseline condition. The apparently weaker effects may have been related to our use of a relatively new methodology, in which GPS positional information was related to a model of the test track to compute lane position. Additional validation of this approach is needed.

We also considered steering reversal rate and steering hold rate. These measures comprise a model of steering developed by MacDonald and Hoffman (1980), according to which steering reversal rate is a measure of driving task difficulty. However, as noted by MacDonald and Hoffman (1980), the direction of an observed effect depends on the "level of total task difficulty relative to the driver's capacity to cope with it" (p. 735). Accordingly, when driving task demands are within a driver's capacity, he or she copes with the additional demands of a concurrent secondary task by increasing total effort, which is reflected in higher steering reversal rates. However, when performing close to capacity, the driver may cope with the additional demands of a secondary task by diverting attention away from steering, which results in decreased steering reversal rates. This latter situation is also likely to be reflected by an increase in steering holds, defined as periods of steering inactivity.

In the present study, we observed increased steering reversal rates on trials involving secondary tasks together with decreasing steering hold rates. This pattern is inconsistent with MacDonald and Hoffman's model, which predicts that secondary tasks will increase both steering reversal rates and steering hold rates. The present results for steering reversal rates are generally consistent with this model; however, the pattern of steering holds indicates higher rates in the baseline condition without a secondary task.

In an attempt to better understand the effects of the secondary tasks on steering behavior, we considered a fourth measure, the proportion of each trial involving no steering activity. We thought this measure might be a more precise indicator of steering inactivity because it includes information about frequency and duration, whereas steering hold rate utilizes the frequency of episodes of at least 400 ms, but otherwise does not consider duration. The results for this measure were stronger than for steering hold rate, however they revealed the same pattern of increased steering inactivity during the baseline trials relative to the secondary task trials. One interpretation of these results is that control inputs made during baseline driving were more accurate and thus increased the time until the next correction was required. In contrast, during secondary task performance, the decreased precision of steering inputs (evidenced as increased steering entropy) may have necessitated more frequent corrective inputs, resulting in more reversals and fewer holds.

We also examined the correlations among the various measures of steering behavior. We found the correlations between steering entropy and the other three steering measures to be relatively small (all less than r = .30), suggesting that steering entropy characterizes different aspects of steering behavior than reversal or hold rates. This may reflect the fact that, unlike the reversal and hold rates, steering entropy is computed by combining information from both the secondary task and baseline trials. As such, steering entropy may be susceptible to contamination by differences in transient influences that may not be consistent across baseline and secondary task trials. However, the robustness of results indicates that this was not a problem in the present study.

5.4.2 Car Following Measures

Based on the experience of Brookhuis (Brookhuis et al., 1994), we expected that drivers would generally be able to follow the lead vehicle adequately while performing a secondary task and that distraction effects would be revealed as increased values of delay (phase shift), reflecting slower responses to lead vehicle speed changes while engaged in secondary tasks. Our results were not entirely consistent with this model in that reduced coherence and increased delay occurred together in many of our comparisons. Results of an earlier study may help explain the nature of these effects. In that study, the pattern of coherence and delay effects differed for different groups of drivers (Ranney et al., 2004a). Specifically, among high-performing drivers, coherence remained consistent across secondary task conditions while delay increased as secondary task demands increased. However, low-performing drivers increased their following distances while performing secondary tasks, which reduced coherence to the point that delay was not interpretable. A similar pattern was observed in the present study among older drivers, who exhibited increased following distances and reduced coherence values, although not to the point that delay was uninterpretable. These patterns suggest a two-level model of performance impairment. In the first level, which represents relatively minor degradation, drivers are generally able to follow lead-vehicle speed changes accurately and there is no significant drop in coherence. At this level, performance impairment is revealed as increased delay, reflecting slower responses to lead vehicle speed changes. However, as impairment becomes more severe, drivers become less able to maintain their car following and as a result the coherence drops. The drop in coherence is due in part to increased headways, which drivers may exhibit while performing secondary tasks. Thus according to this model, significant decreases in coherence reflect more severe performance impairment than increased delay. In the present study, the consistently higher coherence values observed during the simulated phone trials provided the strongest evidence that the 511 tasks were more disruptive than the simulated phone task.

This two-level model of car following performance degradation allows a comparison between the present results and those of the earlier study (Ranney et al., 2004a). If we assume that the participants in the present study are generally equivalent to the high-performing drivers in the earlier study, we may conclude that the secondary tasks performed in the present study were generally more demanding and thus more distracting than those used in the earlier study, based on the degraded coherence that was present in this study but not in the earlier one. This is noteworthy because the secondary tasks used in the earlier study all had at least some manual inputs while those in the present study were performed without any manual inputs. However, the validity of this comparison is tempered by the fact that the car following task in the present study was slightly more demanding than that used in the earlier study, due primarily to faster speeds used in the present study.

Modulus (gain) reflects the following driver's responses at the extreme values of the lead vehicle speed. Specifically, modulus values near 1.0 indicate that the following driver is closely matching the extreme speed values of the lead vehicle. Modulus values significantly greater than 1.0 indicate higher potentially aggressive overcorrection and values smaller than 1.0 indicate undercorrection. Values at either extreme reflect increased potential for safety problems resulting from inadequate following distances. Given this model, the modulus results of the present study are somewhat ambiguous. Modulus values associated with the secondary tasks were consistently lower than for baseline trials, suggesting more conservative responses while engaged in secondary tasks. However, the mean values for the secondary task trials were generally closer to 1.0, which would indicate more accurate following during secondary task trials and overly aggressive following on baseline trials. This pattern suggests that drivers were trying too hard to respond to lead vehicle changes in the baseline condition, however if so, this would also reflect the increased availability of attentional resources in the baseline condition.

5.4.3 Visual Target Detection

Two patterns characterize the target detection results. First, we found lower proportions of targets detected and slower response times for all secondary tasks, relative to the baseline condition. This finding is significant because two of three secondary task conditions had no visual demands. The results therefore support the conclusion that even in the absence of visual demands, secondary tasks with significant cognitive demands can divert attention away from driving, resulting in degraded visual target detection performance (e.g. Strayer and Johnston, 2001). Second, we found consistent differences between the 511-Map and 511-No Map conditions. Specifically, the map condition was associated with lower proportions of targets detected and slower response times, relative to the no-map condition. This finding reflects the peripheral interference associated with the structural conflict between the visual demands of the target detection task and the 511 map task.

We saw above that the 511 map task was associated with significantly higher levels of steering entropy relative to the simulated phone task. We expected to find a similar pattern for PDT performance, reflecting differences in visual demand between the map task and the simulated phone task. However, this pattern was not apparent in our data. Specifically, there were no differences between the simulated phone task and any category of 511 task for the percentage of PDT targets detected. Moreover, the finding that PDT response times were slower for the simulated phone task than for the no-map 511 tasks is contrary to our predictions. There are several possible explanations for this pattern of results. First, while steering entropy reflects all (continuous) steering activity, the PDT measures summarize responses to discrete events presented intermittently. The PDT may therefore be less sensitive than steering entropy for detecting performance impairment that occurs continuously during secondary task performance. Second, differences in pacing between the 511 and simulated phone tasks may have influenced this result. Because the simulated phone task was externally paced, drivers had little control over when their attention was required to interpret sentences or to recall words in sentences. In contrast, the self-paced nature of the 511 tasks may have allowed drivers to perform the secondary tasks in chunks that allowed more efficient monitoring of the target detection area, which led to faster detection. Furthermore, under most conditions target detection could be accomplished peripherally. Additional research, in which pacing and task complexity are manipulated together, will be necessary to determine how these factors interact to modulate distraction effects on visual target detection.

5.4.4 Secondary Task Demand

In a previous study (Ranney et al., 2004a), secondary tasks performed using visual/manual interfaces were more disruptive to driving than identical tasks performed with voice interfaces. The voice interface helped reduce the distracting effects of secondary task performance for measures of vehicle control and target detection, but not for car-following measures. We concluded that the voice interface helped reduce the peripheral (visual and manual) interference, but not the attentional (cognitive) interference. In the present study, all secondary tasks were performed using a voice interface. The consistent pattern of performance decrements observed across all categories of driving performance measures reinforces the earlier conclusion that voice interfaces are not sufficient to eliminate cognitive distraction due to performing secondary tasks while driving.

The present findings also allow us to revisit our interpretation of PDT performance degradation. In the earlier study, we were able to maintain consistent cognitive demands between the two interface conditions, thus isolating the effect of interface. This allowed us to interpret differences in PDT performance between the two conditions as evidence of the structural conflict between the visual demands of secondary tasks performed with the visual/manual interface and those of the PDT. However, the present results are not entirely consistent with this interpretation. Specifically, we found differences in PDT performance between the simulated phone task and the 511-No Map condition, which both required use of the same voice interface. In the absence of visual demand differences between these secondary tasks, this result suggests that differences in PDT performance can also reflect differences in cognitive demands between secondary task conditions. This interpretation is consistent with Strayer and Johnston's (2001) finding that the attentional demands of a phone conversation served to divert drivers' attention from target detection in driving to the phone conversation. Thus, while the PDT task demands are primarily visual, performance degradation may reflect a combination of visual and attentional components.

Our inclusion of a simulated phone task was intended to provide a benchmark to answer the question of whether the navigation system tasks were more disruptive than phone conversation. Unfortunately, it is very difficult to simulate typical phone conversations in experimental settings. To ensure consistent demands across subjects, we chose the Baddeley task, which has been used previously (Alm & Nilsson, 1994; Ranney et al., 2004b) and is generally thought to have demands similar to a moderately intensive phone conversation (Briem & Hedman, 1995). As indicated above, while there were some differences between the simulated phone task and the 511 tasks, we did not find the consistent differences that had been predicted. However, to the extent that our simulated phone conversation was consistently more demanding than a typical phone conversation, we may extrapolate to conclude that the 511 tasks used in this study were more demanding and thus more distracting than a typical phone conversation. We advocate continued efforts to examine naturalistic data to quantify the demands of actual on-road phone conversations.

5.4.5 Implications for Voice Interface Design

Three specific findings have implications for the design of systems that require navigation of hierarchical menu systems. First, based on the consistent differences between the map and no-map conditions, we recommend avoiding incidental or redundant visual displays. Looking away from the road not only impaired target detection but also vehicle control.

Second, the absence of strong and consistent differences between the 511 tasks and the simulated phone task suggests that it may be possible to design information acquisition tasks involving navigation of a hierarchical menu structure without imposing a greater load than that associated with a moderately intense phone conversation. However, while the phone task used in this study was more demanding than typical phone conversations, the simulated 511 system was probably simpler than most existing systems. Therefore, real-world traveler information systems or other systems that utilize hierarchical menu structures are likely to induce greater interference with driving than most routine phone conversations. Designers should be aware of this potential problem and strive to simplify information systems as much as possible.

Our third finding was that there were no differences between error and no error conditions of the 511 task. This suggests that drivers can tolerate less than perfect reliability in voice interfaces without a significant breakdown of driving performance. However, additional research that includes higher levels of errors than used in the present study would be useful to determine a threshold for acceptable system reliability.

Finally, the present results provide indirect support for the idea that self-paced tasks are generally less disruptive to driving than ones of similar difficulty that are externally paced. This implies that message length may influence the driver's level of distraction. Specifically, drivers' freedom to choose when to interact with an in-vehicle information system is partially abdicated once a message is initiated. While drivers may be aware that they can repeat any message, the longer the message and thus the more time they have invested in listening, the less likely they may be to restart a message. Longer messages thus require more concentrated attention and greater short-term memory load than shorter messages. Designers should therefore consider limiting the length of messages in automated systems. They may also want to consider implementing a "pause" function so that drivers can interrupt a long message without having to restart from the beginning.

6.0 CONCLUSIONS

Based on the results of the present study, we conclude that:

1. Performing secondary tasks that require active navigation of hierarchical menu structures, as in a 511 traveler information system, caused significant deterioration of driving performance, even when performed with a hands-free voice interface. The deterioration affected all aspects of driving, including vehicle control, decision making in car following, and visual target detection.
2. Performing secondary tasks that simulate moderately intense hands-free phone conversation also caused significant deterioration of driving performance, although the effects were not as strong as for the 511 system tasks.
3. The simulated phone task used in this study was considered to be more demanding than typical phone conversation. Therefore, our results suggest that use of 511 traveler information systems will likely be more disruptive to driving than typical phone conversation.
4. Differences among age groups in the present study were minimal. They were limited to car-following measures and are consistent with increasing cognitive impairment associated with aging.
5. Secondary tasks that required the use of a visual interface (map) were consistently more disruptive to driving performance than similar tasks that did not require a visual interface. This reflects both the requirement to look away from driving as well as the increased cognitive demands associated with the map tasks.
6. The present results are consistent with earlier results in supporting the conclusion that voice interfaces alone do not eliminate distraction due to secondary tasks like those used in the present study.
7. The absence of differences between 511-task error conditions suggests that drivers can tolerate moderate levels of voice recognition error without suffering significant degradation of driving performance or impacting completion times. Additional research is needed to identify the threshold of acceptable system performance.
8. Differences between 511-task map and no-map conditions suggest that designers should attempt to ensure that secondary tasks like those simulated in this study do not require an additional burden of visual aids, unless essential to task completion.

7.0 REFERENCES

511 Deployment Coalition (2003). Implementation and operational guidelines for 511 services (Rep. No. Version 2.0). U.S. Department of Transportation, 511Deployment Coalition.

Alm, H. & Nilsson, L (1994). Changes in driver behaviour as a function of handsfree mobile telephones. A simulator study, Accident Analysis and Prevention, 26, 441-451.

Baddeley, A., Logie, R., Nimmo-Smith, I., & Brereton, N. (1985). Components of Fluid Reading. Journal of Memory and Language, 24, 119-131.

Barickman, F. S. & Goodman, M. J. (1999). Micro DAS: In-vehicle portable data acquisition system. Transportation Research Record, 1689, 1-8.

Boer, E. R. (2000). Behavioral entropy as an index of workload. In Proceedings of the IEA 2000/HFES 2000 Congress (pp. 3-125-3-128).

Briem, V. & Hedman, L.R. (1995). Behavioural effects of mobile telephone use during simulated driving. Ergonomics, 38, 2536-2562.

Brookhuis, K. A., De Vries, G., & De Waard, D. (1991). The effects of mobile telephoning on driving performance. Accident Analysis and Prevention, 23, 309-316.

Brookhuis, K. A., Waard, D. d., & Mulder, B. (1994). Measuring driving performance by car-following in traffic. Ergonomics, 37, 427-434.

Casali, S. P., Williges, B. H., & Dryden, R. D. (1990). Effects of recognition accuracy and vocabulary size of a speech recognition system on task performance and user acceptance. Human Factors, 32, 183-196.

Federal Highway Administration (2007). U.S. Map for 511 Deployment Activities as of December 31, 2006. Retrieved March 1, 2007, from http://www.fhwa.dot.gov/trafficinfo/511.htm

Harms, L. & Patten, C. (2003). Peripheral detection as a measure of driver distraction. A study of memory-based versus system-based navigation in a built-up area. Transportation Research Part F, 6, 23-36.

Jacko, J. A. & Salvendy, G. (1996). Hierarchical menu design: Breadth, depth, and task complexity. Perceptual and Motor Skills, 82, 1187-1201.

Jahn, G., Oehme, A., Krems, J.F., & Gelau, C. (2005). Peripheral detection as a workload measure in driving: Effects of traffic complexity and route guidance system use in a driving study. Transportation Research Part F, 8, 255-275.

MacDonald, W. A. & Hoffman, E. R. (1980). Review of relationships between steering wheel reversal rate and driving task demand. Human Factors, 22, 733-739.

Michon, J. A. (1985). A critical view of driver behavior models. What do we know, what should we do? In L.E.Evans & R. Schwing (Eds.), Human behavior and traffic safety (New York: Plenum Press.

Myers, J. L. & Well, A. D. (1991). Research Design & Statistical Analysis. (1 ed.) New York: HarperCollins Publishers.

Ranney, T. A., Harbluk, J. L., Smith, L., Huener, K., Parmer, E., & Barickman, F. (2004a). The effects of voice technology on test track driving performance: Implications for driver distraction (Rep. No. DOT HS 809 525). U.S. Department of Transportation, National Highway Traffic Safety Administration.

Ranney, T. A., Watson, G. S., Mazzae, E. N., Papelis, Y., Ahmad, O., & Wightman, J. R. (2004b). Examination of the Distraction Effects of Wireless Phone Interfaces Using the National Advanced Driving Simulator - Final Report on a Freeway Study (Rep. No. DOT HS 809 787). U.S.Department of Transportation, National Highway Traffic Safety Administration.

Rumar, K. (1990). The basic driver error: Late detection. Ergonomics, 33, 1281-1290.

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

Westfall, P. H., Tobias, R. D., Rom, D., Wolfinger, R. D., & Hochberg, Y. (2003). Multiple Comparisons and Multiple Tests Using SAS. Cary, North Carolina: SAS Institute Inc.

8.0 APPENDICES

8.1 <u>Appendix A: Participant Information Summary and Confidential Information</u>

PARTICIPANT INFORMATION SUMMARY

Project Title: The Effects of Voice Technology on Test Track Driving Performance

<u>Investigators</u>: The investigators for this study are listed below. If you have questions at any time regarding this study please contact the investigator(s) at the address and/or telephone number given below:

Thomas Ranney, Ph.D. and Elizabeth Mazzae, MSE
NHTSA Vehicle Research and Test Center
10820 SR 347
East Liberty, OH 43319
Phone: (800) 262-8309 or (937) 666-4511

<u>Study Description:</u> Newer in-vehicle technologies now offer a wider range of functions, including navigational information, retrieval of email messages, or Internet access. In response to concerns that these systems may interfere with driving, manufacturers have developed hands-free interfaces with voice-recognition capabilities. The experiment in which you will participate is being conducted by the United States Government's National Highway Traffic Safety Administration (NHTSA) to determine how well drivers are able to perform in-vehicle tasks using a hands-free phone to interact with voice-recognition-capable systems while driving.

We are inviting you to participate in this research study because you are 18-25, 30-45 or 50-60 years old, have a valid, unrestricted U.S. driver's license (except for corrective eyeglasses and contact lenses), have a minimum of two years driving experience, drive at least 10,000 miles per year, are in good general health, and are a frequent cell phone user, including using your phone while driving. A total of 36 people will participant in the study.

<u>Driving Requirements and Procedures:</u> Your participation in this study will consist of one session lasting approximately 4-5 hours. During the session, you will be given instructions, training, and practice on both the driving tasks and the specific in-vehicle tasks used in the study. You will drive a number of test trials and be asked to answer brief questions about the tasks performed. After completing the test trials the session will end.

All driving will be conducted on a 7.5 mile oval test track at the Transportation Research Center Inc. (TRC). The track is a limited access facility with 4 lanes of traffic, all flowing in the same direction. You will be instructed as to which lane you should drive in and should only change lanes when asked to do so.

The driving task you will be asked to perform will involve driving at normal highway speeds while following a lead vehicle (red sedan) around the track. You should do your best to match your vehicle's speed to the speed of that vehicle and you will occasionally be asked to make stops on the innermost lane of the track. The in-vehicle tasks you will be asked to perform will involve using a wireless phone to complete navigation and conversation tasks while driving. You will complete a specified number of test trials involving the performance of these in-vehicle tasks while driving. After each trial, you will be asked to answer some brief questions about the difficulty of the task performed.

The vehicle which you will be asked to drive is a Honda Accord. The vehicle is instrumented with a MicroDAS vehicle data acquisition system. The MicroDAS contains sensors that measure certain aspects of vehicle operation, vehicle motion, and driver actions. The system also contains video cameras that capture images of driver actions and the environment in which the vehicle is being driven (e.g., driver's hand position on the steering wheel, forward road scene). Additional cameras monitor the driver's eye glance behavior. These sensors and video cameras are located in such a manner that they will not affect your driving, the vehicle's performance, or obstruct your view while driving. The information collected using these sensors and video cameras is recorded onto data storage media for subsequent analysis.
You will be given indoctrination in the rules and procedures before driving on the TRC test track. An experimenter will accompany you at all times during the experiment. The experimenter will give specific instructions about when to begin performing the driving and in-vehicle tasks. Because there will be other traffic on the test track, please make sure that you believe that it is safe to perform any tasks before beginning them. The in-vehicle experimenter will have access to a secondary brake mechanism that can be used to bring the vehicle safely to a stop if needed. However, the experimenter will not be able to ensure complete safety. Therefore, it is very important to always remember that you, as the driver, are in control of the vehicle and you must be the final judge on when or whether to respond to any request made by the experimenter. You should follow a request or complete a maneuver only when, in your judgment, it is safe and convenient to do so. Be aware that crashes can happen at any time when driving. You are required to wear a seat belt at all times while driving.

To allow for the most accurate recording of your eye movements while driving, your entire face must be clearly visible while driving. If your hair hangs in your face, you may be asked to use clips or a rubber band to keep it out of your face. If you require corrective lenses and have contact lenses, we would prefer that you wear them rather than glasses. You will not be permitted to wear sunglasses while driving. In addition, to help the eye tracking system better identify and track your facial features, you will be required to wear several small stickers on your face. The stickers will be put on before you begin driving and cannot be removed or moved until the experimenter informs you that you are finished driving. As a result you may be wearing the stickers for up to 3 hours.

We do not anticipate that any changes to procedures will take place during this study. However, any new information developed during the course of the research that may affect a subject's willingness to participate will be provided to you.

Risks: During your participation in this study, you will be subject to all risks and uncertainties normally associated with driving on the TRC test track, plus any additional risks associated with completing in-vehicle wireless phone tasks while driving. Generally, this will entail light traffic, including a mix of passenger cars and large trucks, all traveling in the same direction at various speeds and with occasional stops in the inside lanes. As a result of the controlled nature of this facility and lack of opposing traffic, the risk of a crash is minimal. While driving you will not be asked to perform any unsafe driving acts.

You will be asked to wear several small, latex stickers on your face while driving. These stickers may cause skin irritation in people with an allergy to latex.

There are no known physical or psychological risks associated with participation in this study beyond those described above.

Benefits: The experiment will provide data on driver behavior and in-vehicle task performance that will be used by researchers to provide a scientific basis for developing recommendations or standards for performing in-vehicle tasks while driving. The compensation that you will receive (see Costs/Compensation section below) is the only direct benefit from participation in the study.

Conditions of Participation, Withdrawal, and Termination: Participation in this research is voluntary. By agreeing to participate, you agree to operate the research vehicle in accordance with all instructions provided by NHTSA and TRC staff. If you fail to follow instructions, or if you behave in a dangerous manner, you may be terminated from the study. At any time you may withdraw your consent and discontinue participation in the study at any time without penalty (you will receive compensation commensurate with the length of your participation).

Costs/Compensation: You will receive a rate of pay of $20 per hour for the time you spend at the TRC facility. In addition, you will have the opportunity to earn incentive pay based on your performance on the driving and in-vehicle tasks. The maximum possible amount of incentive pay is $18. Other than the time you contribute, there will be no costs to you.

Insurance: The contractor assisting with the conduct of this study, TRC, in whose facilities the event will be conducted, will maintain insurance that will cover you in the event of a crash. This insurance will provide coverage for injuries to yourself up to a limit of $10,000.00. Coverage will also be provided for injuries to others, including the driver and any passengers of other vehicles involved in the crash, as well as damages resulting from any crashes occurring during your participation in this study, up to a $1,000,000 limit. Except to the extent covered by such insurance policy, neither the TRC nor NHTSA will be responsible for your actions during this study nor will they indemnify you or otherwise compensate you for any problems arising out of your actions or the normal risks associated with driving. However, you will not be liable for loss or damage to the MicroDAS equipment, the research vehicle, or other equipment during your participation unless there is gross negligence on your part.

Use of Information Collected: In the course of this study certain **NHTSA engineering data** and **NHTSA video data** (video image data recorded by NHTSA) will be collected.

> Information NHTSA may release:
> The **NHTSA engineering data** collected and recorded in this study (including any performance scores based on these data) will be analyzed along with data gathered from other participants. NHTSA may publicly release this data in final reports or other publications or media for scientific, educational, research, or outreach purposes.
>
> The **NHTSA video data** recorded in this study includes your video-recorded likeness and all in-vehicle audio including your voice (and may include, in some views, superimposed information regarding your driving performance). Video and in-vehicle audio will be used to examine your driving performance and other task performance while driving. NHTSA may publicly release video image data (in continuous video or still formats) and associated audio data, either separately or in association with the appropriate engineering data for scientific, educational, research, or outreach purposes.
>
> Information NHTSA may not release:
> Any release of **NHTSA engineering data** or **NHTSA video data** shall not include release of your name. However, in the event of court action, NHTSA may not be able to prevent release of your name or other personal identifying information. NHTSA will not release any information collected regarding your health and driving record.

Informed Consent: By signing the informed consent statement contained in this document, you agree that participation is voluntary and you understand and accept all terms of this agreement. Also by signing the informed consent statement, you agree to operate the research vehicle in accordance with all instructions provided by NHTSA and TRC staff. You may withdraw your consent and discontinue participation in the study at any time without penalty (you will receive compensation commensurate with the length of your participation, as described previously).

Disposition of Informed Consent: NHTSA will retain a signed copy of this Informed Consent form. A copy of this form will also be provided to you at your request.

INFORMED CONSENT STATEMENT:

I certify that I have a valid, U.S. driver's license and that all personal and vehicle information as well as information regarding my normal daily driving habits provided by me to NHTSA, and TRC employees associated with this project during the pre-participation phone interview and the introductory briefing was true and accurate to the best of my knowledge.

I certify that I have been informed about the study in which I am about to participate. I have been told how much time and compensation is involved. I understand that the purpose of this study is to assess driving and in-vehicle task performance. I agree to operate the research vehicle in accordance with all instructions provided to me by NHTSA and TRC staff. It has been explained to me that the study will be conducted on a controlled track and that the risk of a crash is minimal.

I understand and agree that for scientific, educational, research, or outreach purposes, video images of my drive which will contain views of my face and accompanying audio data may be used or disclosed by NHTSA in perpetuity, but my name and any health data or driving record information will not be used or disclosed by NHTSA.

I have been given adequate time to read the attached summary. I understand that I have the right to ask questions at any time and that I can contact the principal investigators at (937) 666-4511 for information about the study and my rights.

I understand that my participation is voluntary and that I may refuse to participate or withdraw my consent and stop taking part at any time without penalty or loss of benefits to which I may be entitled. I hereby consent to take part in this project.

I, ________________________, VOLUNTARILY CONSENT TO PARTICIPATE.

______________________________ ____________

Signature Date

<u>Information Disclosure</u>: By signing the <u>information disclosure statement</u> contained in this document, you agree that NHTSA and its authorized contractors and agents will have the right to use the NHTSA engineering data and the NHTSA video data for scientific, educational, research, or outreach purposes, in perpetuity, including dissemination or publication of your likeness in video or still photo format, but that neither NHTSA nor its authorized contractors or agents shall release your name; and you understand that, in the event of court action, NHTSA may not be able to prevent release of your name or other personal identifying information. NHTSA will not release any information collected regarding your health and driving record, either by questionnaire or medical examination.

INFORMATION DISCLOSURE STATEMENT:

I, _________________________, grant permission, in perpetuity, to the National Highway Traffic Safety Administration (NHTSA) to use, publish, or otherwise disseminate NHTSA engineering data and NHTSA video image data, as defined in the accompanying Participant Information Summary (including continuous video and still photo formats derived from the video recording), and associated in-vehicle audio data collected about me in this study, either separately or in association with the appropriate engineering data for scientific, educational, research, or outreach purposes. I understand that such use may involve widespread distribution to the public and may involve dissemination of my likeness in video or still photo formats, but will not result in release of my name or other identifying personal information by NHTSA or its authorized contractors or agents.

______________________________ ____________
Signature Date

Transportation Research Center Inc.
POLICY & PROCEDURE

CONFIDENTIAL INFORMATION		P&P NO.	153
Volume:	I, General Information	Issue Date:	10/20/03
Function:	Security	Effective Date:	10/20/03
Replaces:	Safeguarding Proprietary Info Issued 05/07/01	Code:	B, D

1. Purpose

To establish standards for the protection of confidential information and a proprietary atmosphere for TRC Inc. and its customers.

2. Scope

This policy applies to all customers and other visitors who have access to testing or other confidential information.

3. Policy

It is the policy of TRC Inc. to protect the identity, objectives, and presence of our customers, their test results, and/or other confidential information by the enforcement of the rules that are outlined herein. These rules are applicable to all personnel at/or within the facilities of TRC Inc.

3.1 You will not be allowed to witness any test or access other confidential information that you are not directly associated with unless prior approval has been given by facility management. This same restriction applies to the photographing of any test or test article.

3.2 In any activity that you are not directly associated with that you do witness, you agree not to disclose any information that you may have obtained.

3.3 Any violation of this policy may result in censure by TRC Inc. and possible punitive legal action through the courts.

I have read and understand the above P&P #153, Confidential Information, and accept my responsibilities in complying with this policy.

Printed Name ______________________ Signature ______________________

Company Name ______________________

Witness Signature ______________________ Date ______________________

8.2 Appendix B. Instrumented Vehicle Orientation

Instrumented Vehicle

This is the vehicle you will be driving today. It is a 1996 Honda Accord that has been modified to collect driving performance data. At this time, please go ahead and get in the driver's seat. I will then tell you more about the vehicle.

First of all, please adjust the seat, mirrors and steering wheel to your comfort level.

The seat controls are on the lower left side of the seat. The side mirror controls are on the door panel, the center mirror can be moved manually without harming our equipment and the steering wheel adjustment lever is on the lower left side of the steering column.

As you can see the vehicle contains many of the standard vehicle features available to most cars, as well as, some cameras and other equipment for this study.

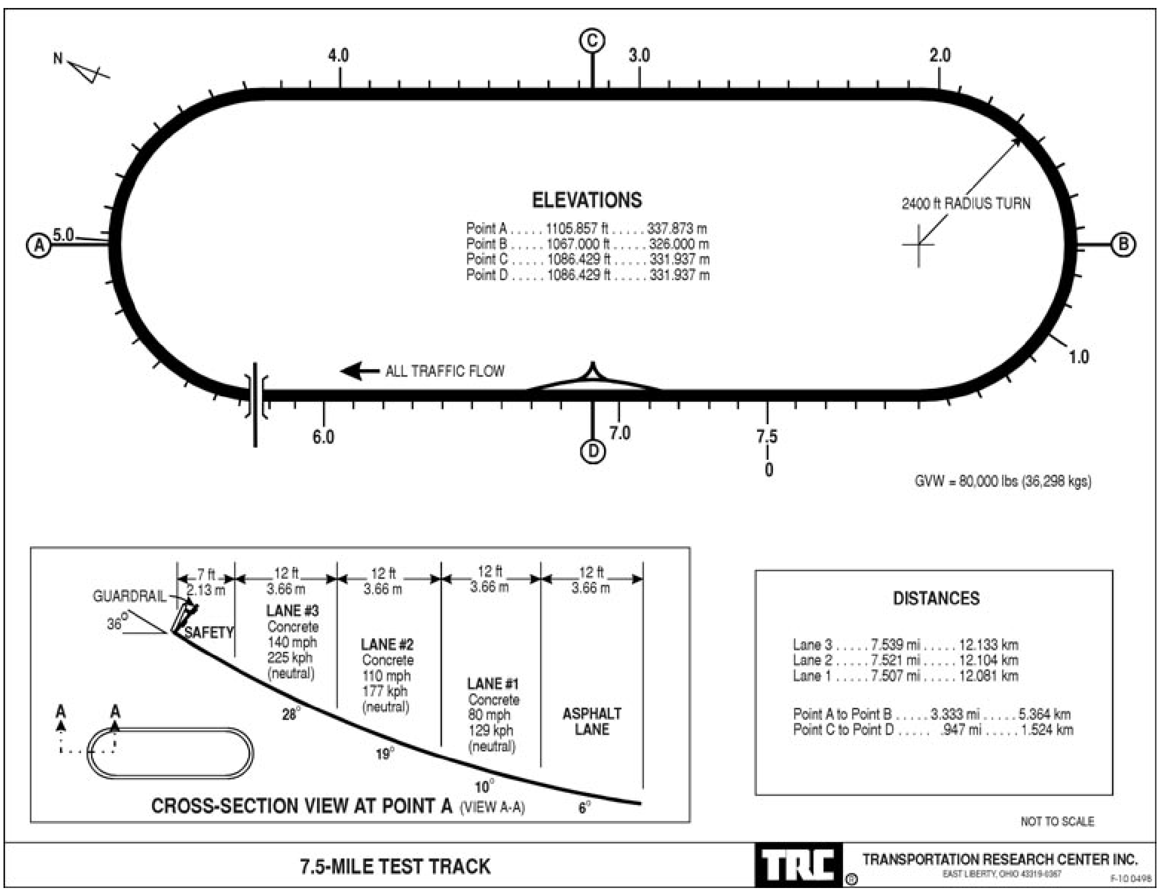
Point Out:
Heat, A/C controls (we ask that you not turn the fan up past #2 during the study)
PDT, light detection task
Eye Tracker cameras
511 Map
Wireless phone with cradle
Headrest speaker and microphone
Hazard light button
As the experimenter, I will be in the back seat when you are driving.

Do you have any questions about the vehicle?

Ok, if you are comfortable and ready to begin training, we would like to start with some task instructions and equipment calibration procedures. After instructions and training are complete, you will be given another opportunity for a break before we start the actual driving task.

8.3 Appendix C: Test Track Guidelines

Here is a picture and diagram of TRC's 7.5-mile test track that you will be driving on today. As you can see, there are four lanes of traffic traveling in a clockwise direction. The innermost lane is a blacktop lane, then lane 1, lane 2, and lane 3.



7.5-Mile Test Track Rules that Vehicles will be following

Here are some of the rules that vehicles follow while driving on the track. These rules are presented to help you understand some of the vehicle actions you may see today while driving.

- Stay in the lane farthest to the right, as your program will allow. Do not run in the 2nd or 3rd lane unless scheduled or to pass. (For your participation, the in-vehicle experimenter will present the speed and lane information that you should follow.)
- Travel in a clockwise direction at all times.
- Utilize flashers during decelerations and stops.
- All stops are to be made in the inner asphalt (blacktop) lane. Vehicles with a gross weight of 8,000 pounds or more will make their stops in the 1st concrete lane. Stops in the 2nd and 3rd concrete lanes are not permitted.
- For mechanical breakdowns, stop in the inner asphalt (blacktop) lane, turn on flashers, and contact the traffic control coordinator.
- Proceed to the inner asphalt (blacktop) lane when being approached by a faster vehicle, curves only, if your speed is 60 mph or below and the vehicle weight is less than 8,000 lbs. It is not necessary to do so in the straight-aways unless traffic becomes congested.
- Passing on the right is only permitted for vehicles running in the inner asphalt (blacktop) lane.
- Controlled braking, without lockup, should be used to avoid an animal on the track. In addition, never swerve in this situation, as a higher speed vehicle may be beside you or behind you, which could lead to a more serious accident.
- Safety is the #1 priority!

There will be construction on the test track today during the experiment. I will show you the location when we first drive onto the test track. Our experiment has been designed so that you should be done with your phone task when we approach the work zone on most laps. However, it is possible that on some laps you will not have completed the task. When we are approximately a half mile before the beginning of construction if you are still working on a task, I will notify you to stop the task and begin to maintain a speed of 50 miles per hour while merging with other traffic to drive past the construction. You should follow the lead vehicle as it changes lanes both before and after the construction. After the work zone, you will not return to the previous task. Rather, I will give you your next task instructions and tell you when to begin.

8.4 Appendix D: Driving Task (Car Following) Instructions

At this time, I would like to explain the driving and car following tasks to you. At all times while driving, you will be behind a lead vehicle (red Chevrolet Lumina).

At certain times while you are driving on the test track, the in-vehicle experimenter will instruct you to begin following the lead vehicle. When this occurs, you should adopt a close but safe following distance and try to maintain that following distance until you are instructed to stop following the lead vehicle. (Note that you will always be behind the lead vehicle but you won't always be performing the close following task.) The term "close" following distance means that we want you to drive as close as you would in moderately congested freeway driving. We ask that you select the shortest following distance at which you are comfortable driving at highway speeds. If the lead car speeds up or slows down, you should also speed up or slow down to maintain the same following distance.

We ask that you try to adopt the same close following distance each time you are instructed to follow the lead vehicle. The experimenter will tell you when to start and when to stop following the lead vehicle. When not following the lead vehicle, you should try to keep your speed near 55 miles per hour, if traffic permits. Your following distance will be monitored continually and your ability to maintain the same following distance will be one of the main measures used to evaluate your driving performance.

While driving on the test track, you should also try to keep the vehicle within the lane at all times. This is important, as there may be other traffic present while you are driving. You should only change lanes when the lead vehicle changes lanes or when instructed to do so by the experimenter who accompanies you in the vehicle.

8.5 Appendix E: Phone Task Intro (with Phone Use Instructions)

You will be asked to perform several wireless phone tasks while driving on the test track. These tasks will require you to make phone calls.

For one type of phone task, a conversation task, you will be determining whether or not a series of sentences make sense, and recalling certain components of those sentences.

In the second type of phone task, you will be obtaining information from an automated phone system. The automated phone system is a "traveler information system called 511," which has information about traffic delays and road conditions. The information you obtain from the traveler information system will help you to answer questions that the experimenter will present to you periodically during the drive.

Phone Use Instructions

At this time, I would like to go over the details of how to make phone calls during your drive today. As you can see, the phone is situated in a cradle. We would like you to place calls today using speed dialing and a hands-free mode (using the headrest speaker and microphone), allowing the phone to stay in the cradle at all times.

As stated, you will be performing two types of phone tasks. For the conversation tasks, you will be using the preset speed dial number 3. For the traveler information system tasks, you will use speed dial number 2.

(Trainer: use finger to point to buttons) To place a call in this mode, you must first choose your speed dialing number by pressing the number 2 (or 3) and then the 'talk' button. You will then hear the phone dialing. You do not need to pause between button presses. Just press them one after the other in sequence. If the system does not dial, press the 'end' button and then try the dialing procedure again.

To end a phone call, press the 'end' button. The in-vehicle experimenter will prompt you when to end phone calls after each phone task.

Phone Call Practice:

Do you have any questions about the phone dialing task? OK. Go ahead and try calling the traveler information system, speed dial number 2. Once we verify it is dialing, you can end the call by pressing the 'end' button.

Now, I would like to play an audio file that will explain in detail the phone conversation task. Then, we will practice the task. (Play Track 2 on CD)

8.6 Appendix F: Conversation (Baddeley) Task Instructions and Practice

This task will simulate a phone conversation. You will perform this task using a portable phone. For this task, you will listen to a number of sentences and answer specific questions about the sentences. Each sentence will have three parts, including: a subject, a verb, and, an object. For example, if you hear the sentence:

The boy hit the ball.
"boy" is the subject, "hit" is the verb, and " ball" is the object.

In the following sentence please identify the subject, the verb, and the object:

The frog ate the fly (PARTICIPANT RESPONDS VERBALLY HERE).

Your task will have two parts. First, you will be asked to determine whether the sentence makes sense or not. In this context, "makes sense" means the action expressed in the sentence could happen. The examples presented previously make sense because a boy could hit a ball, and, a frog could eat a fly. An example of a nonsensical sentence or one that does not make sense, is:

The dog ate the noise.
This sentence is nonsensical because it cannot happen.

Immediately after you hear a sentence, you should try to decide if it makes sense and respond as quickly as possible. If the sentence makes sense, you will say, "YES". If the sentence does not make sense, you will say, "NO". You will have a limited amount of time to respond to each sentence, after which the next sentence will begin, whether or not you have responded.

Sentences will be presented in groups of 4. The second part of your task is to remember a specified word in each sentence, so that you can say these words aloud when prompted at the end of a group of sentences. The specified word will either be the subject of the sentence, or the object. You will be told which word to recall before each group of sentences. You should remember the specified word even if the sentence does not make sense. When all sentences in the group have been completed, you will be prompted with, "NOW," to indicate that you should say the specified words aloud as quickly as possible. You do not need to say them in the order presented. You should just try to recall as many of the subjects or objects as possible.

After you hear the prompt, "NOW," you will be given a limited amount of time to say the cue words (subjects or objects) aloud. Then the next group of sentences will be presented. During each phone call you will be given four groups of sentences without interruption. The cue word type (subject or object) will not change during a phone call. Please do not ask questions or say anything other than the answers to the questions during this time, unless it is urgent.

To summarize, one phone call will consist of four groups of sentences, each of which will have 4 sentences. Before each group of sentences, you will be prompted to be ready, and will be reminded whether you are to recall the subjects or objects for that group of sentences. You will then hear the group of 4 sentences, one at a time. As soon as possible after each sentence, you will say, “YES,” or, “NO,” to indicate whether or not the sentence makes sense. After you have had time to respond to the last sentence, you will be prompted with, “NOW,” which is the signal for you to say aloud the words you were instructed to recall, either the subjects or the objects. After you have been given time to say the subjects or objects aloud, you will be prompted again to be ready for the next group of sentences to start. After four groups of sentences, you will hear “STOP”. This indicates the end of the call and you should hang up.

Here are two examples of two consecutive groups of sentences. You can see your responses in CAPITAL LETTERS and will hear them as a male voice. The first example has subjects as the cue word. The second example has objects as the cue word.

Ready.

Recall <u>Subjects.</u>
    The boy drank the water. YES.
    The girl swallowed the dream. NO.
    The fish ate the ceiling. NO.
    The shortstop caught the ball. YES.
Now.
BOY, GIRL, FISH, SHORTSTOP.
Ready.

Recall <u>Subjects.</u>
    The officer caught the robber. YES.
    The goat ate the ocean. NO.
    The cyclist rode the bicycle. YES.
    The maid boiled the rock. NO.
Now.
OFFICER, GOAT, CYCLIST, MAID.
Stop.

And now for example two.
Ready.
Recall <u>Objects.</u>
    The bear ate the fish. YES.
    The king wore the verb. NO.
    The neighbor entered the paint. NO.
    The girl rode the horse. YES.
Now.

FISH, VERB, PAINT, HORSE.
Ready.
Recall <u>Objects.</u>

The radio played the water. NO.
The hen laid the egg. YES.
The dog chased the tree. NO.
The knife sliced the bread. YES.

Now.

WATER, EGG, TREE, BREAD.

Stop.

Do you have any questions?

Prep Room Practice

Ok. When you are ready, I will present to you the first practice call (by CD instead of phone so we don't have to dial) and you can respond. Notice that it will ask you to recall subjects. Are you ready? (Play Track 3)				
Sentences: **(Play Track 3)**	Judgment / Sensibility		Recollection Memory	
	Meaning	Correct	Recall Word	Correct
Prep Room Practice Call Recall Subjects				
The umpire stopped the game.	YES		umpire	
The ball hit the net.	YES		ball	
The clerk opened the noodle.	NO		clerk	
The cook scrambled the fence.	NO		cook	
The bride wore the dress.	YES		bride	
The chef mixed a city.	NO		chef	
The barber trimmed a bicycle.	NO		barber	
The child rubbed his eyes.	YES		child	
The cat chased the silence.	NO		cat	
The caller left a message.	YES		caller	
The bear climbed the rain.	NO		bear	
The children jumped the rope.	YES		children	
Ok. Good. (comment / feedback as necessary) Now let's try another practice call on CD. This time, note you will be asked to recall objects. Are you ready? (Play Track 4)				
Sentences: **(Play Track 4)**	Judgment / Sensibility		Recollection Memory	
	Meaning	Correct	Recall Word	Correct
Prep Room Practice Call Recall Objects				
The fountain sprayed the water.	YES		water	
The criminal broke the law.	YES		law	
The farmer grew corn.	YES		corn	
The men painted the fear.	NO		fear	
The machine washed the river.	NO		river	
The soldier fired the gun.	YES		gun	
The boy multiplied the tree.	NO		tree	
The chef prepared the mountain.	NO		mountain	
The driver turned the floor.	NO		floor	
The lawyer presented the air.	NO		air	
The boss fired the employee.	YES		employee	
The neighbor entered the house.	YES		house	
Ok. Good. Any questions about the phone conversation task? Next we will train you on the 511 traveler information system task.				

8.7 Appendix G: 511 Traveler Information System Training and Practice

While driving on the test track, you will be asked to make phone calls to obtain information about traffic and road conditions on imaginary roadways. The imaginary roadway system is represented on a map attached to the center console over the radio so that you can see it while driving. Please look at the map now.

As you can see, there are two east/west interstate routes (I-90, I-34) and two north/south interstate routes (I-51, I-75). Various city names are located along these routes. Each route has three consecutive numbered segments, which will be used for some of the questions and answers, as discussed below.

A traveler information system has been created that corresponds to the roadway system shown on the map. This information system has many of the same features as systems that currently exist in several states. The information system consists of an automated phone system, much like those that you may have used to contact businesses or to obtain information about movie show times or business hours.

Periodically, while you are driving on the test track, the experimenter will ask you a specific question about traffic or road conditions on some portion of the imaginary roadway system shown on the map. To answer each question, you will need to contact the traveler information system by speed-dialing number 2 on the mobile phone and use voice commands to move through the system menus.

After being connected to the system, you will hear a welcome message and then the main menu. The main menu has several options, including Traffic and Road Conditions.

- The Traffic option contains information about delays due to accidents, roadway construction, or special events. These are called incidents.
- The Road Conditions option contains information about weather or other conditions that affect the road surface, for example if roads are snow-covered and slippery.

When you are at the main menu, you should choose the option that corresponds to the question that you are trying to answer. Generally, questions will ask either about delays due to traffic incidents or about road conditions. For questions about delays or traffic incidents, you should say "Traffic." For questions about road conditions, you should say "Road Conditions."

Once you have chosen Traffic or Road Conditions, you will be prompted to say the name of a route to hear the detailed reports. For example, if you want traffic information about Interstate 75, after saying "Traffic," you would say "I-75". Then you will hear a list of traffic problems on I-75.

If you want road condition information about Interstate 90, after saying "Road Conditions" you would say "I-90." You will then hear road condition information about I-90.

The detailed reports differ slightly. Traffic incident reports for each route begin with a summary statement that gives the total number of incidents on that route. Then, one by one, the details of each incident are reported. There will also be information about how much delay is caused by each incident.

Road condition reports do not have summary statements. Rather, they describe the road conditions for the various segments of the roadway. Segments are generally identified by the city names at each end of the segment. For example, if a question asks about roadway conditions on the segment numbered 1 on the map, you will listen for information about road conditions on I-90 between Auburn and Bedford. Road condition reports may include information about reduced speed limits or restricted visibility due to weather problems.

There are different types of questions that you will be asked while driving. I will now give you examples. (Examples do not use actual message information.)

1. Which route has shorter total delay due to traffic incidents, I-75 or I-51?

This question is about traffic incidents, so at the main menu, you would say "Traffic." Then you will need to select one of the two routes and listen to the detailed reports for that route. For example, if you take them in the order given, you would say "I-75." You would then hear a summary statement about the total number of incidents and then detailed information about the cause of each incident. Delay information is given at the end of each incident. To get the total delay for I-75 you would need to keep track of the sum of the delays for each reported incident. For example, if I-75 had two incidents and the associated delays were 10 and 15 minutes, you would add them to get a 25-minute total delay for I-75.

After the I-75 message, you will be prompted to say the name of another highway or to say "Menu" to return to the main menu. At this point you still need information about delay due to traffic incidents on I-51, so you would say "I-51" and listen to the detailed reports. Let's assume that I-51 also had two incidents and that the associated delays were 10 minutes each. The total delay would be 20 minutes and since this is less than the 25-minute delay for I-75, the answer to the question would be "I-51 has the shorter total delay."

As soon as you have enough information to answer the question, you should say the answer aloud and then end the phone call by pressing the 'end' button on the phone. Please try to keep your answers brief. Long answers may interfere with the beginning of the next trial.

When you are listening to the reports, you should not worry about which direction you might be traveling on the road and whether the delay occurs only in one direction. For these questions, we just want you to add up the total delay.

Here is an example of another type of question.

2. Which route has better road conditions, I-34 or I-51?

This question is about road conditions, so when you are at the main menu, you would first say "Road Conditions." Then you will need to select each of the two routes and listen to the detailed reports. Road condition reports give information for each segment of the roadway and as shown on the map, each road has three segments. You should listen for key words such as: reduced speed, reduced visibility, snow-covered, slippery, slippery in spots, etc. These suggest significant problems.

Answering questions about road conditions is more difficult than answering questions about traffic incidents because there are no numbers available for direct comparison. Therefore, there is no absolute correct answer and you must use your judgment to determine which route you believe has better road conditions. For example, you might conclude that I-34 has poorer conditions because of reduced visibility than I-51, which is partially snow-covered if visibility problems bother you more than driving in snow. Generally, road conditions that require reduced speed or have restricted visibility are more serious than other conditions. Please remember to make your answer brief. For example, you might simply say: "I-51 has better road conditions."

Here is a question that refers to numbered roadway segments.

3. Which segment has more incident-related time delay, segment 3 or segment 7?

This question is about traffic incidents, so when you are at the main menu, you would first say "Traffic." Then you will need to look at the map to determine which route name has the segments that you are asked to compare. If you look at the map you will see that segment 3 is on I-90, between Cambridge and Franklin. Similarly, segment 7 is on I-51, north of Bedford. Therefore you will need to listen to the detailed reports for I-90 and I-51. When listening to the detailed reports, you will need to pay attention to the location of each reported incident so that you can determine if it is in the segment that you have been asked about. If there is more than one incident for a given road segment, you will need to add the delay for each incident together. If there are no incidents for a given road segment the delay will be zero.

For example, after saying "Traffic," you would say "I-90" and listen to each incident. You would be listening for reports that say that something has occurred between Cambridge and Franklin, or east of Cambridge or west of Franklin. You would then remember the delay associated with any relevant incidents. Next you would select I-51 since segment 7 is on that route. You would do this by saying "I-51" when prompted to say another route name. You would then listen for reports of incidents north of Bedford, adding the delay for each relevant incident.

If you miss some information in a message, or need to hear a message again, you can wait until the message is over and repeat the route name. For example, if you missed some information about I-51, you would say "I-51" when the prompt tells you to say another highway name or say menu. Or, if you don't want to wait until the end of the message, you can say "Menu" at any time, then say "Traffic" and "I-51." Either way will get the I-51 traffic incidents message repeated.

When you have all the information you need to answer the question, you should say the answer aloud and then hang up the phone as soon as possible. You will have a limited amount of time to answer each question. It is therefore possible that if you need to listen to messages more than once, you may not have enough time to obtain all the needed information and will not be able to answer the question. The experimenter will tell you when you need to stop working on the question. At that time you may give your best guess to the question or simply say that you don't know.

Here is another question that uses road segments.

4. Which segments have road conditions that recommend reduced speeds on I-75 or I-90?

This question is about road conditions, so when you are at the main menu, you would first say "Road Conditions." Then you will need to select each of the two routes and listen to the detailed reports. While you are listening to the reports, you will be listening for words that indicate <u>reduced speed</u> recommendations. You need not be concerned about whether the reduced speeds are recommended or mandatory. We are looking for any mention of <u>reduced speeds</u>. You should tell the experimenter the segment number(s) of any road condition reports that mention reduced speeds. You may say the segment numbers at any time, but it is best to wait until the entire message is completed as you may miss some relevant information if you talk during the message.

Other key phrases that might be used in questions include <u>snow-covered roads</u> or <u>reduced visibility</u>.

Summary

When it is time to begin the phone task, the experimenter will notify you by saying: "Please get ready for the next question." At this time you should make sure that the map is available and visible. Several seconds later, the experimenter will read a question aloud. To answer the question, you will need to use the map together with information you obtain from the traveler information system. Below is a summary of the steps required to answer a question:

1. The experimenter will notify you when the next task (question) is about to begin.
2. Several seconds later the experimenter will read the question aloud.
3. When listening to the question, you should listen for the type of information required (Traffic Incidents or Road Conditions). This will tell you which category to select when you reach the main menu of the traveler information system.
4. You should also listen for the specific route names or segment numbers in the question. This will tell you which routes to select when you access the traveler information system.
5. Refer to the map as necessary to determine which routes the question refers to.
6. If you don't understand any part of the question, ask the experimenter to repeat it.
7. Wait for the experimenter to tell you to begin the task.
8. When given the signal to begin, you should contact the traveler information system by speed-dialing number 2.
9. Use the traveler information system to obtain all information necessary to answer the question.
10. Say the answer aloud to the experimenter. Remember to use short answers.
11. Hang up the phone by pressing the 'end' button.
12. Stop working on a question if the experimenter tells you to do so.

Your performance will be scored based on your answer and on the time it takes you to answer each question, so you should answer the question as soon as possible and hang up the phone quickly after answering the question.

If you get confused or lost in the automated phone system, you should say "Menu." You can say this at any time. This will take you back to the main menu. When you first enter the traveler information system, you can bypass the welcome message if you know the category names. This will save some time.

If you make no response after a prompt, the system will repeat the prompt. If you make no response repeatedly, the system will get you back to the main menu.

If you cannot remember the details of the question, you may ask the experimenter to repeat the question at any time.

Do you have any questions?

511 Practice Questions

First, we will try some simple questions. Please feel free to ask questions at any time during this practice. We will call the traveler information system and we can stay connected until all of the questions are completed. Between each question, you should say ‘menu’ as if you were starting a new call / task. Please call the traveler information system now (number 2). Your question is…

1. What is the total delay due to traffic incidents on I-90?
    a. (5 + 5 +10 = 20 minutes)

2. Which segments have road conditions that recommend reduced speed on I-51?
    a. (None)

3. What is the total delay due to traffic incidents on segment 5?
    a. (I-34, Dover to Essex, 5 + 5 = 10 minutes)

OK, now we will try some more complicated questions.

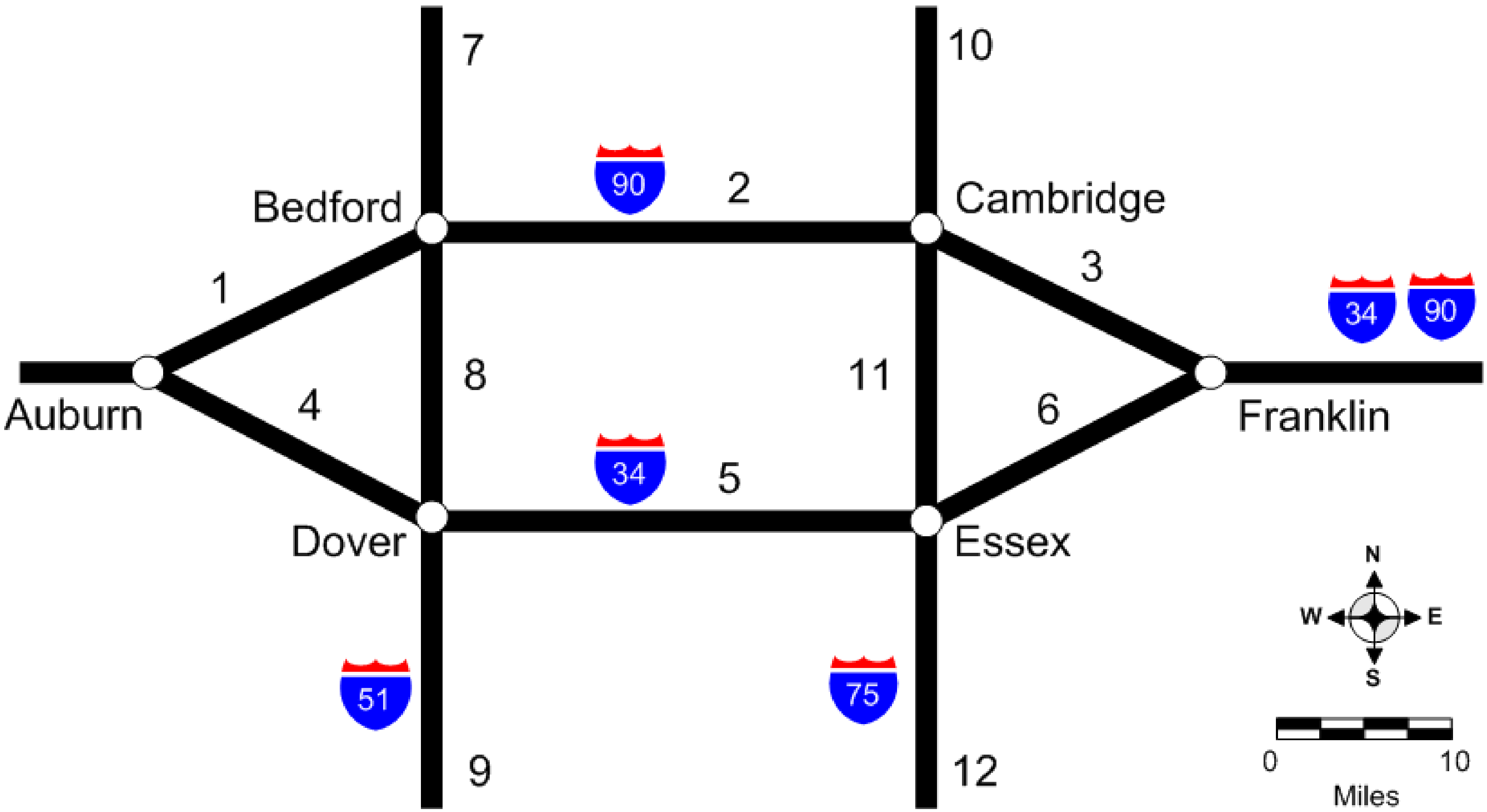
4. Which has better road conditions, I-34 or I-90?
    a. (I-34 has one alert, I-90 has multiple alerts, and therefore, I-34 has better road conditions.)

5. Which has shorter delay due to traffic incidents, segment 3 or segment 6?
    a. Segment 3 (I-90 Cambridge to Franklin, 1 accident, 10-minute delay)
       Segment 6 (I-34, Essex to Franklin, construction, 5-minute delay).
       Therefore, Segment 6 has shorter delay.

6. Which route has more delay due to accidents, I-51 or I-75?
    a. I-51 – 2 accidents (5 + 15 = 20 minute delay)
       I-75 – 1 accident (5 minute delay)
       Therefore, I-51 has more delay due to accidents. Note also, accidents is a subset of incidents.

Do you have any questions?

511 Map



7
10
90
2
Bedford
Cambridge
1
3
34
90
8
11
Auburn
4
6
Franklin
34
5
Dover
Essex
N
W
E
S
51
75
0
10
Miles
9
12

8.8 Appendix H: PDT Instructions

Light Detection Task While you are driving, you will be asked to respond to reflections of small red lights (LEDs), which appear in the left side of the windshield, just above the dashboard. When you see any of the lights appear you should respond as quickly as possible by pressing the micro switch that will be attached to your left index finger. A light will appear approximately every 3 to 5 seconds and will remain lit until the button is pressed, or for about 1.5 seconds if no response is made. You will be scored based on your speed and accuracy in detecting the lights while driving.

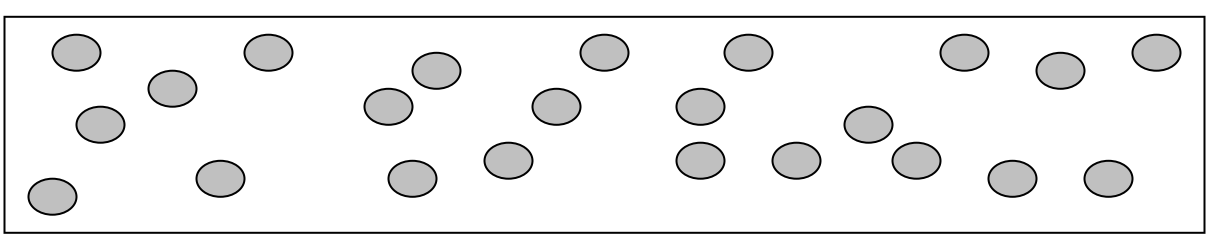
PDT Instruction and Practice – Stationary Vehicle

(*Exp. note: turn transmitter ON*) OK. Now let's try out the light detection task. Go ahead and place the response button on your left index finger and wrist so that it is comfortable and the button can be pressed while you are holding the steering wheel. Now, just try a few button presses in response to the red lights appearing on the left windshield. If you press the button quickly, the light should go out. If you do not respond quickly, it goes out after 1.5 seconds. You will be performing the light detection task while driving on the test track. Are there any questions about this task?

PDT Visual Test

Ok, what I want to do now is confirm how many of the lights you can see on the windshield while sitting, as you normally would while driving. First, what I want you to do is activate all of the lights for me by opening the small light gray compartment door on the left side of the dashboard, to the left of the steering wheel. It looks like an ashtray door or something to that effect. Ok. Do you see a switch inside there? Flip that switch to the left for me; it should turn all of the PDT lights on.

Ok. Now, sit, as you normally would while driving with your hands on the steering wheel. Can you see all of the lights without leaning to the left, by sitting as normal as if you were driving? Please confirm this by leaning to the left slowly, observing whether or not there are more lights in view than when you sit normally. Use this diagram to help determine which lights, if any, are not within your normal view. For some people, the camera blocks some of the lights with most people seeing all of the lights when seated normally. At this time, point to any of the lights on the diagram that you cannot see on the windshield, and I will mark them with an X. Ok. That's it for this task, please flip the switch back to the right and shut its door.



8.9 Appendix I: Monetary Rewards

In addition to your regular pay, we will give you an opportunity to earn a modest amount of money during the experiment. The actual amount of money awarded per trial will be based on your performance in the three tasks shown in the table below. As you can see by the monetary values in the table below, the car-following task is most important, followed by the in-vehicle phone task and light detection task, respectively.

Task	**Good Performance**	**Acceptable Performance**	**Poor Performance**
Car Following	$1.50	$0.75	$0.0
In-Vehicle Phone	$1.00	$0.50	$0.0
Light-Detection	$0.50	$0.25	$0.0
Total	$3.00	$1.50	$0.0

During each session, you will have 6 opportunities to earn the monetary rewards shown in the table. Thus, if your performance is consistently rated as good on all three tasks, you will earn (6 x $3.00) $18.00 in incentive pay during the experimental session. If your performance is acceptable but not good on all three tasks you will earn (6 x $1.50) $9.00 in incentive pay during the session.

Do you have any questions about the monetary rewards?

Ok. Now I will explain a rating scale for mental effort that we will be using today.

8.10 Appendix J: Rating Scale Mental Effort

Instructions

We are interested not only in assessing your performance but also the experiences you will have during the different task conditions. Right now I will describe the technique that will be used to examine your experiences.

Most importantly, we want to assess the mental effort you experience. Mental effort is a difficult concept to define precisely, but a simple one to understand generally. The factors that influence your experience of mental effort may come from the task itself, your feelings about your own performance, how much effort you put in, or the stress and frustration you felt. The mental effort contributed by different task elements may change as you get more familiar with a task, perform easier or harder versions of it, or move from one task to another.

Since mental effort is something experienced individually by each person, there are no effective "rules" that can be used to estimate the mental effort of different activities. One way to find out about mental effort is to ask people to describe the feelings they experienced. We will be using a rating scale to assess your mental effort. Please read the definition of the scale carefully. If you have a question about the scale, please ask me about it. It is extremely important that it is clear to you. The description will be made available to you for reference during the experiment.

Rating Scale Definition
Mental Effort: How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving? How hard did you have to work mentally? How much time pressure did you feel?

After performing a set of tasks, you will be instructed to bring the vehicle to a stop at a specified location. While the vehicle is stopped, the rating scale will be presented to you. You will evaluate the tasks performed (some combination of car following, light detection and phone tasks) since the time when the previous rating scale was administered, by telling the in-vehicle experimenter the number on the scale at the point that matches your experience. Please consider your responses carefully in distinguishing among the different task conditions. Your ratings will play an important role in the evaluation being conducted, thus your active participation is essential to the success of this experiment, and is greatly appreciated.

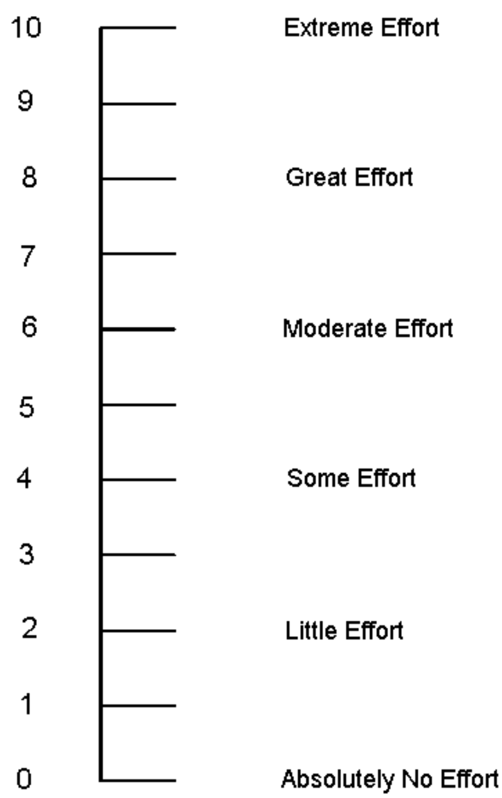
Subject: Date:

Condition: Exp:

Rating Scale Mental Effort

Modified 8/25/03

Please indicate, by marking the scale below, how much effort it took for you to perform the task you just completed



Rating Scale Definition
Mental Effort: How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving? How hard did you have to work mentally? How much time pressure did you feel?

8.11 Appendix K: Computational Methods for Deriving Performance Measures

The measures discussed in this section include measures of steering (steering entropy, steering reversals and holds), and car following (coherence, phase, and magnitude).

K.1 Steering entropy

Steering entropy is a measure of the increase in unpredictability or irregularity of the steering signal due to loading on the driver. It is based on the idea that normal attentive steering involves anticipatory smooth corrective responses to maintain vehicle position. Increased load that distracts the driver results in periods of steering inactivity followed by the need for high frequency corrections. The steering entropy is a measure of the extent and duration of these corrections. It is computed by comparing the steering signal from driving in a loaded (i.e. secondary task) condition with an auto regression model fitted to the baseline (unloaded) drive over the same course. The least squares' residuals are used to compute entropy.

The driver steering angle was filtered with a Butterworth fourth order digital phaseless filter with cut off frequency equal to 3.6 Hz. This cut-off frequency reflects the maximum frequency an average human driver can exert. Extreme maneuvers are inherently low frequency (<1.0 Hz). A cut-off frequency of 3.6 Hz is sufficient to smooth data, keep steering magnitude not reduced within maximum human capacity, and reduce signal noise to negligible values. Doing so, we preserved steering frequency content relevant to drivers' performance.

It has been shown by previous VRTC research that on-center steering corrections for driving straight ahead occur at frequencies below 0.5 Hz (Salaani, Heydinger, & Grygier, 2005)[1], and intensity increases with speed. The steering activities are drivers' adjustments to random disturbances that make the vehicle deviate from its intended path without steering corrections. The random disturbances include tire asymmetrical mechanical properties, road banking, pavement irregularities, vehicle asymmetrical geometrical and compliances properties, etc. The driver's steering corrections for driving straight are small in magnitude and frequency, yet relevant to safety. If the driver does not correct for these disturbances, the vehicle deviates from its intended path.

This filtered hand wheel steering angle is sampled to a sequence of inputs at intervals of 0.15 sec. This time interval reflects average drivers' steering neuromuscular reaction time. This reaction time is used for all drivers regardless of gender and age differences.

The baseline-sampled driver's steering profile is used to estimate the coefficients of the second order Auto Regression (AR-2). The difference between the driver's profile and the one predicted by AR-2 is obtained for baseline and all corresponding loaded runs. This

[1] Salaani, M. K., Heydinger, G. J., Grygier, P. A., "Vehicle On-Center Directional and Steering Sensitivity," SAE Paper # 2005-01-0395

difference is referred to as the steering residual error. The word residual comes from the fact that we are using the least-square technique to estimate auto-regression coefficients. For every baseline, only one coefficient set is obtained and used for the corresponding loaded conditions.

The AR-2 is generated using Matlab with the covariance method. The function to generate the auto regression coefficients is provided next. The least-square technique is used, and the auto-regression function written for this specific project is listed below.

The AR-2 fitting routine

```
function [b,B,nu,s,r] = ar_fit(x,p)
%
% Fit a reference posterior Auto-Regression with order (p) model to the series x
%
%%%%% Written by Mohamed Kamel Salaani at VRTC - Cell-phone study %%%%%%%%%%
% Feb. 2005
%

%
% remove the bias on the steering around the center
%
arx = x-mean(x);
%
% Constract Hankel matrix
%
n=length(arx);
y=arx(n:-1:p+1);
X=hankel(arx(n-1:-1:p),arx(p:-1:1));
%
% get autoregression coefficients
%
B = inv(X'*X);
b = B*(X'*y);
%
% get the residual errors
%
r = y-X*b;
%
% compute variance for the residual error
%
nu = n-2*p;
s = r'*r/nu;
%
% Compute the variance matrix for the y-vector
%
 B = B.*s;
%
```

The estimated residual error is obtained as follows:

$$R_e^i = X^i - \left(b_1 X^{i-1} + b_2 X^{i-2}\right) \qquad (1)$$

Where

b_1, b_2: Auto-regression coefficients

X^i: Driver steering input at time interval *i*

R_e^i: Residual error at time interval i

Using the auto-regression coefficients (b_1, b_2) from the baseline run, residual errors for loaded conditions are then computed using Eq. 1.

The baseline residual error distribution is divided into 10 bins as follows:

$$
\begin{aligned}
B_1 &= (-\infty; -5\alpha] \\
B_2 &= (-5\alpha; -3.5\alpha] \\
B_3 &= (-3.5\alpha; -2\alpha] \\
B_4 &= (-2\alpha; -\alpha] \\
B_5 &= (-\alpha; 0] \\
B_6 &= (0; \alpha] \\
B_7 &= (\alpha; 2\alpha] \\
B_8 &= (2\alpha; 3.5\alpha] \\
B_9 &= (3.5\alpha; 5\alpha] \\
B_{10} &= (5\alpha; \infty]
\end{aligned}
$$

The parameter α is found such that:

$$\alpha = f\left\{\Pr\left((0;\alpha]\right) = 0.0857\right\} \qquad (2)$$

We should also note that,

$$\sum_{i=1}^{10} \Pr(B_i) = 1 \qquad (3)$$

Then the entropy of the baseline is found using this formulation:

$$H_b = -\sum_{i=1}^{10} \Pr(B_i) \log 10\left(\Pr(B_i)\right) \qquad (4)$$

If the residual error from the baseline is symmetric with distribution close to normal, then the entropy of the baseline is very close to 0.5. For any condition different from the baseline, the entropy will either increase or decrease. It increases if the residual errors are distributed more away from the cente,r which indicates a higher standard deviation if we assume a normal distribution. The entropy decreases if the distribution of the residual errors is more compacted at the center, which reflects a smaller standard deviation if we assume a normal distribution. Therefore, steering entropy provides a summary measure for overall steering performance. The entropy is a normalized measure and does not require measure normalization adjustments. Theoretically, the entropy varies from 0 to 1. To attain a value of one, the probability of each bin is identical, and equal to 0.1. Zero entropy is the trivial case of zero probability for each bin. For the maximum entropy condition:

$$\Pr\left(B_i\right) = 0.1 \quad (5)$$

$$H_{Max} = -\sum_{i=1}^{10} 0.1\log 10\left(0.1\right) = -\sum_{i=1}^{10} \log 10(0.1^{0.1}) = -\sum_{i=1}^{10}(-0.1) = 1 \quad (6)$$

The entropy of the loaded conditions is computed in a similar manner as for the baseline, using exactly the bins ($B_{1\to 10}$) and α from the baseline run. The entropy scale is as follows:

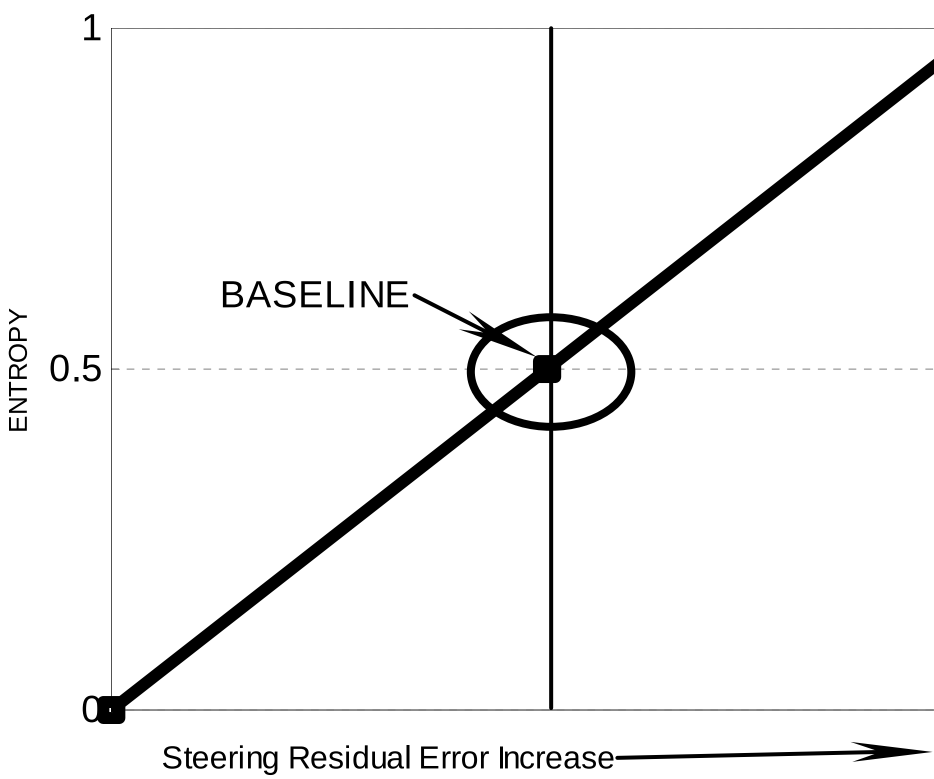


Figure K1. Entropy Scale

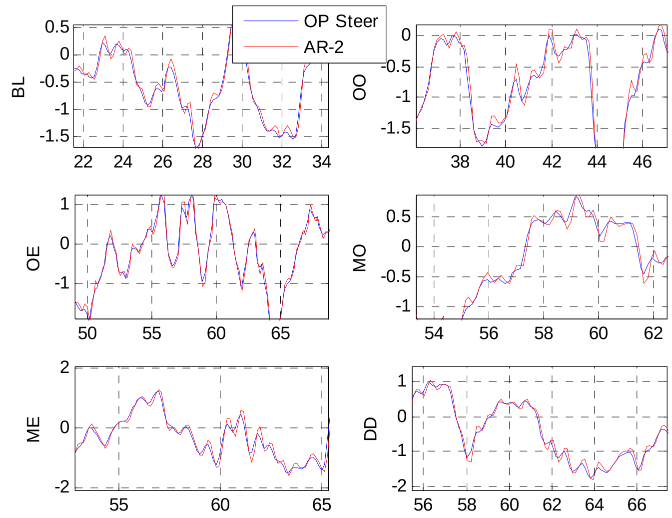


Figure K2. Steering profiles of drivers and AR-2 predicted profile –Shown for six different task loading conditions

Figure K2 shows the steering profile of a baseline run (BL) and runs at different task loading conditions. The actual driver profile and the AR-2 predicted profiles are shown. The entropy calculation is basically a measure of how closely the AR-2 corresponds to the actual profile using regression coefficients obtained from the base line run. Figure K3 shows the cumulative density function of the base line run, and figure K4 shows the same function for a typical task-loading run. Figure K5 shows the distribution of the residual error for each condition. The residual error is the difference between the actual steering profile and the corresponding profile predicted with auto regression (AR-2). Figure K6 shows the probabilities of the error bins and the corresponding entropy values.

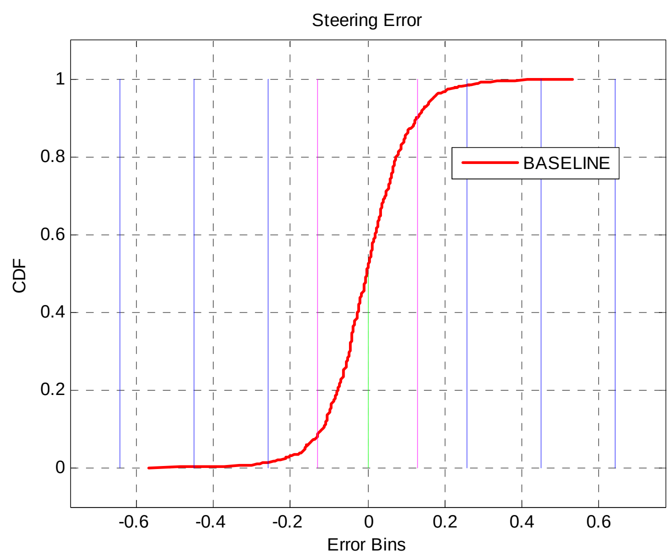


Figure K3. Cumulative density function of steering residual error for a baseline run

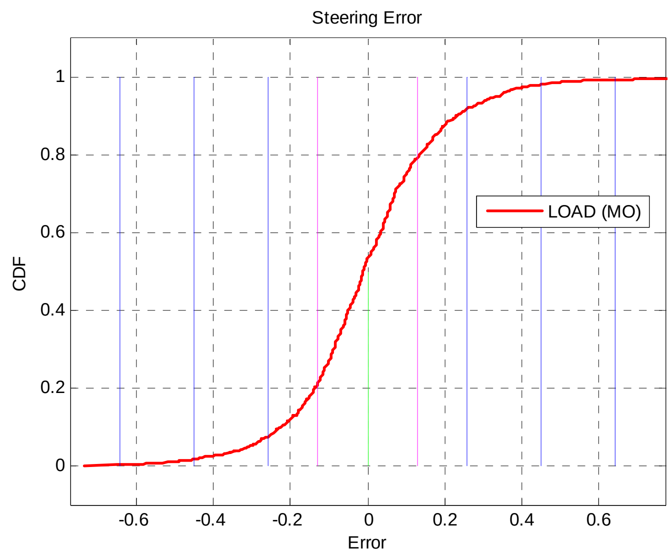


Figure K4. Cumulative density function of steering residual error for a task loaded run

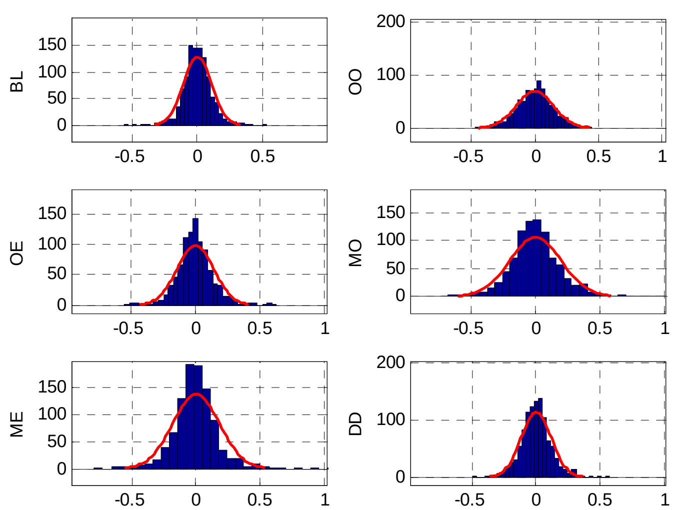


Figure K5. Distribution of steering residual errors for different task loadings

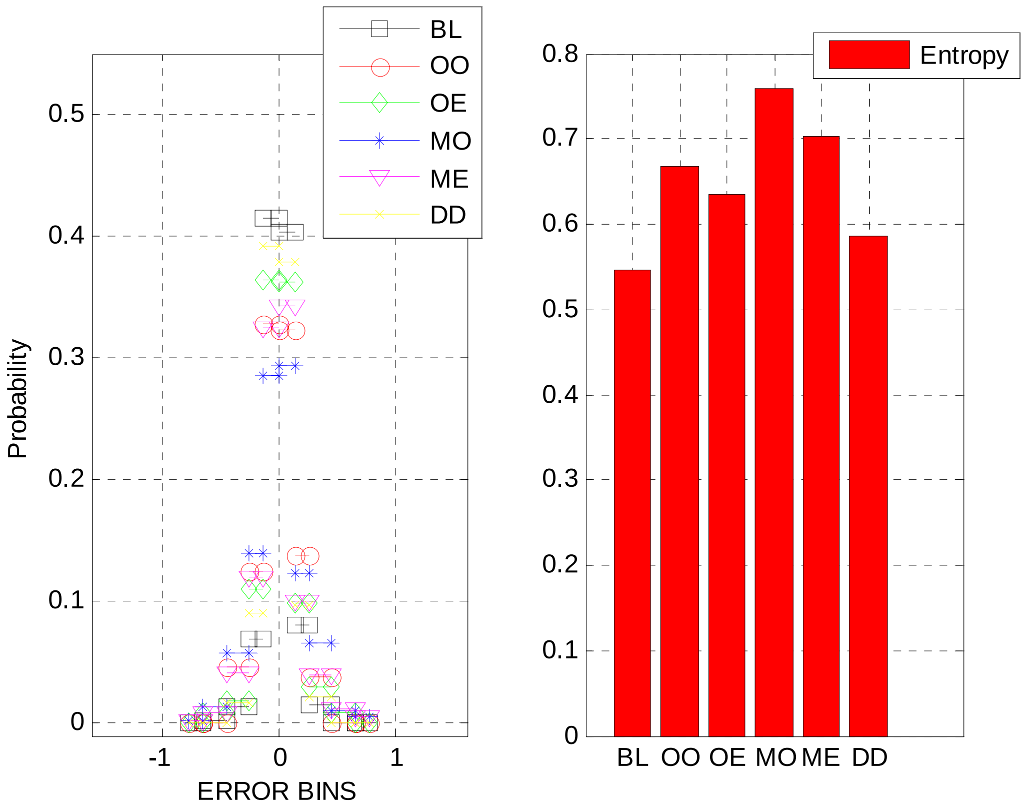


Figure K6. Probability distribution of bins and entropy results for different task loadings

K.2 Steering Reversals and Holds

The hand wheel steering position and positional rate were first filtered using a 4-pole Butterworth low pass phaseless digital filter with a 3.6 Hz cutoff frequency. The rationale for this cut-off frequency is explained above. The data for this section were not sampled at human reaction time rates, because the steering reversals and holds are mainly trend measures. The following measures are computed:

- Standard deviation of steering position
- Steering reversals
- Steering holds (count)
- Probability of steering holds

Steering reversals were determined using both the hand wheel rate and position data. A reversal was defined as when the hand wheel rate passed through the steering rate dead band and the hand wheel position magnitude changed by more than a positional threshold value. The parameters (threshold values) for this measure are defined based on vehicle type and subjective judgments. Steering threshold values used for this project should not be considered universal, since they depend on vehicle testing speed and steering system mechanical properties. For example, a steering system with a higher steering free-play should have a higher steering positional threshold value. A vehicle that responds fast to steering input (e.g. sports car) should have a minimal steering rate dead band. For this project, the steering rate dead band is ±1.5 deg/sec and the positional threshold value is ±1.0 deg.

Steering holds are defined when the driver does not steer for a specified time. Numerically they represent steering segments where the steering rate is within the dead band rate. The hold time interval is set to 0.4 sec. The hold time interval is chosen to be about 2.5 times a drivers' neuromuscular reaction time. The steering hold count measure does not provide the distribution of hold intervals. In other words, if the driver holds for 0.4 sec the hold count is increased by 1, and if he/she holds for 1.0 sec the hold count is similarly increased by 1. Information about the length of steering holds is not also reflected by this counting scheme. To overcome this, we introduced an additional measure, which is the probability of holding more than 0.4 sec. This probabilistic measure includes the effects of holding interval lengths. Figure K7 shows an example of this measure.

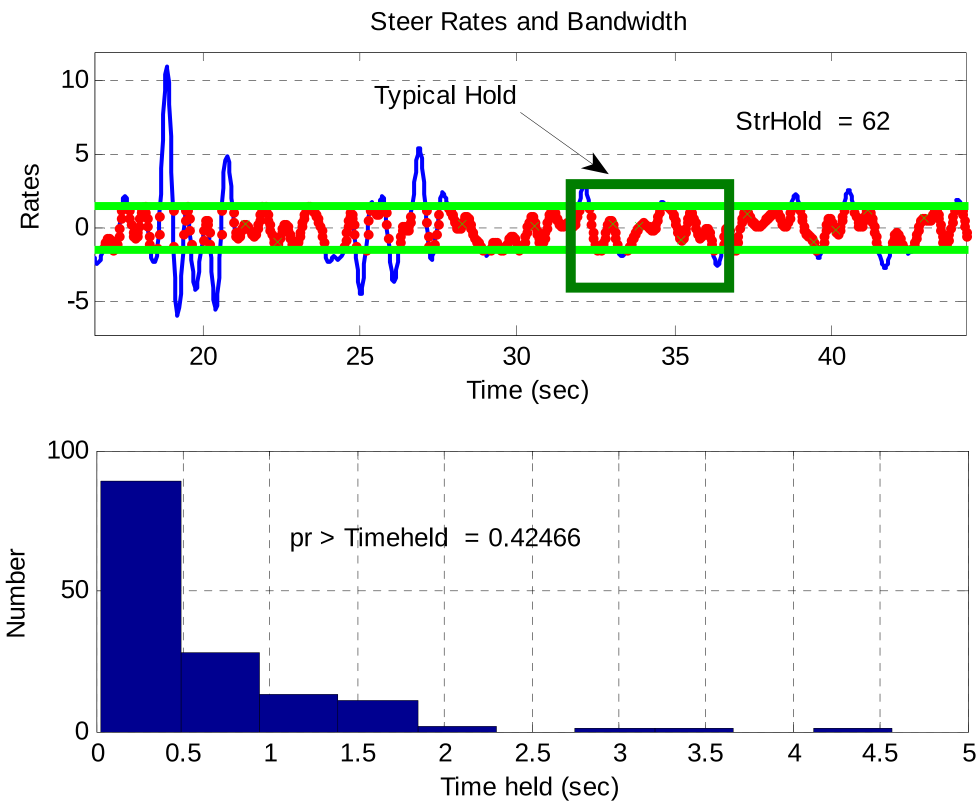


Figure K7. Steering hold and probability of hold

The steering hold and hold probability function

```
function [StrHold, prTimeheld] = get_SteerHold(nxr, t0, show_plots)
%
% Computing steering hold
%
%
% Written by Kamel Salaani at VRTC - VRTC Track Data Study
%
global Vnoise dead_band Holdtime Driver_time Sampl_time p1 HoldTime1
%
% finding number of reversals
%
indx  = find(nxr < Vnoise/2 & nxr  > -1*Vnoise/2);
%
nn = length(indx);
%
if length(nxr) == length(indx)
    StrHold = -1;
    prTimeheld = -1;
    return
end
%
StrHold = 0;
Timeheld = 0;
tc = 0;
j  = 1;
for i=1:(nn-1)
    if (indx(i)+1) ~= indx(i+1)
        if tc ~= 0
            Timeheld(j) = tc;
            j = j +1;
        end
       tc = 0;
    else
```

```
        tc = tc + Sampl_time;
        if tc > Holdtime && tc < (Holdtime + Sampl_time)
            StrHold = StrHold + 1;
            indx2(StrHold) = indx(i);
        end
    end
end
%
[fi xi] = ecdf(Timeheld);
i = max(find( xi < HoldTime1));
if ~isempty(i)
    pr = fi(i);
else
    pr = 1;
end
prTimeheld = 1 - pr;
```

The steering reversal function

```
function StrRev = get_SteerRev(x, xr, t0, show_plots)
%
% Computing steering reversal
%
%
% Written by Kamel Salaani at VRTC - VRTC Track Study
%
global Vnoise dead_band Holdtime Driver_time Sampl_time p1 HoldTime1
%
% finding number of reversals
%
indx  = find( xr > Vnoise/2 | xr  < -1*Vnoise/2 );
%
nn = length(indx);
StrRev = 0;
str1 = 0;
for i=1:(nn-1)
    if (xr(indx(i))*xr(indx(i+1)) < 0)
        if str1 == 0
            str1 = x(indx(i));
        else
            str2 = x(indx(i));
            if abs(str2-str1) > dead_band
                StrRev = StrRev + 1;
                indx2(StrRev) = indx(i);
            end
            str1 = 0;
        end
        %
    end
end
```

K.3 Car Following Analysis

K.3.1 Sine speed profile

The lead vehicle speed can be either sinusoidal or random. The sinusoidal signal is intended to test the subject at a constant frequency (around 0.03 Hz) that mimics a steady flow of traffic. This load might not be what people are subjected to on highways, but it provides a simple analysis. Figure K8 provides a comparison of lead and following vehicle sine speed profiles. The measures of this signal are taken at the fundamental frequency of 0.03 Hz. The power spectrum signal, transfer function and coherence are meaningful at the 0.03 Hz, and the computational 'leaks', values at other frequencies away from the central 0.03 Hz, are ignored. If we increase the number of cycles, the computational 'leaks' would decrease. However, due to the fixed length data collection interval, this was not possible.

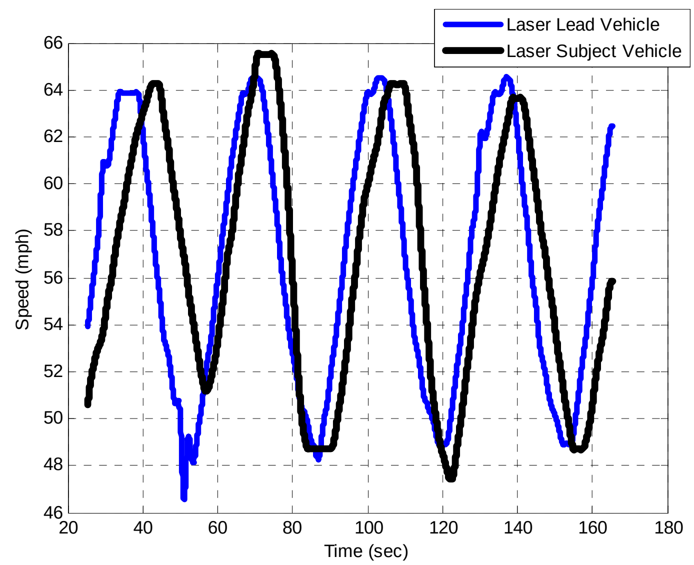


Figure K8. Speed profiles of lead and subject vehicle for a typical run –Sine speed profile

Figure K9 shows the power spectrum of a typical baseline run of a sine speed profile. The sine wave was at 0.03 Hz. The power spectrum around this frequency is power leakage associated with the computational method used. Figure K10 shows the transfer function, with phase expressed in degrees. Figure K11 is the coherence of the lead and following vehicle. Figure K12 shows the distance between the lead and the subject vehicle.

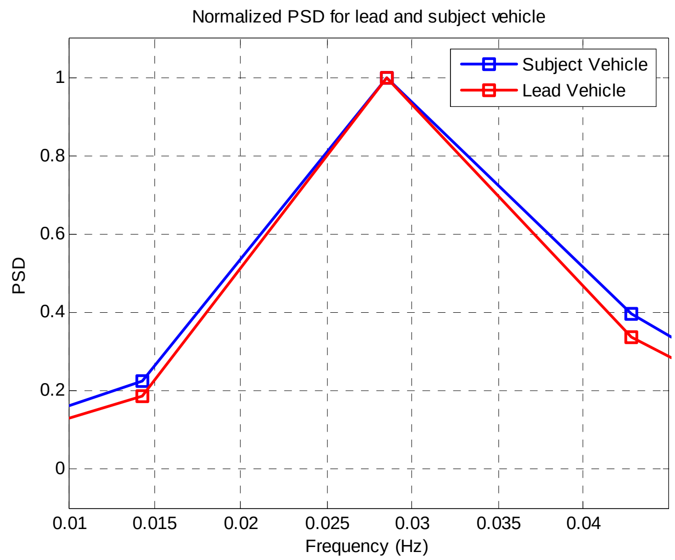


Figure K9. Normalized speed power spectrum of lead and following vehicles for a typical run – Sine speed profile

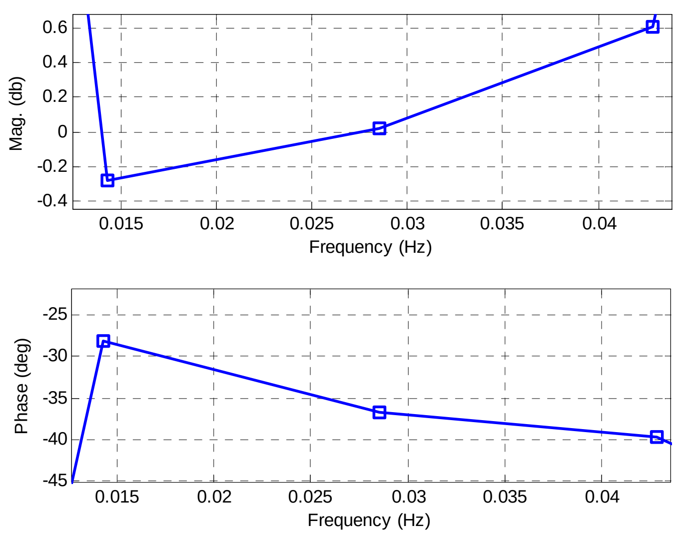


Figure K10. Speed transfer function (Lead/Follow) for a typical run –Sine speed profile

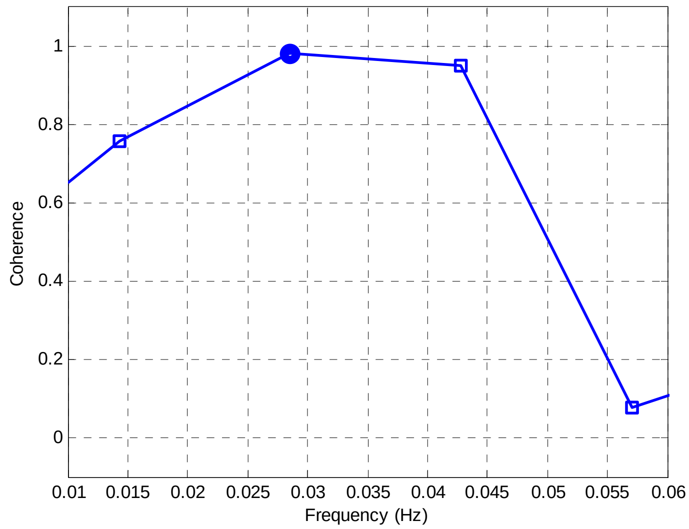


Figure K11. Speed Coherence (Lead-Follow) for a typical run –Sine Speed profile

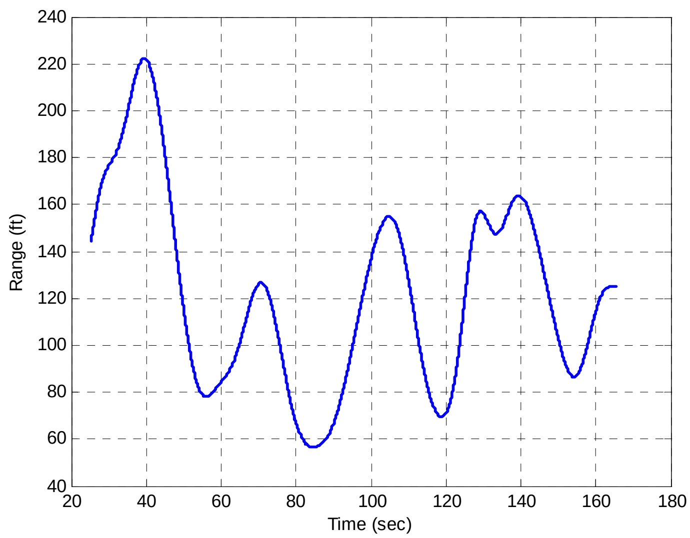


Figure K12. Range between lead and follow vehicle for a typical run

K.3.2 Random speed profile

The random signal was obtained by band filtering a white noise signal with a band frequency that mimics the variations of traffic frequencies on highways. The band frequency was set to [0.02 – 0.045] Hz. The lead speed signal was randomly varied as shown in Figure K13. The random signal can have many shapes, by trying different random seed levels. Due to space limitations, and the resulting testing time limit, the shape of the random signal chosen was the one that provided the best (almost flat) power-spectrum signal within the band frequency with the maximum number of data points. This random signal tests the driver at different frequencies, and provides a more realistic approach to car following human factors testing. The lead vehicle and subject vehicle time histories must be stationary random processes, that is, the mean and covariance are independent of time. The lead vehicle time history satisfied this criterion since its speed is obtained from filtering a Gaussian noise source. We assume the subject vehicle is maintaining the same mean. We assume the subject vehicle velocity fluctuations have covariances independent of time. The spectral analysis is done using the Matlab spectral analysis toolbox. The random speed is obtained from a white noise signal having a Gaussian distribution, and filtered with band pass filter.

$$V_N = normrd(0, Speed\_Std, Total\_Number\_Samples, 1) \qquad (7)$$

This random speed is filtered with a band pass, [*f1 f2*], fourth order filter and added to the mean speed values.

$$V_l = V_0 + filter(V_N, [f1; f2]) \qquad (8)$$

Where,

V_l : Lead vehicle speed

V_0 : Mean speed

$[f1; f2]$: Frequency bandwidth of interest

The computation of an ensemble average of spectral properties provides smooth results with significantly reduced random error. The spectral properties are obtained using a windowing function. A Hanning window is used because of its ability to reduce ‘leakage’ of power around the main lobe. Without using an appropriate window the spectral estimates are distorted. Since the Hanning window discards relevant information at the beginning and end of each record, the ensemble-averaged records were overlapped by at least 80% of window duration.

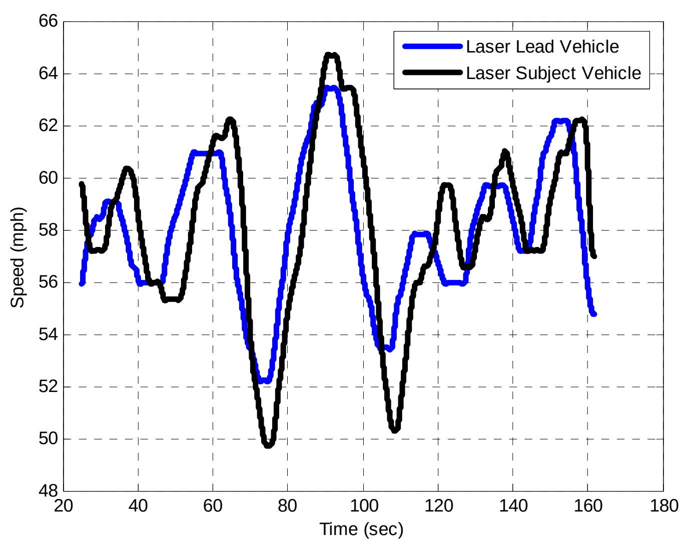


Figure K13 Speed profiles of lead and subject vehicle for a typical run – Random speed profile

Figure K14 is the power spectrum of both the lead and the follow vehicles. Since the speed spectrum is designed to have varying frequencies between 0.02 and 0.045 Hz, the power around 0.03 Hz is not power leakage, and should be taken into account in determining the complete power spectrum properties. Figure K15 shows the transfer function, where the magnitude (db) and phase (deg) are shown. Figure K16 is the coherence, which is reasonably high for the frequency band of interest. Figure K17 presents the range in feet between the lead and the following vehicle.

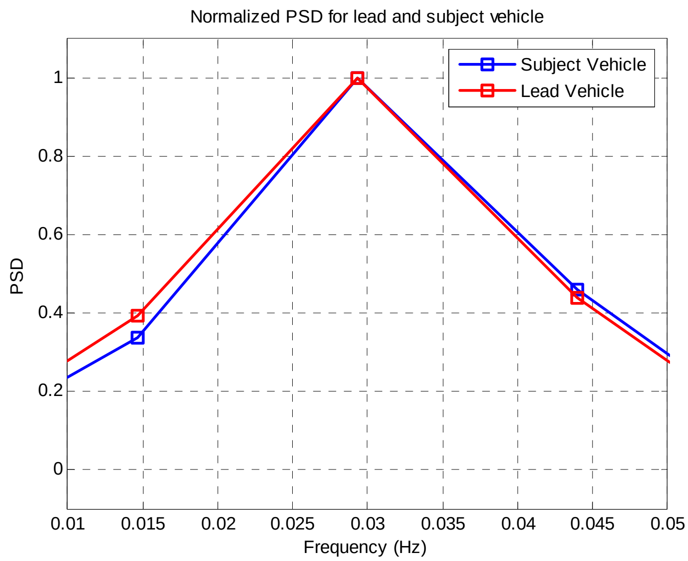


Figure K14. Speed Power Spectrum for a typical run – Random speed profile

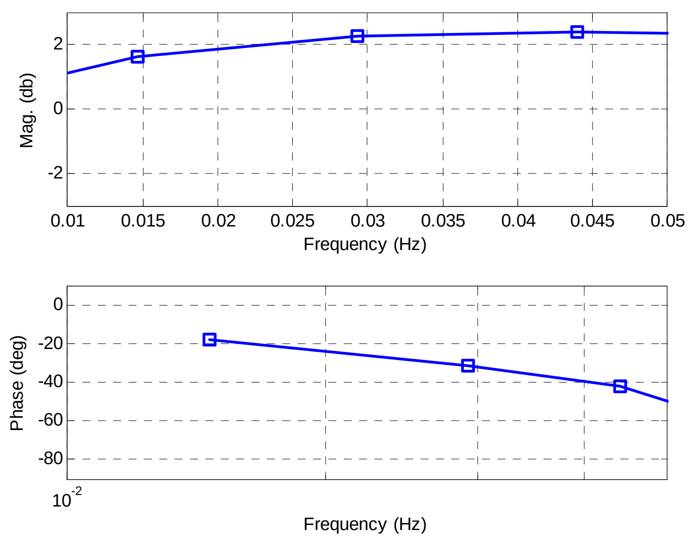


Figure K15. Speed Transfer function for a typical run – Random speed profile

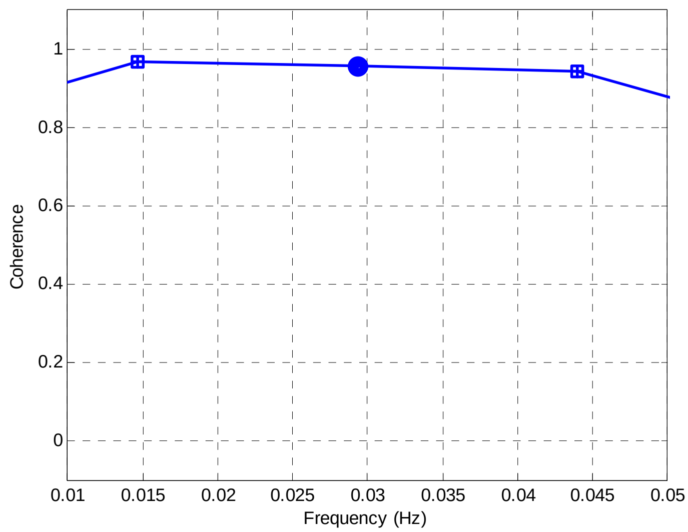


Figure K16. Speed coherence function for a typical run – Random speed profile

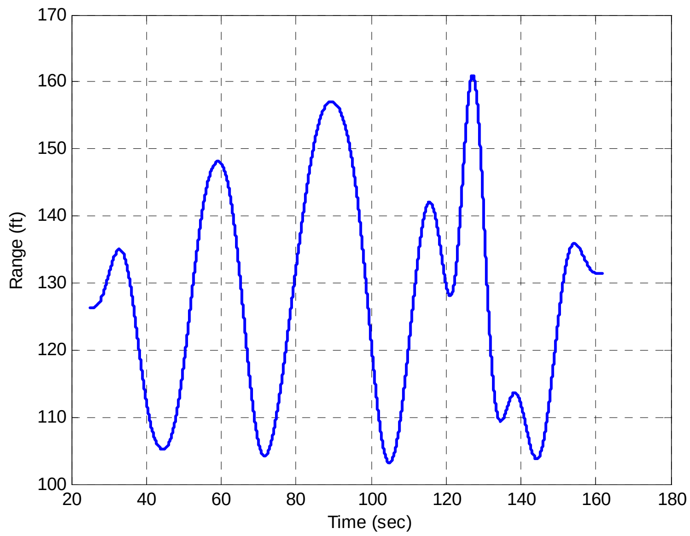


Figure K17. Range between lead and subject vehicle for a typical run – Random speed profile

Figures K18 and K19 show the histograms of the sine and random speeds and their corresponding accelerations. The random speed profile provides a distribution with more variation around the mean speed, while the sine wave provides more speed variations away from the mean speed. The acceleration profile for the sine wave is also more distributed away from the center.

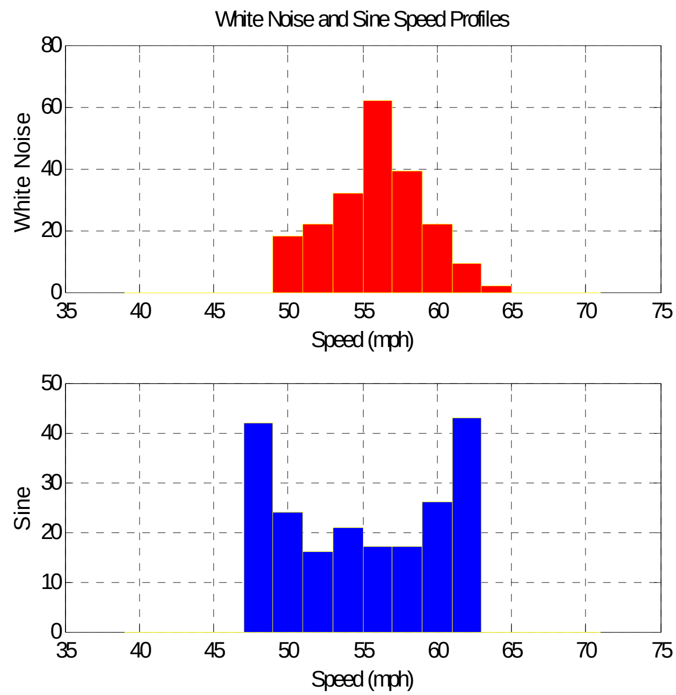


Figure K18. Typical speed histograms for sine and random speeds

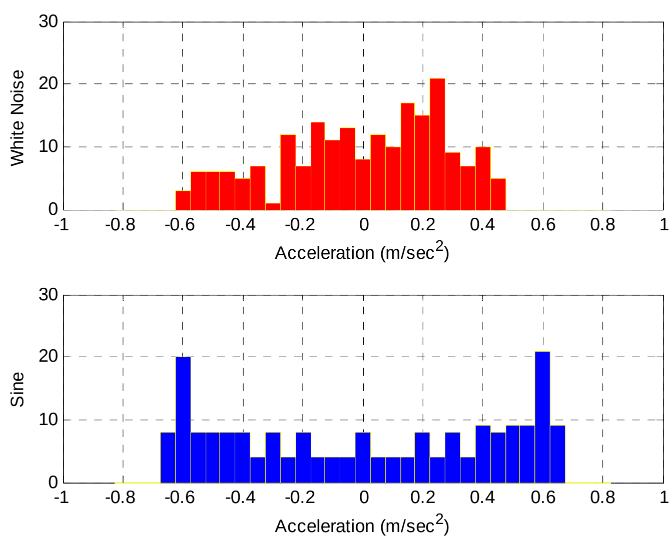


Figure K19. Typical acceleration histograms for sine and random speeds

Figure K20 shows a typical difference of sine and random speed variations for a long testing profile, which was not used in this research but is added for clarification of the differences between the two methods. In this figure, the power spectrum of the sine profile has less power leakage and provided good results at the fundamental frequency, while the power spectrum for the random speed provided results for wide bandwidth, from 0.02 Hz to 0.04 Hz. It is clear from this result that random speed tests the driver at more frequencies than the sine wave. If we decrease the number of sine waves, the power spectrum will flatten more at the fundamental frequency (0.03 Hz) due to numerical reasons, which is called power spectrum leakage. It is therefore important to distinguish between signal properties and numerical power leakage.

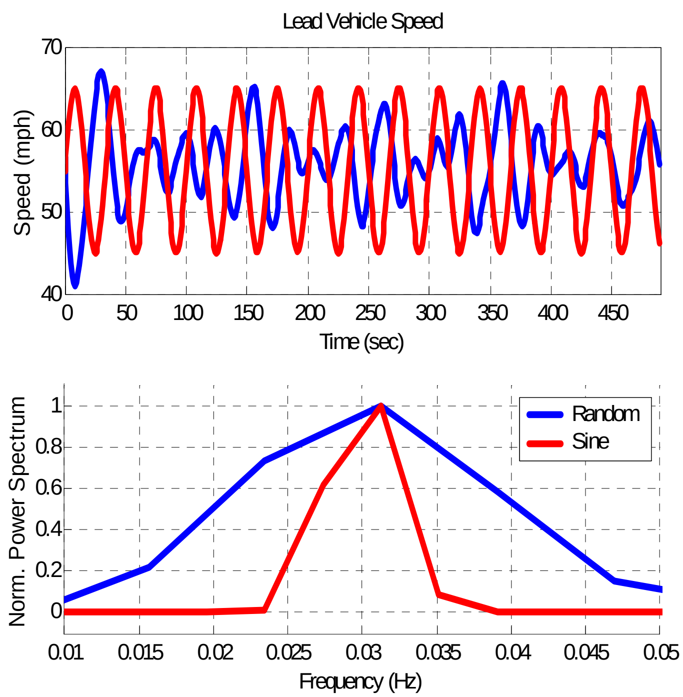


Figure K20. Typical normalized power spectrum for random and sine speed

K.4 References

Salaani, M. K., Heydinger, G. J., Grygier, P. A. (2005) Vehicle On-Center Directional and Steering Sensitivity, SAE Paper # 2005-01-0395. Society of Automotive Engineers, Detroit.