U.S. Department
of Transportation

**National Highway
Traffic Safety
Administration**

DOT HS 809 525 **December 2003**

The Effects of Voice Technology on Test Track Driving Performance: Implications for Driver Distraction

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 for graphical images contained in this report to satisfy Section 508 of the Americans With Disabilities Act (ADA).

Technical Report Documentation Page

1. Report No. DOT HS 809 525	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle
The Effects of Voice Technology on Test Track Driving Performance: Implications for Driver Distraction

5. Report Date
December 2003

6. Performing Organization Code
NHTSA/NVS-312

7. Author(s)
Thomas A. Ranney, Transportation Research Center Inc. (TRC)
Joanne L. Harbluk, Transport Canada
Larry Smith, Kristen Huener, Ed Parmer, Transportation Research Center Inc. (TRC)
Frank Barickman, National Highway Traffic Safety Administration (NHTSA)

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 Heath Albrecht (TRC) for technical support, Elizabeth Mazzae (NHTSA) and W. Riley Garrott (NHTSA) for their administrative support, and Y. Ian Noy (Transport Canada), Moshe Eisenman (El Mar. Inc.) and Patricia Trbovich (Transport Canada) for their technical contributions.

16. Abstract

The objectives of this research were: (1) to compare the distraction potential of voice-based versus visual/manual interfaces for selected transactions undertaken while driving, (2) to examine the effect of performing tasks of differing complexity on driving performance, and (3) to evaluate the potential of using eye-tracking technology to make inferences about changes in subjects' allocation of attention while performing specified in-vehicle tasks while driving. Twenty-one subjects completed two sets of eight laps around a 7.5-mile test track during two four-hour sessions. They drove an instrumented vehicle while performing a combination of car following, peripheral target detection, and secondary (in-vehicle) tasks of varying complexity. Subjects performed one set of laps with each of two interfaces, voice-based and visual/manual. Secondary tasks comprised three categories including baseline tasks (radio tuning, phone dialing), simple tasks (message retrieval plus voice memo creation), and complex tasks (simple task components plus phone dialing and information retrieval from automated phone systems). Measures of driving performance, target detection, secondary task performance and eye movements were recorded. Analyses were conducted to determine whether the voice-based interface reduced the relative distraction potential for secondary tasks of varying complexity. Performing secondary tasks while driving resulted in significant decrements to vehicle control, target detection and car-following performance. The voice-based interface helped reduce the distracting effects of secondary task performance. Improvements were relatively minor and limited to vehicle control and visual performance measures. There was no effect on car-following measures, suggesting the voice interface had little effect on cognitive distraction. The results suggest that voice interfaces may not provide enough help to overcome the cognitive distraction associated with secondary tasks of increasing complexity, particularly in driving situations that require time-space judgments and tactical decision-making.

17. Key Words
Driver Distraction, Driver Performance, Voice-interface, Eye tracking

18. Distribution Statement
Document is available to the public from the
National Technical Information Service
Springfield, VA 22161

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 78	22. Price

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

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

Despite the significant limitations of crash data for determining the incidence of distraction as a contributing factor in crashes and the absence of matched exposure data necessary for assessing the crash risk associated with various distracting activities, the scientific evidence continues to accumulate in support of the conclusion that wireless phone use while driving can compromise safety. Within this context, the next generation of in-vehicle technologies is viewed with increasing concern that the adverse effects resulting from drivers' interactions with such multi-function devices as navigation systems and systems that allow PC functionality, including email and Internet access, will be greater than those associated with wireless phone use. The objectives of this research were therefore: (1) to compare the distraction potential of voice-based versus visual/manual interfaces for selected transactions undertaken while driving, (2) to examine the effect of performing tasks of differing complexity on driving performance, and (3) to evaluate the potential of using eye-tracking technology to make inferences about changes in subjects' allocation of attention while performing specified in-vehicle tasks while driving.

Twenty-one subjects completed two sets of eight laps around a 7.5-mile test track during two four-hour sessions. Eleven subjects were unskilled test drivers (Group 1); 10 subjects were recruited from NHTSA's Vehicle Research and Test Center (VRTC) engineering staff (Group 2). They drove an instrumented vehicle while performing a combination of car following, peripheral target detection, and secondary (in-vehicle) tasks of varying complexity. Subjects performed one set of laps with each of two interfaces, voice-based and visual/manual. Secondary tasks comprised three categories including baseline tasks (radio tuning, phone dialing), simple tasks (message retrieval plus voice memo creation), and complex tasks (simple task components plus phone dialing and information retrieval from automated phone systems). Measures of driving performance, target-detection, secondary task performance and eye movements were recorded. Car-following performance was analyzed using coherence between lead and following vehicle speeds and associated measures (phase shift or delay and modulus or gain). Target-detection was measured with the Peripheral Detection Task (PDT), in which subjects responded to LEDs reflected on the windshield of the experimental vehicle. Analyses were conducted to determine whether the voice-based interface reduced the relative distraction potential for secondary tasks of varying complexity.

The results indicated strong and consistent driving performance decrements associated with secondary task performance, relative to the driving-only condition. Despite instructions to maintain constant following distances, subjects drove at longer following distances when performing secondary tasks. This allowed them to allocate less attention to car following, which was reflected in lower coherence scores. In addition, the car following delay, represented by the phase shift between the two vehicle speed signals in the coherence analysis, increased by more than one second during secondary task performance, relative to the trials with no secondary task. This increased delay reflects slower response time over the entire secondary task interval in contrast to response times for discrete events, which characterize behavior over an interval of only several seconds. Increased steering reversal rates during secondary task performance suggest increased total task workload, while increased steering hold rates indicate withdrawal of attention from steering during secondary task performance. Secondary task performance also affected both lane keeping, as reflected by increased lane-position variability, and target detection, as reflected by reduced percentage of PDT targets detected.

The results generally did not support the hypothesis that increasing task complexity would exacerbate distraction effects associated with performing the secondary tasks while driving. The performance of complex secondary tasks did not consistently interfere more with driving performance than did the baseline tasks of continuous radio tuning and phone dialing.

Observed differences between interface conditions were less evident and generally weaker than differences between secondary task conditions, most notably for the car-following measures. Subjects drove at longer following distances while using the visual/manual interface for two of three secondary task conditions relative to the voice interface; however, this difference was considerably smaller than differences between secondary task conditions. There were no differences observed between interface conditions for the car-following performance measures (coherence, phase shift, modulus). In contrast, the observed reductions in steering wheel holds and standard deviation of lane position reflected improved vehicle control performance in the voice interface condition relative to the visual/manual interface condition.

We found consistent differences between the subject groups. Group 1 drivers had longer following distances, detected fewer PDT targets and generally had more difficulty performing the required tasks together. Based on questionnaire responses, we found that Group 1 drivers had less education, less experience with personal computers and Internet activities, less experience in multi-tasking while driving, and were less willing in their everyday driving to perform secondary tasks like those used in this study than Group 2 drivers. They were also less likely to drive in heavy freeway traffic, which requires following at shorter headways. These results suggest that individual differences can be expected to influence the extent to which engaging in complex secondary tasks will impact real-world driving behavior.

The pattern of results indicates that both the car-following paradigm and the PDT were sensitive to performance decrements associated with secondary task performance. Performing secondary tasks while driving resulted in significant decrements to vehicle control, target detection and car-following performance. Drivers generally increased their following distances when performing secondary tasks. The voice-based interface helped reduce the distracting effects of secondary task performance, but the improvements were relatively minor and were limited to the manual and visual aspects of the tasks. The voice-based interface did not improve performance on the car-following task, which is more cognitively demanding than other driving task components. Although the voice interface used in this study may not be representative of future products, the results suggest that voice interfaces may not provide enough help to overcome the cognitive distraction associated with secondary tasks of increasing complexity, particularly in driving situations that require time-space judgments and tactical decision making.

The results of the present study provide information concerning the relative potential for interference with driving, when the secondary tasks are undertaken in situations that have primary task demands comparable to those used in the experiment. The real-world effects of secondary task engagement on driving behavior and safety also require consideration of factors that influence drivers' willingness to engage in the secondary tasks, the resulting incidence of secondary task engagement, and the associated crash risk.

1.0 INTRODUCTION

1.1. Driver Distraction and Wireless Phones

Currently over 137 million Americans use wireless phones (IIHS, 2002). It is further estimated that 3.9% of passenger car drivers are using their wireless phones while driving at any point in time (Utter, 2001). The accelerating growth of wireless phone use while driving has fostered interest among safety professionals and more recently among the public at large in the associated problem of driver distraction. Despite the significant limitations of crash data for determining the incidence of distraction as a contributing factor in crashes (e.g., Goodman, Tijerina, Bents, & Wierwille, 1999) and the absence of matched exposure data necessary for assessing the crash risk associated with phone use (Stutts, Reinfurt, Staplin, & Rodgman, 2001), the scientific evidence continues to accumulate in support of the conclusion that wireless phone use while driving can compromise safety (e.g., Redelmeier & Tibshirani, 1997; Laberge-Nadeau et al., 2001). Perhaps more importantly, it is also becoming clear that the problem of amassing sufficient scientific evidence to rigorously demonstrate the role of distraction as a causal factor in crashes is a significant and daunting problem, requiring convergent evidence from numerous studies (e.g., Tijerina, 2000).

1.2. Multi-Function In-Vehicle Technologies

Within this context, the next generation of in-vehicle technologies is viewed with increasing concern for we can readily imagine that the adverse effects resulting from drivers' interactions with such multi-function devices as navigation systems and systems that allow PC functionality, including email and Internet access, will be greater than those associated with the cellular phone. In particular, we assume: (1) that the potential for adverse effects will increase monotonically with the increasing availability of in-vehicle technologies; (2) that there may be no limits to what some drivers will attempt to do while driving, and (3) that these tendencies will be exacerbated by the increasing complexity of transactions involving such new technologies. The present research represents a step toward providing empirical support for the hypotheses embodied in these assumptions.

2.0 OBJECTIVES AND APPROACH

2.1. Study Objectives

The emergence of mobile technologies that permit message retrieval and Internet access allows for the integration of a variety of complex task capabilities into moving vehicles. The safe operation of these technologies is predicated on the assumption that voice-activated and voice-based interfaces are sufficient for preventing significant distraction among drivers performing complex transactions while driving. However this assumption is not well tested. One objective of this research was therefore to compare the distraction potential of voice-based versus visual/manual interfaces for selected transactions undertaken while driving.

As PC-based technologies expand the types of tasks that can be conducted while driving, it is important to determine whether increases in task complexity will provide a significantly greater potential for distraction, relative to the more simple tasks that are commonplace among drivers today. In particular, tasks that require navigation of hierarchical menu systems with many options, or searching for a specified item in a database of messages or phone numbers, may impose a significant burden on the driver's working memory, which could result in an increase in distraction. The second objective of this experiment was to examine the effect of performing tasks of differing complexity on driving performance.

Measuring distraction in a moving vehicle implies the need to monitor rapidly changing attributes of the driver's attention. The characteristics of the driver's point of gaze, including visual search patterns, have implications for his or her distribution of attention while driving and engaging in secondary tasks. The third objective of this study was to evaluate the potential of using eye-tracking technology to make inferences about changes in subjects' allocation of attention while performing specified in-vehicle tasks while driving.

2.2. Terminology

Driving is referred to as the primary task, while the in-vehicle tasks, including message retrieval, phone dialing, and radio tuning are referred to as the secondary tasks. The focus of the present study is on driving performance, which reflects the fact that drivers are instructed and rewarded for performing the primary and secondary tasks to the best of their abilities. This is in contrast to driving behavior, which reflects what drivers choose to do on the road when they are free to choose how much attention to allocate to primary and secondary tasks. The experiment thus measures the distraction potential of the various secondary tasks, defined as the decrement in driving performance associated with a particular secondary task under a given set of driving conditions.

2.3. Experimental Approach

The real-world incidence of distraction and its effects on driving behavior and safety depend on numerous motivational factors that contribute to the driver's willingness to engage in secondary tasks together with the demands of the associated driving situations. Because they typically eliminate drivers' motives, experimental studies are best suited for evaluating the relative distraction potential associated with various in-vehicle tasks. The experimental approach taken in this study was to control the primary (driving) task demands and measure the relative level of interference associated with secondary tasks with different task characteristics and performed with different interfaces.

A common experimental paradigm used to assess the distraction potential of in-vehicle tasks utilizes a car-following task, in which the lead vehicle decelerates unexpectedly while the driver is engaged in a secondary task (e.g., Lee, McGehee, & Brown, 2000). The main performance measure is the subject's response time to the initiation of lead vehicle deceleration. Distraction is inferred from increases in the time required to respond to the lead-vehicle deceleration, relative to an undistracted condition. One potential problem with this methodology is the possibility that driver behavior may change as a result of the use of repeated surprise trials (Muto & Wierwille, 1982). This is particularly likely if the unexpected event involves a near-critical situation, which elicits a relatively strong emotional response and a correspondingly strong motivation to avoid future occurrences. In response, drivers may develop strategies for reallocating their attention to anticipate subsequent surprise events. Indeed, Lee, et al., (2000) found strong evidence that drivers modified their attention following an initial surprise event. If the strategies adopted by drivers to avoid subsequent near-critical situations involve allocations of attention that are fundamentally different from those associated with normal driving, it can be argued that the data obtained from repeated surprise trials may not apply to real-world surprise events, which typically occur so infrequently as to lose their influence over one another (Muto & Wierwille, 1982).

There are several possible solutions to this problem, each of which has associated strengths and weaknesses. Perhaps most valid would be to include only one surprise event per subject, however this adds a significant burden to the data collection by increasing the number of subjects required for a particular experimental design. The second approach is to eliminate the first surprise trial on the assumption that all subsequent trials benefit equally from the information obtained during the first surprise. However, this solution does not address the possibility that drivers' attentional strategies differ fundamentally between the experimental and corresponding real-world situations. A third option is to select a methodology that allows assessment of distraction potential in non-critical situations and thus does not require the use of repeated surprise event trials. This latter option was the approach taken in this study.

The method, based on the work of Brookhuis, de Waard and Mulder, (1994), utilized a car-following task in which the speed of the lead vehicle was varied systematically and the speeds of the lead and following vehicles were subjected to a transfer and coherence analysis. The coherence between the speeds of the two vehicles is a measure of squared correlation, reflecting the accuracy of the following driver's adaptation to changes in the lead vehicle speed. When coherence is relatively high (e.g., = .70), the driver is adequately following the lead vehicle's speed changes. Brookhuis, de Waard and Mulder (1994) have shown that distraction due to wireless phone use while driving increased the phase shift of the two speed signals, reflecting an increase in the lag or response time in the car following task, which they refer to as delay. They argued that this measure of delay incorporates both perceptual and cognitive factors, in contrast to other measures that focus primarily on operational-level behaviors, which are more automatic and less susceptible to effects of distraction. Moreover, because the delay is computed using data obtained over the entire interval during which a driver is engaged in the secondary task, it represents a broader assessment of driver distraction than would result from the use of a discrete stimulus and single response time measure, which applies only to the very short interval during which the surprise event stimulus and secondary task performance co-occur. This difference becomes increasingly relevant for assessing the distraction potential associated with complex secondary tasks, which have components with different attentional demands (e.g., searching for, retrieving, interpreting, and responding to messages). The use of data characterizing the driver's response to a single discrete stimulus may be expected to provide unreliable results if the secondary task demands are not consistent across all task components.

The Peripheral Detection Task (PDT) (Harms & Patten, 2003) was used to measure workload associated with secondary task performance. This task required drivers to respond to simple visual stimuli presented near the driver's line of sight. The task was performed continuously, with targets appearing for one second every 3-5 seconds. The proportion of targets missed and response time have been shown to be sensitive indicators of workload. Because the PDT task demands remain constant throughout the experiment, we considered it to be a part of the primary task, reflecting momentary changes in drivers' awareness of events occurring in the forward view.

Eye-movements were recorded so that eye-glance behavior could be analyzed to determine how drivers allocate their attention while performing secondary tasks together with driving under relatively constant levels of demand. Specifically, analyses were conducted to identify changes of allocation of visual attention (e.g., changes in the percentage of time the drivers' eyes are off the road) associated with the two interfaces and whether the visual functional field of view is reduced as the cognitive load associated with the secondary task is increased (Recarte & Nunes, 2000).

The study objectives were addressed in an experiment in which subjects performed a series of in-vehicle (secondary) tasks both manually and with a voice-based interface while driving an instrumented vehicle. Specifically, the combination of car-following, PDT performance and secondary task performance was performed on a closed course with minimal traffic traveling in the same direction. Secondary tasks were selected according to the following criteria: (1) tasks were available via the Clarion AutoPC; and (2) tasks could be performed using either a visual/manual interface or voice-activated/speech-based interface. Specific secondary tasks meeting these criteria include searching a database of messages or phone numbers, retrieving messages, phone call initiation, and radio tuning.

2.4. Experimental Design

The experimental design had one between-subjects factor, Driver Group (Group 1, Group 2, explained below), and two within-subjects factors: (1) Secondary Task (none, baseline, simple, complex) and (2) Interface (voice, visual/manual). Simple secondary tasks required subjects to search for a specified message (e.g., "Movies") and record a voice memo. Complex secondary tasks included these same components plus additional intervening components, including finding and dialing a phone number, and retrieving information from an automated phone system. The baseline task involved continuous radio tuning or phone dialing.

3.0 METHOD

3.1. 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 our data collection. Due to the relatively low speeds used in this experiment, the two vehicles used the innermost lane. Occasionally, stopped traffic was present, necessitating a passing maneuver. Data collection was suspended when a passing maneuver was required. Data collection was also 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.

3.2. Participants

Twenty-one subjects participated in the experiment. Ten subjects were female, 11 male. Participant age ranged from 22 to 67 years (M = 40.3, SD = 13.9). Eleven subjects were TRC Level-1 test drivers (Group 1) and 10 subjects were staff engineers at the National Highway Traffic Safety Administration's (NHTSA) Vehicle Research and Test Center (Group 2). TRC Level-1 test drivers have no special driving skills. Subjects were required to have prior experience with computers, Internet access, and email. All subjects received their normal pay for the time during which they participated in the study plus incentive pay based on their performance.

We had planned to use TRC Level-1 test drivers exclusively as participants in this experiment. However, preliminary results indicated that some subjects were not able or willing to perform the tasks as instructed. Specifically, due in part to their training and indoctrination concerning test track safety procedures, the test drivers had difficulty maintaining a close following distance. In addition, two TRC test drivers declined to participate because they were not comfortable performing the three tasks simultaneously. As a result, the pool of available test drivers was not adequate to complete the original design matrix and we recruited subjects from the Vehicle Research and Test Center (VRTC) engineering staff to increase the number of subjects tested. Differences observed between the groups are discussed in the Results section.

3.3. Apparatus

A 1995 Chevrolet Lumina served as the lead vehicle (LV), and a 1996 Honda Accord was the subject vehicle (SV). Both vehicles were equipped with Micro Data Acquisition System (MicroDAS) (Barickman & Goodman, 1999) and GPS receivers. GPS position readings were used to derive vehicle speed. A Vorad radar device on each vehicle measured range (inter-vehicle spacing) and range rate to the other vehicle. The two platforms collected data independently at a 30Hz sample rate. Data collection was started independently at the beginning of the trials by an experimenter in each vehicle. The SV had a secondary brake for emergency activation by the experimenter accompanying the subject. The SV also had an event switch, which produced stepped voltages to mark the beginning and end of secondary task performance in the continuous data stream.

3.3.1. Subject Vehicle

The SV MicroDAS was configured to collect vehicle speed, range and range rate, lateral position, hand-wheel position, GPS timing signals, and subject responses to the PDT. The primary SV data collection channels are displayed in Table 1.

Table 1. Subject Vehicle Data Collection Channels

Data Channel	**Description**	**Units**	**Resolution**
Vehicle Speed	Ground speed	kph	1 kph
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)	Degrees	.1 degree
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

The PDT consisted of an array (3 x 20 cm) of 23 LEDs positioned on the dashboard and shielded from direct view of the driver. 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 5-25° to the left of the driver’s line of sight and 2-4° above the dashboard. Drivers responded via a wireless micro switch attached to the index finger of their left hand.

Secondary tasks were performed using the Clarion AutoPC and a Motorola Piper cellular telephone. We used the El-Mar VISION 2000 eye tracking system (Eizenman, Jares, & Smiley, 1999) to obtain eye-movement data from subjects while they were participating in the study. The driver-side airbag was deactivated to permit safe use of the head-mounted eye tracker.

3.3.2. 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	kph	1 kph
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 µs

The LV speed was controlled to approximate a sine wave with a mean of 40 mph, amplitude of 7.5 mph, and a 25-s period (0.04 Hz). The LV speed controller was created by interfacing a laptop with the vehicle's built in cruise controller and running a basic proportional integral derivative control loop in software. The PC consisted of a 486DX laptop with a data acquisition board. The data acquisition board captured analog signals from the vehicle.

Input/ouput (I/O) lines link the vehicle's cruise controller with the vehicle's standard cruise control buttons. These I/O lines control the Cruise Enable, Cruise Speed Set, Cruise Accel, Cruise Coast, and Cruise Cancel functions. The laptop was connected via a digital to analog converter to the I/O lines. The vehicle speed input, as measured by the LV's transmission speed sensor, provided feedback for the system.

LV speed was controlled by sending signals from the laptop through the D/A converter and sending the resulting analog voltage pulses on to the appropriate I/O line. To make the vehicle accelerate, the Cruise Accel line was pulsed; to make the vehicle decelerate, the Cruise Coast line was pulsed. In preliminary testing, it was determined that just pulsing the Cruise Coast I/O line to slow the vehicle did not provide enough deceleration authority to provide the desired speed variation period. It was determined that placing the LV in second gear would provide the desired amount of deceleration authority when commanding the cruise control system to coast.

To operate the automated cruise controller, the experimenter selected second gear and accelerated the vehicle to approximately 35 MPH. The experimenter then engaged the cruise controller through the PC.

3.4. Procedure

Each subject completed two sessions, each lasting approximately four hours and conducted on different days. During the first session, the experiment's purpose and terms of participation were explained and the subject's informed consent was obtained (Appendix A). The subject was then trained in a stationary vehicle using one interface condition. Half of the subjects received the voice-interface condition first, while the other half received the visual/manual interface condition first. The subject then received practice in the stationary vehicle and 30 minutes of practice driving. Two experimenters accompanied the subject during all driving trials: one experimenter prompted subjects and monitored data collection; the second experimenter monitored surrounding traffic.

After a short break, the subject completed four laps of the test track with the eye tracker and four laps without it. Data collection was restricted to the two straight segments on each test track lap.

The duration of each trial was determined largely by the requirement that the eye tracker not be worn continuously for more than approximately 20 minutes. In addition, all eye tracker trials were grouped in the middle of the driving trials to minimize the time required to calibrate the eye tracker. Before each trial, the experimenter read instructions describing the task. The experimenter notified the subject to begin the secondary task at the beginning of each straight section and recorded the beginning and end of secondary task performance period so that task-completion times could be computed. Drivers were instructed to maintain a constant following distance and to complete the designated secondary tasks as quickly as possible.

At three points during each session (see Table 3.), the subject completed a workload rating form (NASA TLX, see Appendix C) (Hart & Staveland, 1988), without the weightings (Moroney, Biers, Eggemeier, & Mitchell, 1992). Performance feedback, in the form of monetary incentive earnings, was presented at the same intervals.

The second session was identical to the first, except that the introductory segment was not included and a different interface condition was used.

Table 3. Driving Trial Structure

Trial	**Number of laps**	**Straight segments**	**Eye tracker**	**Feedback/ Workload rating**	**Comments**
1	2.5	1-5	No	Yes	Calibrate eye tracker after event 5
2	1	6-7	Yes	No	Short duration to compensate for calibration time
3	2	8-11	Yes	Yes	
4	2	12-15	Yes	No	Segments 14 and 15 for repeats, Eye tracker off after trial
5	2	16-19	No	Yes	Segment 19 for repeat, Workload rating for entire session

3.5. Tasks

3.5.1. Driving Task

Although car-following performance measures are sensitive to differences in following distance, Brookhuis and colleagues (Brookhuis & Waard, 1994) found that subjects were not able to follow at a specified distance. They have had better success allowing subjects the freedom to select a following distance. However, the use of different following distances creates potential problems for data analysis, which are discussed, along with approaches to minimize their impact, in the data analysis section. Following their recommendation, subjects in the present study were instructed to follow the LV at a close but safe following distance and to maintain the same distance under all driving conditions. To motivate accurate car-following performance, the subjects were instructed that their performance would be evaluated on how well they could maintain their performance within a specified range.

3.5.2. Peripheral Detection Task

At intervals that varied randomly between three and five seconds, one of the 23 LEDs was illuminated. Each activation lasted one second, unless terminated by the driver's response.

Drivers were required to respond 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.

3.5.3. Secondary In-Vehicle Tasks

Subjects used either the voice-activated or visual/manual interface for an entire session. At the beginning of each driving trial, the subjects were given specifications for a set of messages stored in the AutoPC message database. Each message contained specific task instructions. Subjects were instructed to retrieve the messages in a specified order and to complete each task before retrieving the next message.

Secondary tasks were composed of the components presented in Table 4. Each task had an initial retrieval component, which required subjects to select the appropriate message and then read or listen to it. Similarly, each task had an action component, in which subjects executed the instructions contained in the message. Simple tasks required subjects to create voice memos using information contained in the message. Complex tasks had additional components, in which subjects found a specified phone number, initiated a call, and retrieved information from an automated phone system. An example is presented in Appendix G.

Table 4. Secondary Task Components

Secondary Task	Task description	Variations	Frequency per Session
Simple	Open message containing list of items, create voice memo	Shopping List, Errands	2
Complex	Open message, open phone book, autodial automated system, retrieve information, create voice memo	Store Hours, Airline Schedule, Movies theatre schedule, Package shipping information	8
Continuous Baseline	Dial sequence of phone numbers, tune sequence of radio stations	Familiar/unfamiliar phone numbers	4
None	Driving Only		2

Table 5 presents the components for one complex task trial and the implementation with visual/manual and voice interfaces. Shaded actions were accomplished in the same manner for both interface conditions. Specifically, manual scrolling through messages, which allowed the driver to view each message header, was used for both conditions because the voice-based scrolling required listening to the body of each message before moving onto the next message. Voice messages were recorded in the same way in both interface conditions. Finally, information from automated phone systems was obtained aurally in both interface conditions.

Subjects were allowed to complete secondary tasks only during straight roadway segments, each of which is approximately two miles long. After a preparatory signal, the experimenter told the subject when to begin the secondary task. At the same time, the experimenter activated the start button to mark the secondary task in the MicroDAS data files and begin recording the secondary task completion time. When the subject completed the secondary task, the experimenter terminated event timing and marked the end of the secondary task in the data set via button press.

Table 5. Driver Actions Associated with Complex Secondary Tasks and Implementation by Interface Condition

Task Component	**Action**	**Interface**	
		Visual/Manual	**Voice**
1. Select specified message	a. Activate “Messages” program	Manually	Voice command
	b. Scroll to specified message	Manually	Manually
	c. Select message	Manually	Voice command
2. Obtain message content	Read/Listen to message	Visually	Aurally
3. Find phone number of specified contact	a. Open “Address Book” program	Manually	Voice command
	b. Scroll to specified contact	Manually	Voice commands
4. Place phone call	a. Dial selected number	One touch dialing	One voice command
	b. Retrieve specified information	Aurally	Aurally
5. Create voice memo	a. Open “Voice Memo” program	Manually	Voice command
	b. Record voice memo	Vocally	Vocally

3.5.4. Monetary Rewards

Subjects were given performance incentives to motivate acceptable car following, PDT, and secondary task performance. Monetary rewards were computed based on experimenter ratings of subject performance using rating scales shown in Appendix D. The tasks were weighted to motivate consistent priorities across subjects. The maximum possible reward during each session was: Car-following: $4.50, In-vehicle task: $3.00, and PDT: $1.50. Subjects received feedback concerning their performance following each 20-minute trial. The monetary rewards were paid directly to the subject in addition to the hourly wage they were earning for their participation in the study.

3.6. **Data Reduction**

3.6.1. Data Collected

Overall, data were obtained for 647 straight segments, or approximately 31 per subject. The average duration was 133.9 seconds (SD = 47.9).

3.6.2. Data Synchronization

Data were synchronized between the SV and the LV in post processing. Since both the SV and the LV had GPS sensors, this was accomplished by using the coordinated universal time (UTC) in conjunction with the pulse-per-second (PPS) timing signal. The UTC message is an asynchronous message that is received by the MicroDAS through an RS-232 message. Because of latencies in interrupt handlers and other delays (message parsing, other serial messages, etc.) the UTC message can lag actual UTC time. Accurate timing can be derived by using the message in conjunction with the PPS signal. The PPS signal is specified by the GPS manufacturer to be accurate within +/- 1 microsecond on the rising edge when selective availability is turned off. Figure 1 describes the signal timing. As shown, the UTC times (edges) do not necessarily align with each other, even though their time values are the same. If these values were used alone, synchronization of two data sources could produce errors. However, it can be observed that the PPS timing signal does occur at approximately the same time in each data set(+/- 1 µs). Accurate synchronization between independent data sources is accomplished by finding the common UTC time values and then temporally aligning the data sets when the PPS signal transitions from a low to a high.

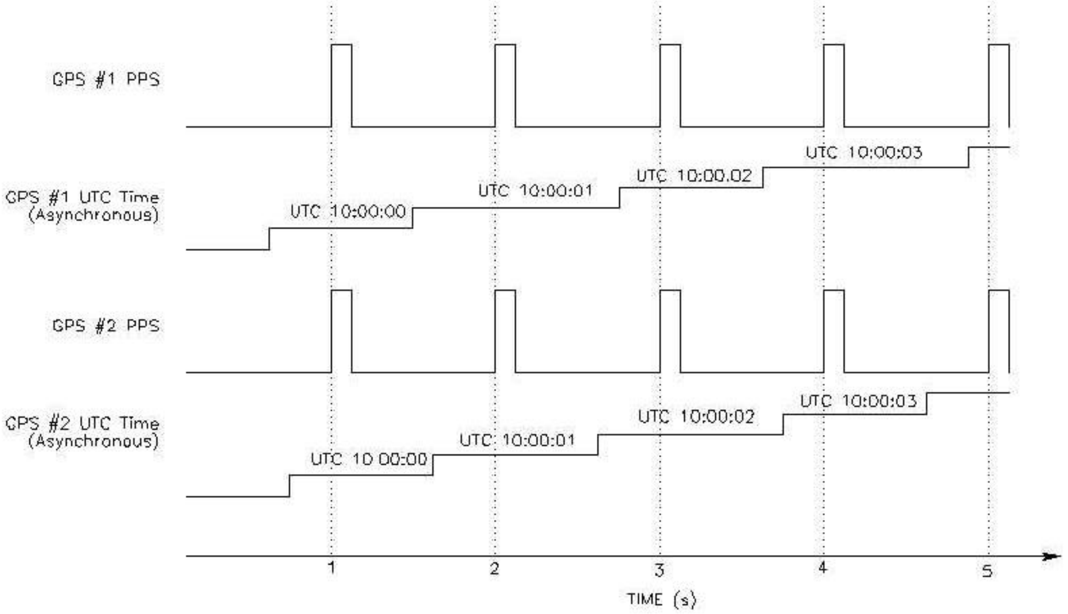


Figure 1. GPS synchronization timing diagram.

Using this method, the SV and the LV data were synchronized. We wrote a software routine that found the first common time between the files and then aligned each data set with the corresponding rising edge of the PPS signal from both sets. Checking the PPS signal periodically ensured complete synchronization between both systems.

3.6.3. Transfer and Coherence Analysis of Car-Following Data

Both test vehicles were equipped with an Eaton Vorad EVT-200 Radar system. The SV had the EVT-200 antenna mounted on the front of the vehicle to measure the headway distance to the LV. Since radar data is typically noisy and prone to dropouts, a second EVT-200 was mounted on the LV facing towards the rear to measure the tailway distance to the SV.

The following channels were directly output from each of the Vorad EVT-200 Radar electronic control units (ECU):

- Raw Vorad Range Channel 1
- Raw Vorad Range Channel 2
- Raw Vorad Range Channel 3
- Raw Vorad Rate Channel 1

- Raw Vorad Rate Channel 2
- Raw Vorad Rate Channel 3
- Vorad Target ID 1
- Vorad Target ID 2
- Vorad Target ID 3

The Vorad ECU is capable of outputting data for up to three targets simultaneously. Data pairs (Range and Rate) are output according to the ECU's determination of the three closest targets, with the one channel being the closest, the two channel being the next closest, and the three channel being the third closest. Each target is assigned an ID for tracking purposes. When the Vorad ECU determines that a target has switched from one channel to another, the corresponding Vorad Target ID is supposed to switch accordingly. If the target is determined to be a different object, the Vorad Target ID is incremented. Although the target ID numbers are sometimes updated correctly, often the internal tracking software has problems and the same target is assigned a new Target ID.

To overcome this limitation, VRTC developed a processing function that combines these outputs to track the target of interest (the LV). The processing routine scans a block of the raw radar data and attempts to eliminate outliers. Values are deemed unreasonable when their range rate is equal to or greater than the host vehicle's current velocity. Those data points thus deemed to be erroneous in the range, range rate, and target ID channels are set to zero. This method of filtering is not effective for use in a collision-avoidance system, but is valid for measurement purposes in this test where the LV is known to be moving at a similar velocity to the SV. The block of data is then scanned for similar target IDs. If a target ID is found to be present in greater than 10% of the data block it is held in memory. The corresponding data with target IDs present in less than 10% of the block are set to zero. An array for range and range rate is then constructed based on the frequency counts of the target ID data. When several data points exist for the same time value, the data point with the most common target ID is given preference.

The data are output into combined Vorad_Range and Vorad_Rate channels that represent the target vehicle being tracked. The following figures display the output of the function. Figure 2 displays the unprocessed range and range rate data, in which the traces are made up of data from all three Vorad channels (for both Range and Range Rate). Figure 3 displays that data after the removal of outliers and completion of processing.

Finally, the data were filtered for small dropouts, or spikes. For this processing, the maximum width of a spike to be removed was set to be 65 points (2.17 seconds) for range or 35 points (1.17 seconds) for range rate. The start of the spike was assigned as follows: (1) for range - when a data point was more than three meters past the threshold average of the last seven points; (2) for range rate - when a data point was more than 2 meters/second past the threshold average of the last seven points. The end of the spike was assigned when the data returned to within three meters (range) or 2 meters/second (range rate) of the moving average value. If the start and end of a spike were found within the allocated width of the spike (65 or 35 points, respectively) a linear interpolation was performed between the initial known valid point and the point that determined the end of the spike. An example of spike removal is shown in Figure 4.

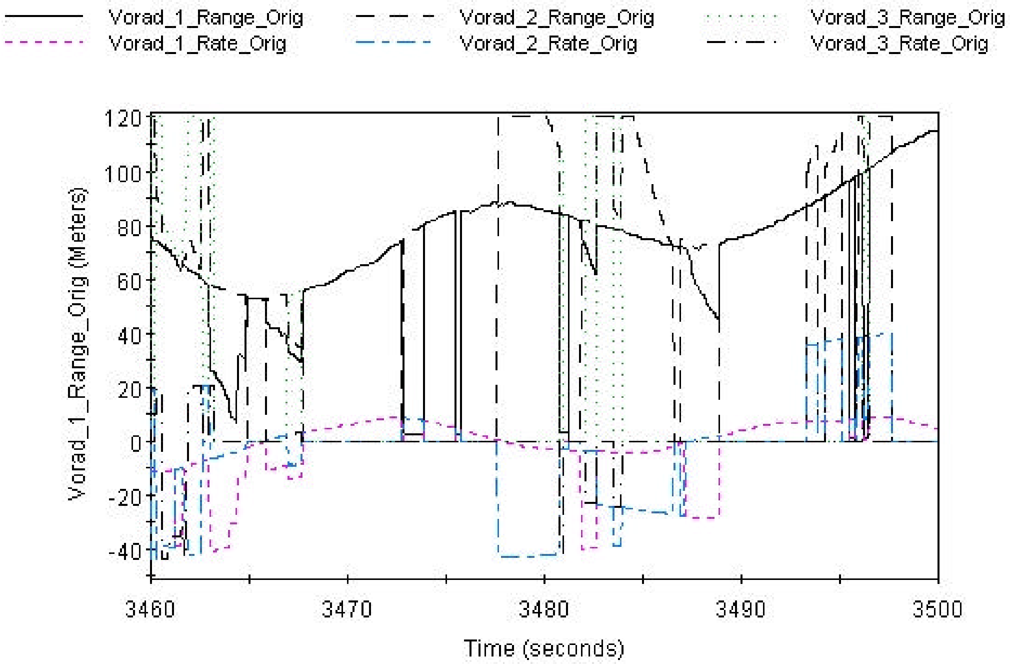


Figure 2. Raw Vorad data

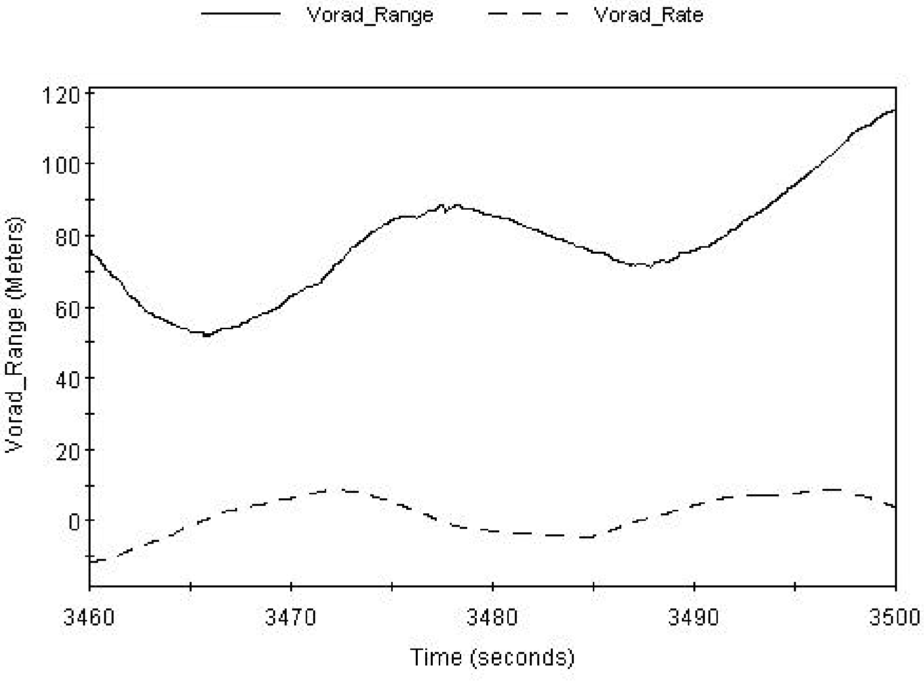


Figure 3. Processed Vorad data (combined)

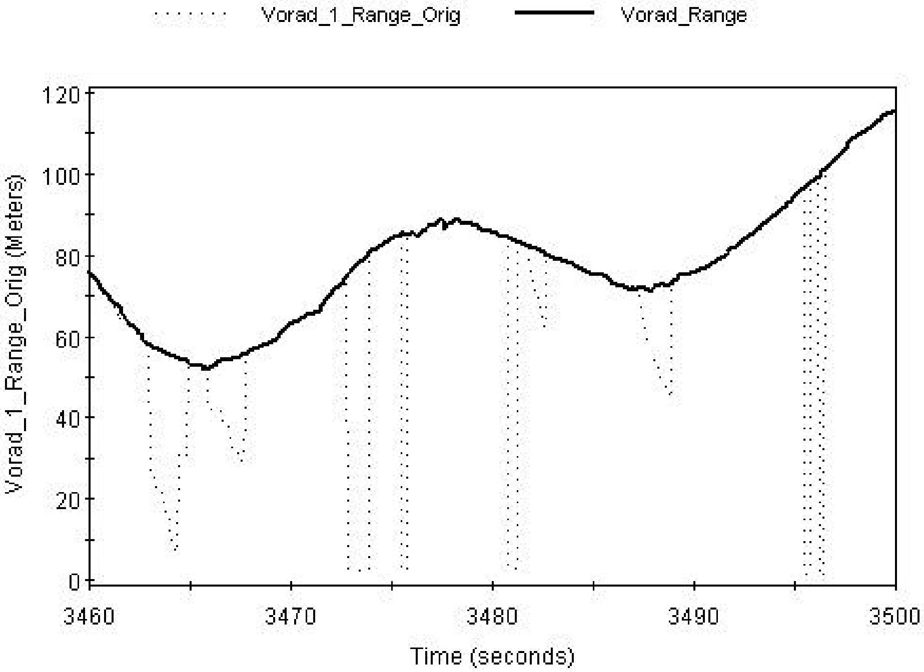


Figure 4. Example of the spike removal function to filter Vorad data.

In addition to the Vorad range rate measurement, two ground speed data channels were collected from each vehicle. *Speed_Event1*, as measured by the vehicle's transmission speed sensor, and *GPS_Speed1*, from the onboard GPS, were the channel names for the lead vehicle data; *Speed-_Event* and *GPS_Speed* for the SV. Each curve is read from the sensor at a sampling rate of 29.97 Hz. The *Speed_Event* curves change only in steps of 1 kph although the reading is updated faster than the sampling rate. The *GPS_Speed* curves have better accuracy but are not updated as frequently. The *GPS_Speed* data were time-shifted to match that of the *Speed_Event* data in order to remove the delay inherent in the *GPS_Speed* data. Filtering with a Butterworth low-pass phaseless filter smoothed both speed curves. After the filtering, the *GPS_Speed* curves were used for scaling the *Speed_Event* curves, using the mean ratio between the two curves for each vehicle. The data from both platforms were synchronized through use of the Global Positioning System (GPS) timing data.

The actual period, amplitude, and mean velocity (midway between the mean peak and the mean trough) were calculated for the lead vehicle's velocity curve, and the RMS deviation between the two vehicles' velocity curves was calculated. The data were then normalized, or *detrended*, by subtracting the mean velocity of the lead vehicle's curve from each point for both vehicles.

Following the example of Brookhuis and de Waard (Brookhuis et al., 1994) and using the equations derived in Bendat and Piersol (1986) the lead vehicle and SV speed curves were compared in the frequency domain to get *coherence*, *modulus*, and *phase*. For the Fourier transform, the curves were divided into at least two equal segments with a 50% overlap, with the actual number of segments depending on the total length of the test. The Hanning window was applied to each segment and then the segments were zero-padded before performing the discrete Fourier transform in order to standardize the spacing of the entries in the resulting frequency table at intervals of 0.005 Hz.

The coherence is a weighted average over a band of frequencies which includes the lead vehicle's primary frequency of approximately 0.04 Hz, using the power spectrum density of the lead vehicle as a weighting factor: For each frequency included, the weight is Pxx(f)/sum(Pxx(f)), so the weighted coherence is

$$\boldsymbol{coh} = \mathrm{sum}(\mathrm{coh}(f)*\mathrm{Pxx}(f))/\mathrm{sum}(\mathrm{Pxx}(f)).$$

To determine which frequencies to include, we truncated the series on both sides of 0.04 Hz to exclude all frequencies beyond the first entry found for which coh(f) is less than 0.35. Also, the series was further truncated after the first summation if any of the weights were found to be less than 0.005. These cut-off values are arbitrary, but experience has shown that terms outside these limits are too small to contribute significantly to the sum.

Following Bendat and Piersol (1986), the modulus is the sum of the ratios of the magnitude of the cross-spectral densities to the power spectrum density of the lead vehicle, that is,

$$\mathbf{M} = \mathrm{sum}(|\mathrm{Pxy}|/\mathrm{Pxx})$$

and the phase is the arctangent of the ratio of the sums of the imaginary and real components of the cross-spectral density of the two cars' speed curves,

$$f = \mathrm{atan}(\mathrm{sum}(\mathrm{IPxy})/\mathrm{sum}(\mathrm{RPxy}))$$

which was converted from radians to seconds by multiplying by the period of the lead vehicle's speed curve and dividing by 2π. The sums in the equations for modulus and phase were done over the same frequency interval used for the weighted coherence.

Error estimates were computed from the following equations:

$$s_{\mathrm{coh}} = (1 - \boldsymbol{coh}) * (2 / (\boldsymbol{nw} * \boldsymbol{nf} * \boldsymbol{coh}))^{1/2}$$

$$s_{\mathrm{M}} = ((1 - \boldsymbol{coh}) / (2 * \boldsymbol{nw} * \boldsymbol{nf} * \boldsymbol{coh}))^{1/2}$$

$$s_{f} = s_{\mathrm{M}} * \mathrm{Period} / (2\pi)$$

where ***nw*** is the number of windows, or segments, used in the discrete Fourier transform, and ***nf*** is the number of frequencies in the band used in the weighted coherence calculation.

3.6.4. Steering Reversals and Holds

Hand wheel position data were processed to create a hand wheel rate channel, which was used to calculate steering reversals and holds. The hand wheel position data were first filtered using a 4-pole Butterworth low pass phase-less digital filter with a 2 Hz cutoff frequency (100 pad points). The filtered position data were then differentiated to derive the hand wheel rate data.

The start of a steering hold was defined when the hand wheel rate remained within +/- 1 deg/s for a minimum of 400 ms. The end of the hold was defined when the hand wheel rate exceeded the hold value by +/- 1 deg/s. The +/- 1 deg/s threshold acts as a dead band that compensates for noise in the derived rate channel. This value was chosen based on the signal-to-noise ratio of the data.

Steering reversals were determined using both the hand wheel rate and position data. A reversal was defined when the hand wheel rate passed through the dead band of +/-1 deg/s and the hand wheel position magnitude changed more than 2 degrees.

3.6.5. Peripheral Detection Task (PDT) Data

PDT data were used to compute response times for each light activation trial. The data were grouped by driving trial. The number of light activations during the trial, which varied with the duration of the trial, and the number of correct detections were used to compute the probability of a correct response. Response times were computed for each correct detection and means and standard deviations were computed for each secondary task trial.

3.6.6. Eye Tracking Data Reduction

Data from the videotape were analyzed by an automated fixation analysis software package (FAST, EL-MAR Inc.). The automated analysis system provides two streams of data. The first is associated with eye movements and pupil dynamics (i.e., saccades, pupil-diameters, blinks) while the second is associated with the fixation behaviour on objects in the participant's field of view. For each participant, visual scanning data associated with each trial were first identified on the tape. This was done by detecting codes predefined to indicate the beginning and the end of each trial (i.e., trial duration) that were sent to the VISION 2000 from the instrumented vehicle during the test.

In order to determine the visual scene elements fixated by the participant, the automated analysis system calculates the intersection of the line-of-sight vector with the observed scene. To facilitate this analysis three distinctive reference targets were placed along the car's dashboard in a way that maximizes the probability that for the expected head movements at least two reference targets will appear within the field-of-view of the scene camera in each video. The analysis procedure starts by manually identifying the reference markers and the objects-of-interest (Forward View, Peripheral Detection Task, and AutoPC) in a single video frame. The video frame is captured from the recorded videotape by a frame grabber that is part of the FAST system. Following this initial procedure the analysis software tracks automatically the coordinates of the reference targets in all subsequent video frames. Using the reference targets as point correspondences, the automated analysis software calculates the point-to-point mapping between any video frame and the initial frame (i.e., the coordinates of the objects-of-interest in each video frame are calculated). Using the calculated boundaries of the objects-of-interest and the eye-position data from the eye-tracker, the system calculates fixation statistics on objects of interest, including fixation time, fixation frequency and the time-line data. Fixation time on a specific object is defined as the total time spent within the boundaries of the object. Fixation frequency on a specific object is defined as the number of times that this object is scanned and re-scanned by the driver. The time line data provide a list of time-intervals that each object was viewed. In this study the above parameters were calculated for three objects-of-interest: the Forward View, the area associated with the Peripheral Detection Task and the area associated with the AutoPC.

The number and amplitude of saccades within each object were also calculated for each trial, both for the entire duration of the trial and for each 5-second interval within the trial. Saccades were detected automatically by identifying eye-movement sequences with absolute peak-velocities that exceed 30 deg/s. Use of this threshold guarantees that saccadic eye movements with amplitude larger than 1 degree will be easily detected (mean peak velocity for 1 degree saccade is 50 deg/s). The beginning and end of each saccade were determined by the two first zero-crossings to the left and right of the peak-velocity. If the duration of the detected saccade was less than 120 ms (larger than the duration of a 30 degrees saccade) the sequence was identified as a saccade and was included in the statistical analysis. Eye velocity estimates were obtained by differentiating the eye-position data with a 5-point FIR differentiator. The differentiator has a bandwidth exceeding 20 Hz. In this study the number of saccades and the mean amplitude of saccades in each interval (sub-interval) were calculated for three objects-of-interest: the Forward View, the area associated with the Peripheral Detection Task and the area associated with the AutoPC.

4.0 DATA ANALYSIS

4.1. Data Analysis Approach

The data were analyzed using PROC GLM, in Version 8.02 of SAS. For most performance measures, a mixed design was utilized to compute analyses of variance (ANOVAs). Subject Group was the between-subjects' factor. Within-subjects' factors included Interface (visual/manual and voice), the Secondary Task (none, continuous baseline, simple, complex). The main hypotheses concerning differences between the interfaces were addressed through examining the main effects of Interface and the Secondary Task x Interface interactions. The ANOVAs were computed using the multivariate approach, which allows for the testing for non-sphericity, the absence of homogeneity of variances of difference scores in experimental designs involving repeated measures (Myers & Well, 1991). This problem arises because of expected correlations among scores when subjects are used as their own controls. We applied the Huynh-Feldt adjustment to the degrees of freedom when the data exhibited this problem, as recommended by Myers and Well (Myers &Well, 1991). In order to perform the multivariate analyses, data were grouped by Secondary Task categories so that there was a balanced design. It should also be noted that most of the performance measures were not normally distributed. Opinions vary about how to address this situation. Many researchers transform their data to improve normality, but most who adopt this approach also interpret differences on the original scale. This is not appropriate since once the data have been transformed, differences tested cannot be related to the original performance measure. Because the F test in the analysis of variance is robust to moderate deviations from normality, in fact, some have found that the F test works better for skewed distributions, we decided not to transform the data so that we can interpret differences in their original scales. This approach necessitates a caution, particularly for differences that tests show to be marginally significant. Such differences may not reflect real differences.

5.0 RESULTS

Five categories of data were collected, including: (1) vehicle control measures, (2) PDT performance measures, (3) secondary task measures, (4) eye-movement measures, and (5) workload ratings. Usable data were obtained from 635 (94%) of the 672 (21 subjects x 32 trials/subject) trials run. Results of analyses in the above categories are preceded by a discussion of the findings concerning differences between subject groups.

5.1. Subject Groups

We had planned to use TRC test drivers exclusively as participants in this experiment. However, after looking at preliminary results from several subjects and reviewing experimenters' comments, it became clear that many of these subjects were not able or willing to perform the tasks as instructed. Specifically, due in part to their training and indoctrination concerning test track safety procedures, which require test drivers to adopt and maintain long following distances, the TRC test drivers found it difficult to follow the lead vehicle closely enough to respond appropriately to the speed changes of the lead vehicle. In addition, two TRC test drivers declined to participate because they were not comfortable performing the three requested tasks simultaneously. At this point, we decided to recruit subjects from the Vehicle Research and Test Center (VRTC) engineering staff. Members of this group only occasionally drive on the TRC test facilities and were thus expected to approach the driving tasks more like real-world driving tasks.

The grouping was not planned, but because preliminary analyses showed consistent differences between the groups, including longer following distances, generally poorer target detection performance, and slower secondary task performance, we concluded that we could not combine data from the two groups. We therefore decided to include the subject group as a factor in the experiment. We believe that the differences between subject groups provide insights about individual differences relevant to driving performance. In particular, our data reveal consistent differences between the subject groups, reflecting better performance among Group 2 drivers. These results will provide guidance for the selection of test participants for future studies of this type.

5.2. Car-Following Performance

Subjects were instructed to select and maintain a single following distance during their participation in the experiment. They were also instructed that car-following was their first priority among the three tasks performed concurrently. Analyses of car-following performance involved use of two categories of performance measures. The first category included the means of car-following distance, referred to as range. The second set of measures is derived from the transfer and coherence analysis described by Brookuis and colleagues (1994). These results are presented in subsequent sections.

5.2.1. Car-following Distance (Range)

We computed an ANOVA on mean following distance using the statistical model outlined above. We found differences between subject groups, *F*(1,20) = 24.11, p = .0001. Specifically, Group 1 drivers drove at longer following distances (M = 86.3 m, SD = 26.2), than Group 2 drivers (M = 49.9 m, SD = 14.5). The main effect of Secondary Task was significant, *F*(3,57) =

27.43, p = .0001, as was the interaction between Secondary Task and Interface, *F*(3,57) = 8.65, p = .0017. This latter effect is presented in Figure 5.

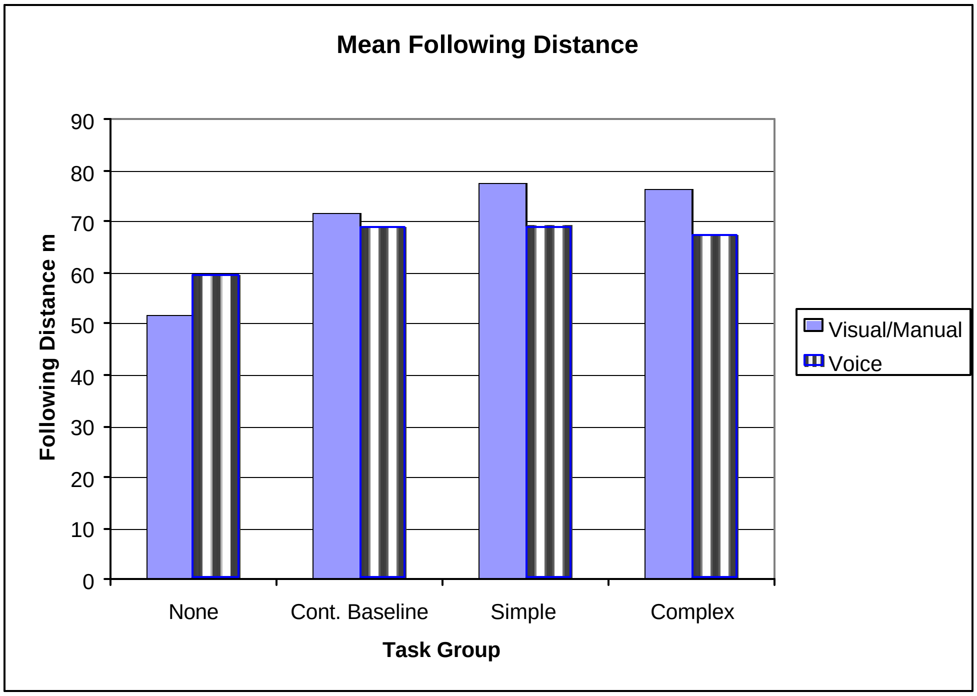


Figure 5. Mean following distance by Interface and Secondary task

Figure 5 shows that subjects generally drove at longer following distances when performing any secondary task, relative to the trials without secondary tasks (Secondary task = none). For simple and complex tasks, subjects drove at longer following distances when using the visual/manual interface than when using the voice interface.

5.2.2. Car- Following Coherence, Phase Shift, and Modulus

Coherence is a measure of squared correlation, representing the percentage of lead-vehicle speed variation accounted for by the following-vehicle speed. Coherence values vary between 0 and 1, such that larger values represent stronger correlations and thus better car-following performance. We computed coherence and the associated measures of phase shift, and modulus for each trial, following the approach used by Brookhuis, de Waard and Mulder (1994). Coherence was correlated with inter-vehicle spacing (r = -.75, p < .0001), reflecting the fact that drivers tracked lead-vehicle speed changes more accurately when following closer behind the vehicle. Accordingly, more than half of the variance associated with coherence ($r^2 = 0.56$) is attributable to the following distance. We found that mean following distance was also correlated with the phase shift (delay), although the correlation was not as strong (r = .22, p = .0001). Approximately 5% ($r^2 = 0.048$) of the variance in car following delay was determined by following distance. Details of the analyses of the individual measures are presented in the following sections.

We computed an ANOVA on coherence using the model described above. We found significant differences between subject groups, *F*(1,19) = 20.62, p = .0002. Group 2 drivers followed the lead vehicle speed changes more accurately (M = 0.93, SD = .05) than Group 1 drivers (M = 0.74, SD = .19). The main effect of Secondary Task was significant, *F*(3,57) = 15.58, p = .0001, reflecting the finding that performing any secondary task reduced coherence. This effect is shown in Figure 6.

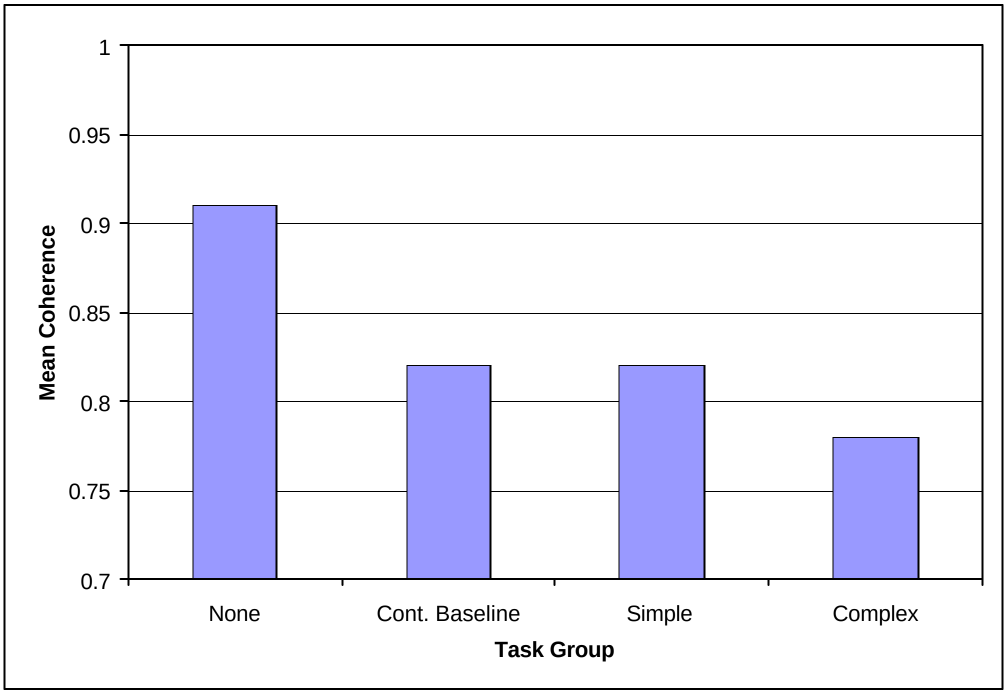


Figure 6. Mean coherence in car following by Secondary Task

Drivers followed the speed changes of the lead vehicle most attentively when there was no secondary task. Their car-following performance suffered most when performing complex tasks, with the simple and continuous baseline tasks having values between these two extremes.

The overall differences between interface conditions were not significant, *F*(1,19) = 2.18, p = .16, however the Secondary Task x Interface interaction was significant *F*(3, 57) = 4.02, p = .02. This effect is shown in Figure 7.

Figure 7 shows that drivers performed car following better when using the voice interface for the simple and complex tasks, reflecting a potential, but small benefit of the voice interface. Differences between the interface conditions for the baseline and no secondary task conditions were smaller and in the opposite direction.

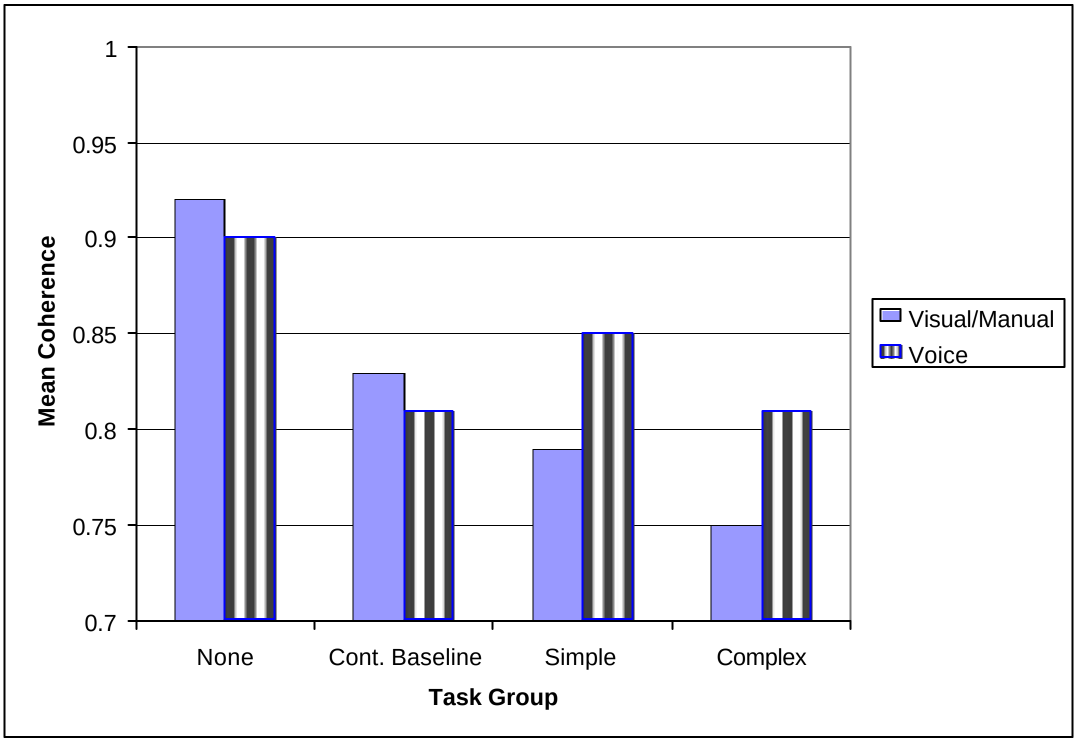


Figure 7. Mean coherence in car following by Secondary Task

We computed an ANOVA for the phase shift or delay, which was the main performance measure used by Brookhuis, de Waard and Mulder (1994). Generally, effects were weaker in our analysis, reflecting the wide range of coherence values, particularly among Group 1 drivers. The main effect of Secondary Task was significant, *F*(3,57) = 3.56, p = .038, as was the Secondary Task x Subject Group interaction, *F*(3,57) = 4.24, p = .022. Means for the Secondary Task main effect are shown in Figure 8. Means for the Secondary Task x Subject Group interaction are shown in Figure 9.

The delay associated with the No Secondary Task condition (M = 3.14 s, SD = 1.36) was significantly shorter than that associated with the three conditions involving secondary tasks (Cont. Baseline: M = 3.76 s, SD = 1.24; Simple: M = 3.52 s, SD = 1.27; Complex: M = 3.73 s, SD = 1.16). There were no differences between interface conditions for this measure, *F*(1,19) = .25, p = .62.

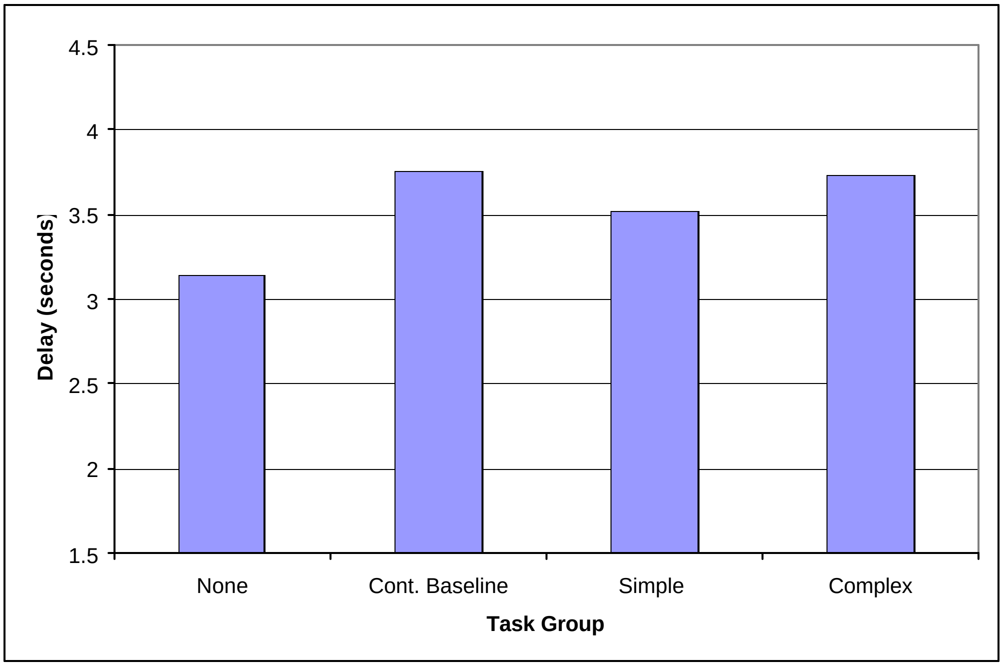


Figure 8. Mean car-following delay by Secondary Task

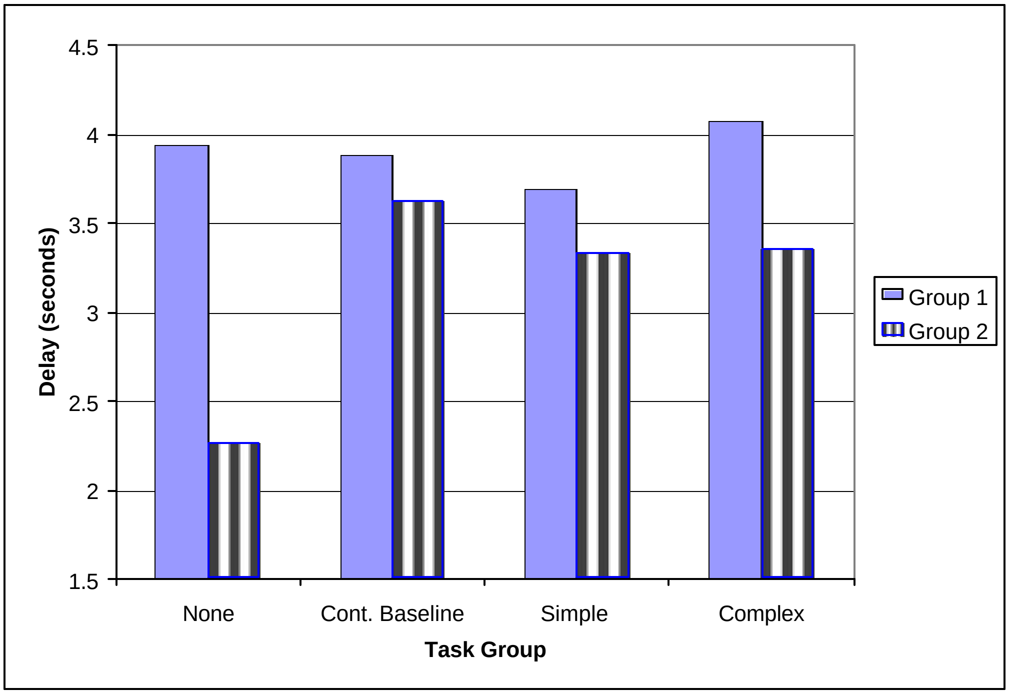


Figure 9. Mean car-following delay by Secondary Task and Subject Group

Because Group 1 drivers were less engaged in car following, the delay measure was considerably less sensitive to differences between secondary task categories than for Group 2 drivers.

A separate analysis was computed using data from Group 2 drivers only, because of their better car-following performance. The main effect of Interface was not significant, *F*(1, 9) = 0.32, p = .58. The Secondary Task main effect was significant, *F*(3,27) = 18.83, p = .0001 (see Figure 10). The Secondary Task x Interface interaction was significant, *F*(3,27) = 4.18, p = .015, however this effect primarily reflects the difference between interface conditions in the No Secondary Task condition, which are not readily interpretable (see Figure 11).

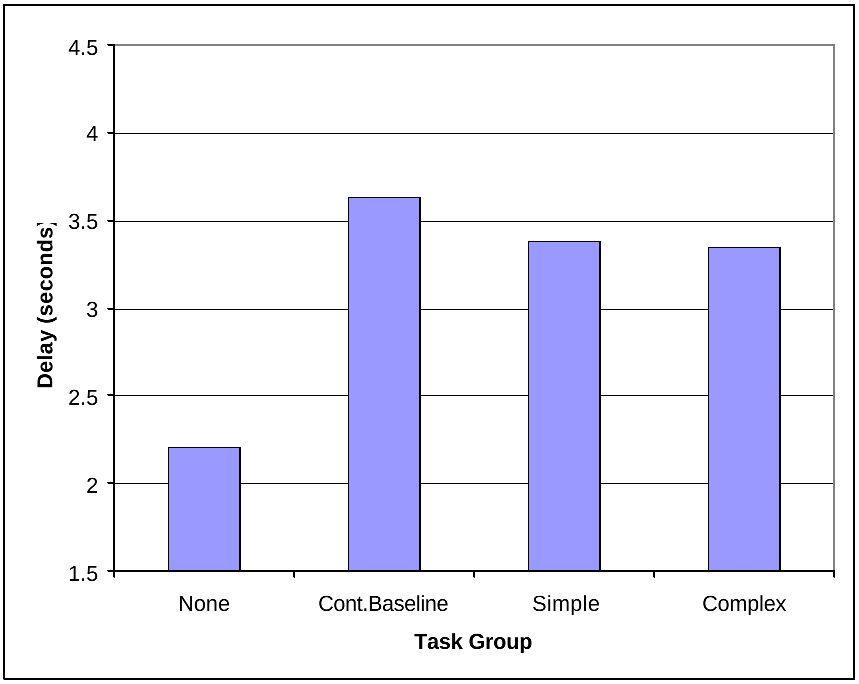


Figure 10. Mean car-following delay by Secondary Task (Group 2 drivers only)

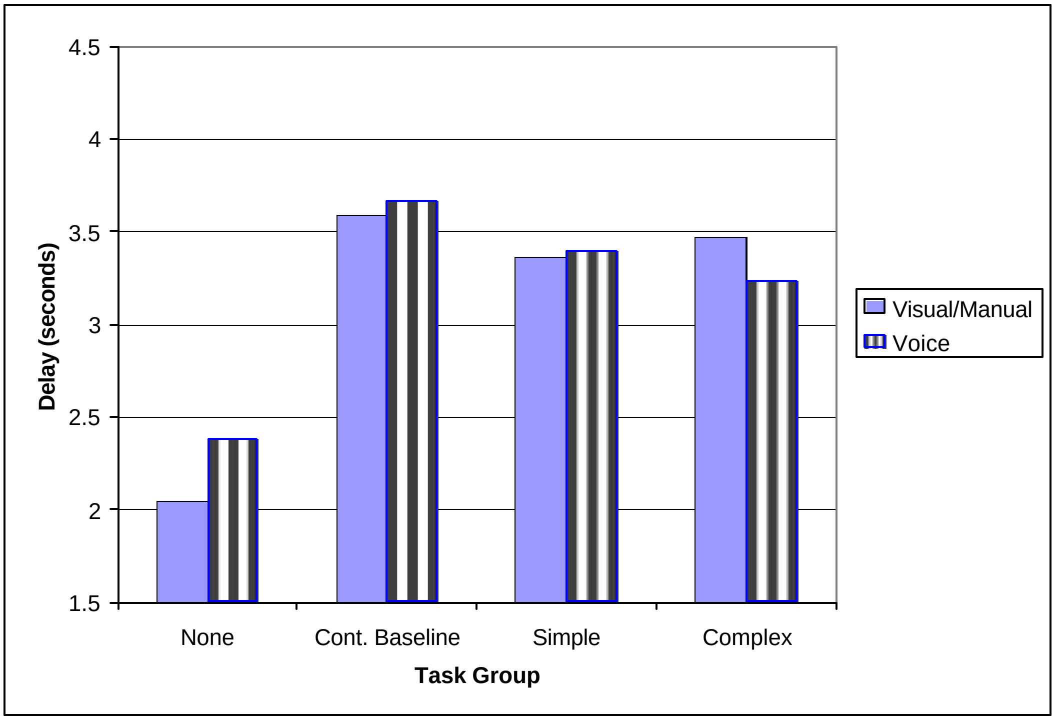


Figure 11. Mean car-following delay by Secondary Task and Interface (Group 2 drivers only)

Modulus is the third and final parameter of the coherence model of car following performance. The modulus is the output/input or gain associated with the two signals. Modulus values significantly greater than one indicate overreaction by the following driver to changes in the lead vehicle speed. In close following situations, this could lead to problems if the lead vehicle stopped suddenly. Similarly, extremely low modulus values indicate that the following driver is under-correcting in response to lead vehicle speed changes. In the present experiment, the pattern of modulus values reflected the fact that the Group 1 drivers did not follow close enough to detect changes in the lead vehicle speed. As such, the relatively low modulus values for Group 1 drivers are considered not meaningful. Therefore, an analysis was done using data from Group 2 drivers only. The main effect of Secondary Task, *F*(3,27) = 3.09, p = .044, reflected significantly higher modulus values for the No Secondary Task condition (M= 1.01, SD = 0.17), relative to the three conditions involving secondary tasks (Cont. Baseline: M = 0.88, SD = 0.22; Simple: M = 0.89, SD = 0.25; Complex: M = 0.94, SD = 0.46). The means are shown in Figure 12.

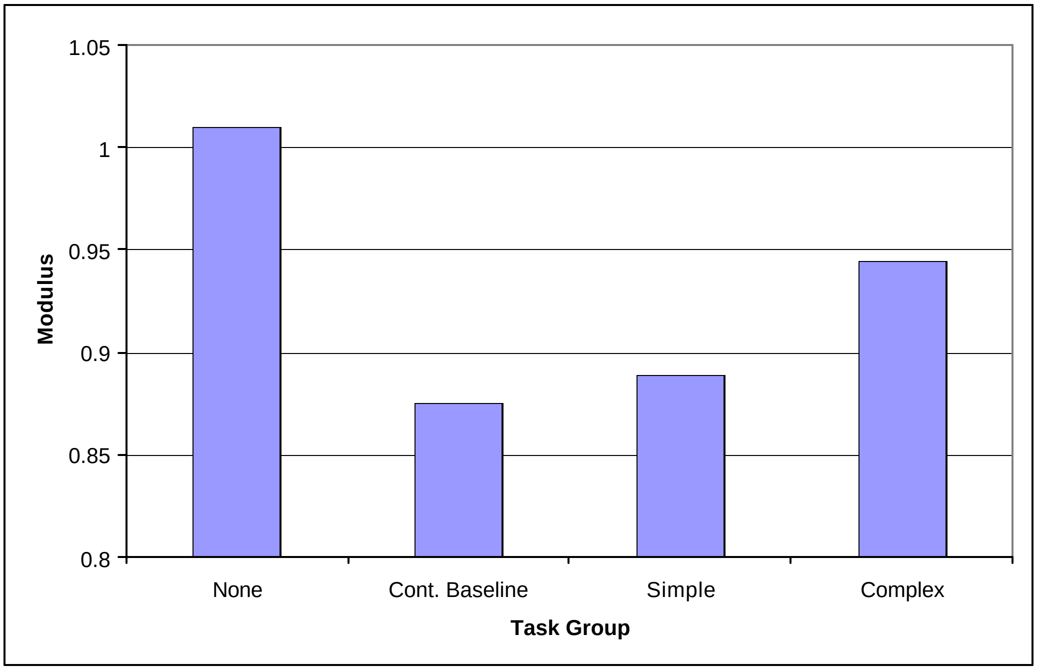


Figure 12. Mean car-following gain (modulus) by Secondary Task (Group 2 drivers only)

Group 2 drivers did not exhibit significant over-correction in any condition. The mean value for the No Secondary Task condition, which is very close to one, indicates that the drivers were tracking the extreme values of the lead vehicle speed quite accurately. While the results do indicate impairment associated with secondary task performance, the ordering of secondary task means is not consistent with the hypothesis that the complex tasks created more interference than the simple or continuous baseline tasks. There were no differences between the interface conditions for this measure, *F*(1,9) = 1.01, p = .34.

5.3. Lane Maintenance Performance

We computed the minimum, maximum, mean and standard deviation of lane position for each trial. Trials with extreme values (lateral position > 180 cm from lane center), indicative of lane changes or data loss, were eliminated from the analysis. Acceptable data remained for 625 of 642 (97%) trials. An ANOVA was computed using standard deviation of lane position as the dependent measure. Group 2 drivers exhibited less lane position variability on average (M = 19.4 cm, SD = 5.7) relative to Group 1 drivers (M = 24.2 cm, SD = 8.8), however this difference was marginally statistically insignificant, *F*(1,18) = 4.17, p = .056. Secondary tasks performed with the voice interface were associated with less lane position variability (M = 20.9 cm, SD = 7.5) than those performed with the visual/manual interface (M = 22.8 cm, SD = 7.9), *F*(1,18) = 15.15, p = .001. This trend was observed in all secondary task conditions as shown in Figure 13.

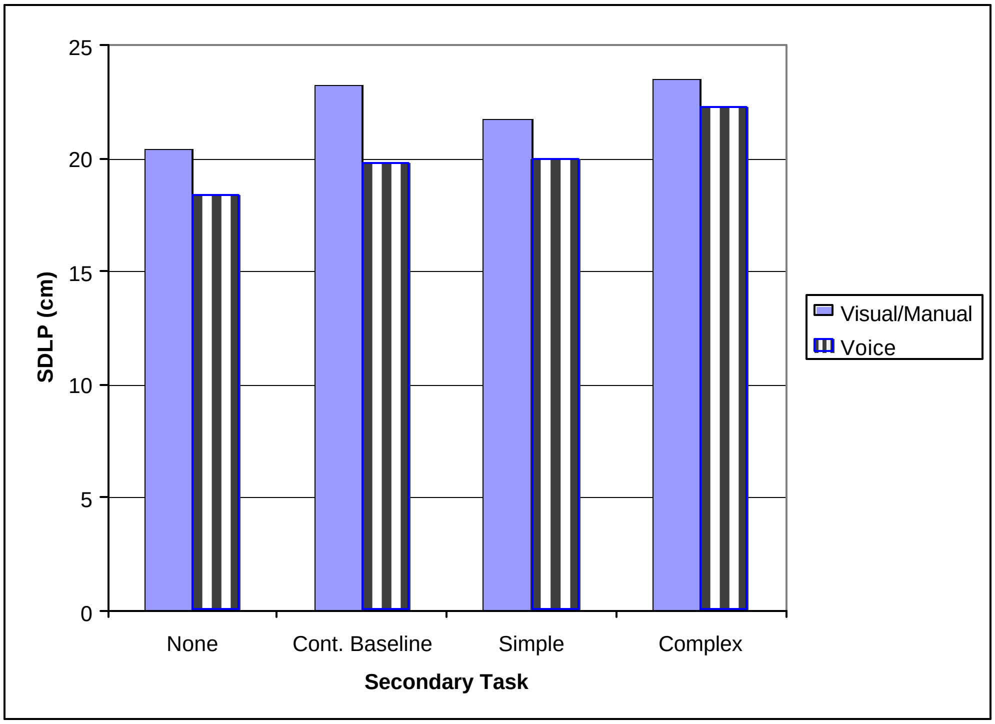


Figure 13. SDLP by Interface and Secondary Task

5.4. <u>Steering Performance</u>

A steering reversal was defined to begin when the steering velocity left a zero-velocity dead band and ended when the steering velocity entered a zero-velocity dead band such that the magnitude of the reversal was 2 degrees or greater (Tijerina, Kiger, Rockwell, & Tornow, 1995). Reversal rates (reversals/s) were higher for trials involving secondary tasks, relative to the No Secondary Task condition, *F*(3,57) = 8.39, p < .0001, reflecting higher total task workload when secondary tasks were performed (see Figure 14). There was no consistent difference between interface conditions, *F*(1,19) = 0.02, p = .89, however the Interface by Secondary Task interaction approached significance, *F*(1,19) = 3.57, p = .07, reflecting slightly higher mean values for visual/manual interface in the three conditions involving secondary tasks versus a small difference in the opposite direction for the No Secondary Task trials.

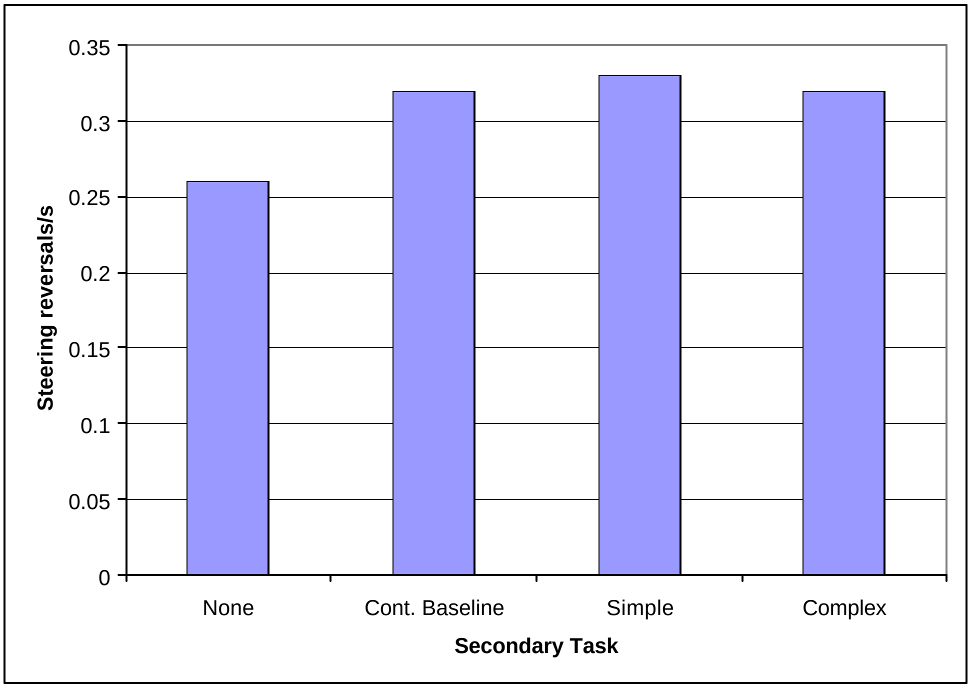


Figure 14. Steering reversals by Secondary Task

Steering holds are defined as periods of at least 400 ms involving no steering activity. MacDonald and Hoffman (1980) have interpreted steering holds as evidence of withdrawal of attention from the steering task, presumably reflecting increased secondary task demands. The correlation between reversals/s and holds/s ($r = -.55$, $r^2 = 0.30$) indicates that the measures reflect largely different behaviors. The visual/manual interface condition (M = 0.44 holds/s, SD = 0.13) was associated with more steering holds/s than the voice condition (M = 0.40 holds/s, SD = 0.11), *F*(1,19) = 4.31, p = .052. The No Secondary Task trials were associated with more steering holds/s than the trials involving secondary tasks, *F*(3,57) = 3.84, p = .037, which is not consistent with predictions of MacDonald and Hoffman (1980).

5.5. Peripheral Detection Task

PDT data were obtained continuously during the portion of each straight road segment in which subjects performed secondary tasks and during baseline driving segments, which generally included an entire straight segment of the closed course. On average there were 20.9 (SD = 9.1) target (LED) activations during a trial, which lasted approximately 132 seconds, on average.

Group 2 subjects detected a larger proportion of targets (M = .72, SD = .24) than Group 1 subjects (M = .62, SD = .21), *F*(1, 20) = 3.87, $p < .06$. The marginally non-significant statistical result reflects relatively large variation among subjects. The main effect of interface was statistically significant, *F*(1,19) = 5.81, $p < .03$, reflecting the finding that subjects detected more targets while performing secondary tasks using the voice interface (M = .71, SD = .20) than they did while using the visual/manual interface (M = .62, SD = .24). The main effect of Secondary Task was also statistically significant, *F*(3,57) = 40.86, $p < .0001$. Subjects were most likely to detect targets (M = .87, SD = .10) when not performing a secondary task. Subjects were least likely to detect targets while performing the simple secondary tasks, (M = .51, SD = .23).

Intermediate values were associated with the complex tasks (M = .63, SD =. 20) and the continuous baseline tasks (M = .64, SD = .19). The interaction between Secondary Task and Interface was statistically significant, *F*(3,57) = 3.59, p < .02. This effect is shown in Figure 15. Subjects detected proportionately more targets while using the voice interface in all task conditions. Differences between interface conditions were greatest for the simple and complex tasks (.15 and .10, respectively) and least for the continuous baseline tasks and the driving segments without secondary tasks (.04 and .05, respectively).

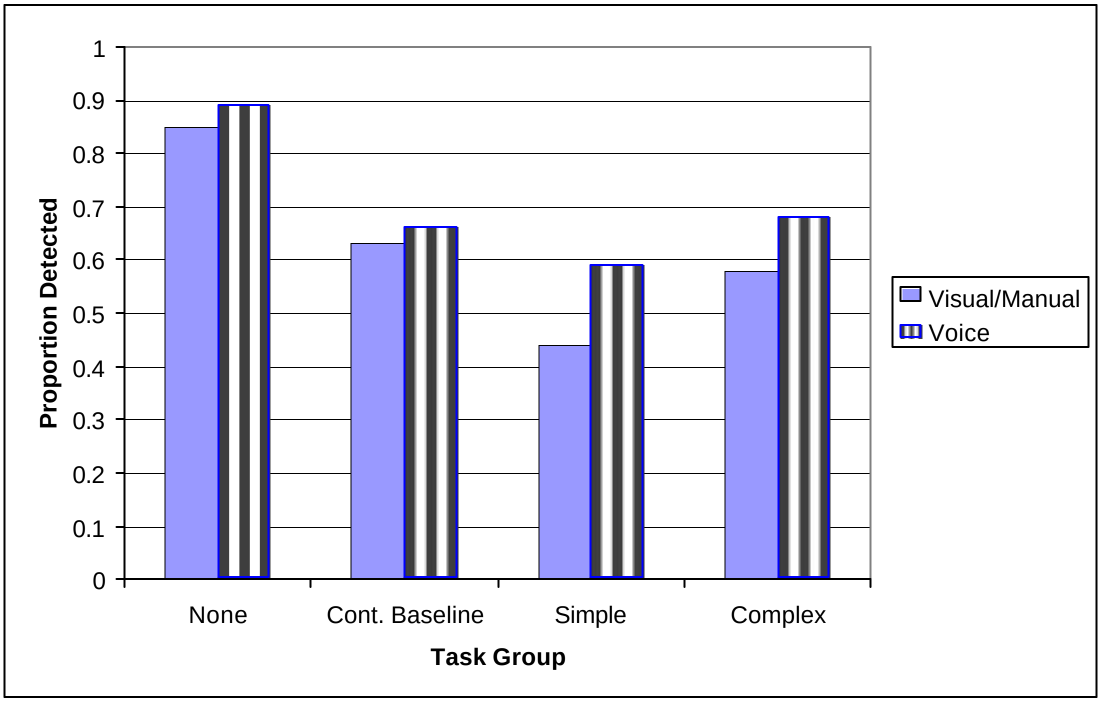


Figure 15. Proportion of PDT targets detected by Secondary Task and Interface

We computed an ANOVA for the target-detection response times. Response time data were grouped by trial so that each data point represented the mean RT for the respective driving trial. Response times were significantly faster among the Group 2 drivers, relative to the Group 1 drivers (M = .77 vs. M = .95, respectively), *F*(1,19) = 5.94, p < .02. Drivers responded more quickly to PDT targets while performing secondary tasks using voice interface (M = .80 s, SD = .32) than in the visual/manual interface condition (M = .87 s, SD = .24), *F*(1,19) = 7.05, p < .02.

The interaction between Secondary Task and Interface condition was not statistically significant, F(1,19) = 2.79, p = .086.

5.6. Secondary Task Performance

Task completion times were computed for each secondary task. The overall mean completion time was 132.8 s (SD = 46.2 s). Group 1 subjects were slower in completing secondary tasks (M = 140 s, SD = 47.9) than Group 2 subjects (M = 116.5 s, SD = 44.6), *F*(1,20) = 15.69, p = .0008. Subjects took longer to complete complex tasks (M = 148.9 s, SD = 48.0) than simple (M = 71.0 s, SD = 33.0) or continuous baseline tasks (M = 117.5 s, SD = 43.7), *F*(2,38) = 227.0, p < .0001. There were no differences between Interface conditions, *F*(1,18) = 0.36, p = .55.

The continuous baseline tasks required subjects to either dial a sequence of phone numbers or tune the radio to a sequence of radio stations. The use of a sequence was intended to extend the task so that it lasted long enough for the coherence car-following measures to be computed. The in-vehicle experimenter recorded the number (to the nearest half) of phone numbers and radio stations completed during the data collection interval.

Using the number of phone numbers completed and the total task time, we computed the average time required to dial each phone number. Dialing time was considerably longer in the voice condition (M = 59.3 s, SD =28.4) than in the visual/manual condition (M= 21.2 s, SD = 6.3), *t*(46) = 20.4, p = .0001. This difference, particularly the large standard deviation in the voice condition, reflects the considerable difficulty experienced by some subjects with the AutoPC voice recognition. Subjects were able to dial familiar numbers (M = 20.0 s, SD = 4.7) slightly more quickly than unfamiliar numbers (M = 22.2 s, SD = 7.3) in the manual condition, but this difference was not statistically significant, *t*(27) = 2.42, p = .13. In the voice interface conditions, subjects dialed familiar numbers (M = 65.5 s, SD = 16.5) slower than unfamiliar numbers (M = 54.9 s, SD = 34.7) on average, with the difference approaching statistical significance, *t*(17) = 4.46, p = .06.

We computed the time required to tune a single radio station for each trial. Radio tuning was faster in the visual/manual condition (M = 16.4 s, SD = 4.8) than in the voice interface condition (M = 22.1 s, SD = 8.3), *t*(74) = 3.02, p = .001.

5.7. Subjective Workload Ratings

Subjects completed workload ratings (NASA –TLX) at four times during each session: (1) following practice with the interface in a stationary vehicle; (2) when the eye tracker was initially put on and calibrated, (3) when the eye tracker was taken off for a rest, and (4) at the end of the session. Accordingly, the first rating was considered a practice. The second rating relates to completion of the tasks without the eye tracker, while the third rating relates to the completion of tasks with the eye tracker. The fourth set of ratings was a summary measure, intended to refer to the entire data collection session.

We created a single composite measure for each rating and subjected this to analysis of variance. There were no differences in ratings between interface conditions, including interactions. The only difference observed was between the four replications, *F*(3,54) = 6.57, p = .0007. Examination of the means revealed that the largest difference was between trial one and the other trials. We eliminated trial one scores and repeated the analysis. The results indicated a significant difference, *F*(2,40) = 5.25, p = .01. Examination of the means revealed a marginally non-significant difference between the eye-tracker and no eye-tracker ratings, *F*(1,20) = 3.53, p = .07, such that the eye tracker trials were associated with slightly higher workload (M = 44.7, SD = 20.7) than the non-eye tracker trials (M = 42.6, SD = 20.5).

We also looked at the six individual scores that were combined to create the composite measure (Definitions are presented in Appendix C). Our focus was on differences between the two interface conditions. We found no significant differences between interface conditions on any of the individual scales.

5.8. Eye Tracking Data Analysis

Data obtained from the eye tracker were reduced under supervision of Transport Canada.

Eye tracking tapes were available for 16 (76%) subjects. Each subject completed 16 trials (8 per interface condition) while wearing the eye tracker, resulting in a total of 256 eye-tracking trials. Data for 117 of these trials (45.7%) were suitable for analysis. In these trials more than 95% of the eye-movement data was analyzed.

Most of the rejected trials (116 trials or 83%) occurred in sections of the track where the sun was interfering with either the eye-tracker performance or with the illumination of the reference targets (i.e., there was direct sun illumination of the subject's face). Twenty-three trials were rejected due to improper eye-tracker calibration or improper set-up of the eye-tracker on the driver's head.

Consequently, the eye tracking analyses are based on a subset of the total data set. Because there were four trials involving complex secondary tasks for each interface condition, this condition offered the greatest likelihood of obtaining at least one trial of usable eye tracking data. Specifically, at least one complex trial provided data for 10 subjects, four from Group 1 and six from Group 2. For conditions in which multiple trials were available, a single mean was created for each measure. Control condition (no secondary task) data were also available for this set of subjects. This approach to the data resulted in a workable data set appropriate for repeated measures analyses. Because the subset of trials included in the eye tracker data set is not representative of the larger main data set, it is likely that there may be inconsistencies in the results. Accordingly, the complete set of results is presented in Appendix F. A summary is presented here, which should be interpreted with caution, due to the differences between the two data sets.

5.8.1. Eye Tracking Analyses Performed

The analyses of drivers' visual behavior focus on three main types of questions. To address these issues the following analyses of visual behavior were performed as a function of interface type and driver group:

1. How long do drivers look off-road (at AutoPC) or at the roadway (forward view) while performing tasks? The analyses of Percentage of Task Time spent looking at regions of interest address these issues.

2. How often do drivers look at AutoPC (or out to the roadway) while carrying out the tasks? To address this issue, analyses were done on the number of glances made per 10 seconds to the specified areas while driving and performing secondary tasks.

3. How long are the glances that drivers make to the device (or out to the roadway) while performing the tasks? The Analyses of Glance Durations provide this information.

In addition, a series of analyses were performed examining drivers' visual behavior while carrying out the peripheral detection task (PDT).

5.8.2. Summary of Analyses for Percent Time Looking in Different Areas

Other than the finding that drivers spent more time looking in the forward view when not performing any in-vehicle task, the results of these analyses were not very informative as to the relative benefit/cost of either type of interface. Although some of the means were in the expected direction (i.e., Group 1 drivers spent more eyes-off-road-time (EORT) overall, and the voice-based interface affording the drivers more time in the forward view), none of the comparisons was statistically significant. Details are presented in Section 9.6.2.

5.8.3. Summary of Analyses of Number of Glances per 10 Seconds

The primary finding of these analyses was in the comparison between trials with and without a secondary task. Drivers made significantly fewer glances to the forward view when not performing an in-vehicle task, reflecting the fact that they did not have to alternate their attention between the forward view and the AutoPC. Although the Group 1 drivers made numerically more glances per 10 s to the AutoPC, this difference was not statistically reliable. Additional details are presented in Section 9.6.3.

5.8.4. Summary of Glance Duration Analyses

A glance was defined as the time from which a driver's direction of gaze moved toward a target, such as the AutoPC, to the moment it moved away from it. Consistent with ISO 15007 (2002), the glance duration includes the transition time to that target. Glance durations are of particular interest from a safety perspective in that long off-road glances (exceeding 2 s) are considered a safety hazard (e.g., Zwahlen et al., 1988). A number of analyses of glance duration were performed and are provided in Appendix F. The main findings are summarized here.

The analysis of glance durations to the AutoPC revealed a significant driver group by interface type interaction. The interaction indicated that Group 1 drivers using the visual/manual interface, on average, made the longest glances to the AutoPC (M = 1.01 s) and consequently took their eyes off the road for longer periods of time. Group 2 drivers' mean glance durations to the visual/manual interface were .72 s. In the case of the voice-based interface, mean glance durations were comparable for both groups (Group 1 M = .84 s; Group 2 M = .83 s).

Some of the most important safety-relevant data comes from the analyses of the distribution of glance durations. Overall, a significantly greater percentage of long glances (> 2 s) were made using the visual/manual interface (7.32%) than when using the voice-based interface (2.12%).

5.8.5. Percentage of Glances as a Function of Glance Duration- Visual/Manual Interface

Figure 16 shows the distribution of glance durations to the AutoPC in the visual/manual interface condition for both driver groups. A higher percentage of short glances (those up to .5 s) was made by Group 2 drivers (M = 31.58%) compared to Group1 drivers (M = 17.83%). The two groups were comparable in the percentage of midrange (.51-2 s) glances. Conversely, a greater percentage of long (exceeding 2 s) glances was made by Group 1 drivers (M = 13.78%) compared to Group 2 drivers (M = 3.00%). The glances exceeding 2 s are further broken down in the graph according to their duration. These data indicate that when using the visual/manual interface, Group 2 drivers made more short glances to the device than Group 1 drivers; however, Group 1 drivers made more long glances than Group 2 drivers. It is important to note that over half of the long glances made by Group 1 drivers exceeded 2.5 s.

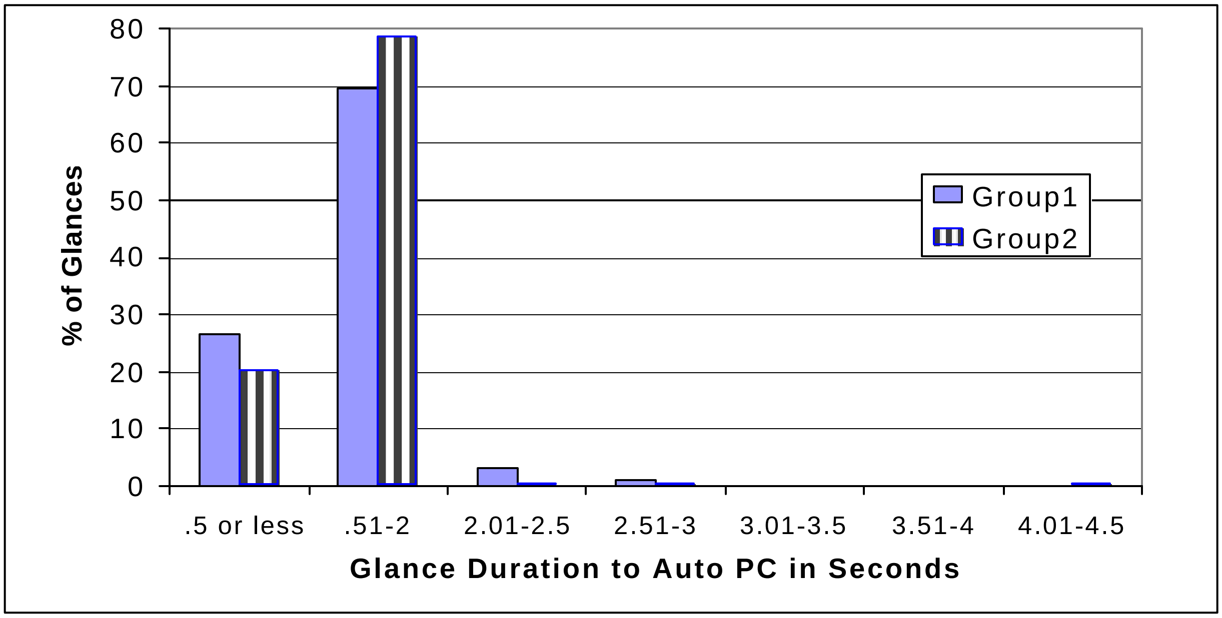


Figure 16. Distribution of glance durations to the AutoPC: visual/manual interface

5.8.6. Percentage of Glances as a Function of Glance Duration-Voice Interface

Figure 17 shows the distribution of glance durations to the AutoPC when the voice interface was used for both driver groups. The mean percentage of short glances (up to .5 s) made by Group 1 drivers was 26.76% versus 20.24% for Group 2 drivers. The percentage of midrange glances (between .51 and 2 s) made by Group 2 drivers (M = 78.60%) was greater than that for Group 1 drivers (M = 69.66%). Long glances (> 2 s) were made by both groups (Group 1 = 3.57%; Group 2 = 1.16%). As in the previous graph, glances exceeding 2 seconds are further categorized according to their duration. Here as with the visual/manual interface, both groups of drivers made glances exceeding 2 seconds, although they constituted a lower percentage of the overall glances. The examination of the distribution of maximum glance durations to the AutoPC revealed that long glance durations were a problem for both interfaces and both groups of drivers. When using the visual/manual interface, Group 1 drivers made more dangerously long glances than Group 2 drivers. However, when the voice-based interface was used, long glances were also observed, although not to such a great degree. These findings were unexpected. It was expected that use of the voice-based interface would free drivers from making long glances to the AutoPC.

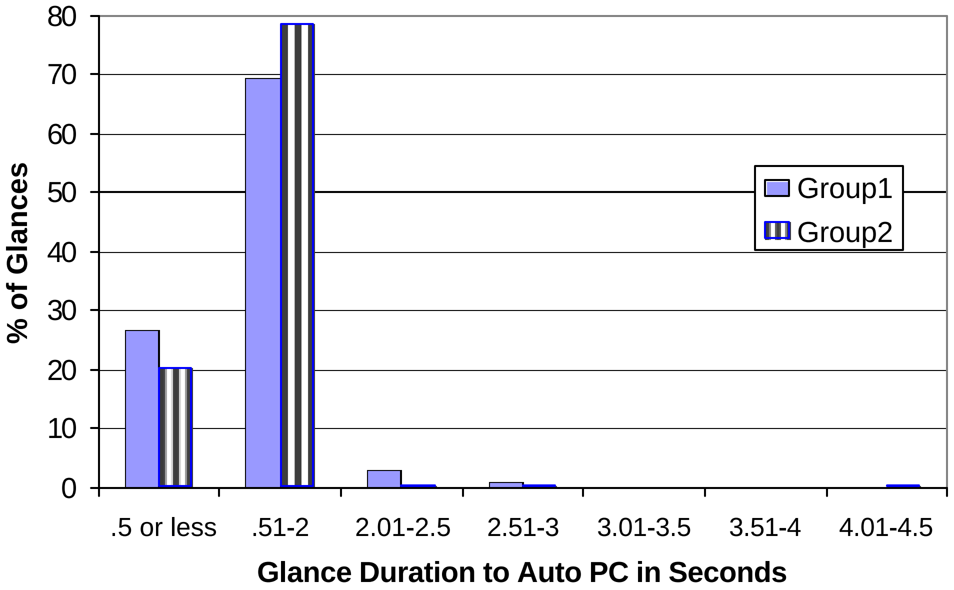


Figure 17. Distribution of glance lengths to the AutoPC: voice interface

5.8.7. Summary of Glance Behavior During the PDT Task

The two groups of drivers differed in the amount of time spent looking at the PDT and in the number of glances made (per 10 s interval) to the PDT. Group 1 drivers spent a greater percentage of task time looking at the PDT and made more glances (per 10 s interval) to the PDT. However, their greater amount of time looking at the PDT was not reflected in their performance. Group 2 drivers (eye-tracking sample) detected more of the PDT targets (M = .88) than Group 1 drivers (M = .69) and they were significantly faster at doing so (mean RT: Group 1 = .81 s, Group 2 = .63 s).

Compared with when they used the visual/manual interface, drivers using the voice-based interface, spent a greater percentage of their time looking at PDT, made more glances (per 10 s interval) to the PDT and made longer glances to the PDT. These results are consistent with the finding that drivers detected marginally more (p = .06) targets while performing secondary tasks using the voice interface (M = .80) than they did while using the visual/manual interface (M = .69). However, drivers were most likely to detect targets (M = .92) when not performing any secondary task.

5.8.8. Discussion of Visual Behavior Results

Most of the useful information concerning drivers' visual behavior and the AutoPC comes from the analyses of glance durations. Considering glances away from the road (to the AutoPC), use of the visual/manual interface was associated with longer mean glance durations for the Group 1 drivers and also with a greater percentage of glances exceeding 2 s for all drivers. These findings suggest that a visual/manual interface may be more distracting than a voice-based interface, at least under the conditions tested in this study.

However, the examination of the distribution of maximum glance durations raised concerns about both interface types. When using the visual/manual interface, both groups of drivers made dangerously long glances away from the road, although Group 1 drivers showed the strongest effect. However, when the voice-based interface was used, long glances were observed in both groups as well, although to a lesser degree. Reading a message, visually scanning message headers, and correcting mistakes appear to be the primary activities associated with these long glances. The safety implications of off-road glances exceeding 2 seconds are discussed in greater detail in Appendix F.

The data from the PDT analyses suggest that there may be some benefits associated with the voice-based interface. Compared with when they used the visual/manual interface, drivers using the voice-based interface spent a greater proportion of their time looking at the PDT, made more glances (per 10 s interval) to the PDT, and made longer glances to the PDT. This increased inspection of the PDT was reflected in their performance: greater and faster target detection with the voice-based interface. The safety implication of this finding is that the voice-based interface may lead to better event detection by allowing more heads up time.

Overall the analyses of visual behavior suggest that there are safety concerns associated with the use of both interfaces, although the voice-based interface appears to have some benefits.

6.0 DISCUSSION

The results indicate strong and consistent driving performance decrements associated with secondary task performance, relative to the driving-only condition. Despite instructions to maintain constant following distances, subjects drove at longer following distances when performing secondary tasks. The consistency of this effect suggests behavioral adaptation (Smiley, 2000) to increasing total task demands, however it is not possible to determine whether drivers deliberately increased following distances or whether the increased following distances resulted from increased neglect of the car-following task, as drivers shifted their attention to the secondary tasks. Nevertheless, the longer following distances were reflected in lower coherence scores. The car-following delay, represented by the phase shift between the two vehicle speed signals in the coherence analysis, increased by more than one second during secondary task performance, relative to the trials with no secondary task. This finding is noteworthy because the increased delay reflects slower response time over the entire secondary task interval (M = 132 s) in contrast to response times for discrete events, which characterize behavior over an interval of only several seconds. Increased steering reversal rates during secondary task performance suggest increased total task workload, while increased steering hold rate indicates withdrawal of attention from steering during secondary task performance (MacDonald & Hoffman, 1980). Secondary task performance also affected both lane keeping, as reflected by increased lane-position variability, and target detection, as reflected by reduced percentage of PDT targets detected.

The observed differences between interface conditions were less evident and generally weaker than differences between secondary task conditions, most notably for the car-following measures. Subjects drove at longer following distances while using the visual/manual interface for two of three secondary task conditions, relative to the voice interface, however this difference was considerably smaller than differences between secondary task conditions. There were no observed differences between interface conditions for any of the car-following performance measures (coherence, delay, gain). In contrast, the observed reductions in steering wheel holds and standard deviation of lane position (SDLP) reflect improved vehicle control performance in the voice interface condition relative to the visual/manual interface condition. One possible explanation is that the required use of the hands in the visual/manual interface condition created a conflict with the manual input demands of steering, which impacted measures of steering and lane-position performance. Accordingly, this conflict did not extend to car-following performance (i.e., speed control), which did not require use of the hands.

The results generally did not support the hypothesis that increasing task complexity would exacerbate distraction effects associated with the secondary tasks. Relative to the other secondary task categories, the complex tasks were associated with markedly poorer performance only for a single measure, coherence. Otherwise, the performance on trials involving simple secondary tasks was more likely to have differed from that for trials involving continuous baseline and complex tasks. This pattern was evident for the car-following delay, SDLP, and to a lesser extent for the proportion of PDT targets detected. This pattern may have been an artifact of differences in completion times associated with the simple secondary tasks relative to the other categories of tasks. However, this potential problem did not confound the finding that the performance of complex secondary tasks did not consistently interfere more with driving performance than did the baseline tasks of continuous radio tuning and phone dialing.

The overall pattern of results can be interpreted within the three-level hierarchy proposed to underlie cognitive control in driving (Michon, 1985). According to this model, driving consists of concurrent activity at three levels: (1) strategic, (2) tactical/maneuvering, and (3) operational/vehicle control. In the present study, measures of steering and lane position maintenance represent low-level operational/vehicle control behaviors, which are generally executed automatically and require minimal conscious attention. In contrast, car following represents a mid-level tactical or maneuvering behavior, which requires more active interpretation of visual cues and decision making. Within this context, the pattern of results suggests that performing the AutoPC secondary tasks influenced driver performance at both of these levels. However, the apparent benefits associated with the voice interface appear limited to the operational level, influencing steering and lane-position maintenance, while having essentially no effect on the (tactical-level) car-following measures. This is consistent with the fact that the secondary task performance requirements differed considerably between the two interfaces with respect to the degree of physical manipulation, but not appreciably with respect to the cognitive demands.

PDT performance was consistently better in the voice interface condition relative to the visual/manual condition, both in terms of percentage of targets detected and response time. This reflects inherent differences in visual requirements between the visual/manual interface, which requires reading and manipulating, and the voice-based interface, which requires listening and speaking. The eye glance data analyses revealed trends consistent with this conclusion, but did not have sufficient power to obtain statistical significance. Differences between interface conditions in PDT performance are consistent with the distinction between operational and strategic level behaviors and the associated implication with respect to the hypothesized locus of performance improvements due to the voice interface. Specifically, the PDT is largely visual and requires very little cognitive activity, since the signals do not have to be interpreted (there were no distractors) and there is no decision making about how and when to respond. Thus, the improvement in PDT performance associated with the voice interface is likely to have been primarily visual and not cognitive. The absence of differences in subjective workload ratings between interface conditions further supports the conclusion that the voice interface did not appreciably reduce the cognitive burden associated with the secondary task. Moreover, the differences between interface conditions observed on several measures when there was no secondary task were suggestive of a continuing cognitive load associated with the voice interface, however if present, this may have been due to problems related to the reliability of the voice interface. This interpretation is speculative and requires additional support to establish its credibility.

We found consistent differences between the subject groups. Group 1 subjects drove at longer following distances, detected fewer PDT targets and generally had more difficulty performing the required tasks together. Based on questionnaire responses, we found that Group 1 drivers had less education, less experience with personal computers and Internet activities, less experience in multi-tasking while driving, and were less willing to perform secondary tasks of the sort used in this study than Group 2 drivers. They were also less likely to drive in heavy freeway traffic, which requires following at shorter headways. These results suggest that individual differences can be expected to influence the extent to which engaging in complex secondary tasks will impact driving behavior. Specifically, the results suggest the influence of two related factors, both represented by the TRC test drivers. First, there is a segment of the population, whose members are not interested in using complex systems while driving. We refer to this as a lifestyle factor. Second, it appears that individual differences in attention-switching ability influenced the level of distraction potential observed in this study. This observation recalls earlier attempts to develop tests to predict selective attention (Avolio, Alexander, Barrett, & Sterns, 1981). While these efforts produced some success with respect to prediction of crash involvement for individual drivers (Ranney, 1994), they were limited by the difficulty of tying all crash involvement to problems of attention switching. The inherently more direct relationship between attention switching and distraction due to engaging in secondary tasks involving in-vehicle technologies suggests that such an approach may be considerably more successful in predicting distraction potential among individual drivers.

One potential weakness of the car-following paradigm, as noted by the Brookhuis et al. (Brookhuis et al., 1994), is that the delay (phase shift) during secondary tasks was affected by the time-headway between the vehicles. They conducted at least three experiments using this paradigm and initially addressed this problem by instructing all subjects to maintain the same following distance. They found that subjects had considerable difficulty maintaining a pre-specified following distance and removed this restriction in their subsequent experiments, allowing subjects to select their own following distances. They found that at shorter following distances, drivers responded more quickly to speed changes in the lead vehicle, resulting in shorter delay values. In the present experiment, we had considerable difficulty, particularly among Group 1 drivers, encouraging subjects to drive at close following distances. This introduced unwanted variability into the distribution of delay, which we addressed by conducting separate analyses using Group 2 only. As also reported by Brookhuis and colleagues (Brookhuis et al., 1994), this problem did not introduce enough variation to mask the impairment effects observed in the present study. Furthermore, rather than attempting to force a following distance that some drivers would find uncomfortable, we chose instead to use following distance as a measure of behavioral adaptation (Smiley, 2000), which is a natural response to increasing task demands.

We had intended to decompose the secondary tasks into more basic components, such as searching for a message or phone number, reading/listening to a message, dialing a phone, etc. However, this proved to be too much effort for the in-vehicle experimenters, particularly given the considerable difficulty experienced with the voice interface condition, which necessitated repeating steps several times. We hope that in future work we will be able to decompose the complex tasks so that we can assess the distraction potential of individual task components. However, such segmentation will require the development of comparable measures of driving or visual performance that can be compared across task components. The coherence measures are ideal for summarizing the impact of the entire (2 minute) secondary task on driving performance but could not be used to compare segments. Eye movement data, specifically the percentage of time spent looking at different locations, would be ideal for this comparison.

A number of procedural recommendations concerning the use of the El Mar eye tracking system can be made for future research. The percentage of successful trials, and consequently the amount of analyzable data, could be increased substantially if the experimenter controlled the time-of-day of the experiment, the direction of the drive and is more experienced with the calibration and set-up routines of the eye-tracker. A short pilot study, prior to the main study exploring these issues, would likely increase the percentage of usable data.

7.0 CONCLUSIONS

1. Performing secondary tasks like those used in this experiment while driving on a closed course resulted in significant decrements to vehicle control, visual target detection, and the decision making involved in car following. The pattern of results indicates that the combination of tasks used in this study generally created a relatively high level of workload for the drivers.

2. The voice-based interface helped reduce the distracting effects of secondary task performance, but the improvements were relatively minor and limited to measures of vehicle control and visual target detection. The absence of improvements in car-following performance indicates that the voice-based interface did not significantly reduce the cognitive distraction associated with the secondary tasks.

3. The car-following paradigm and the PDT were sensitive to performance decrements associated with secondary task performance. Additional vehicle control measures related to steering and lane maintenance were more sensitive to the differences between the visual/manual and voice-based interfaces.

4. Although the voice interface used in this study may not be representative of future products, the results suggest that voice interfaces may not provide enough help to overcome the increasing distraction associated with secondary tasks of increasing complexity, particularly in driving situations that require time-space judgments and tactical decision making, such as car following.

5. Differences observed between subject groups in the propensity and willingness to perform secondary tasks suggest that the problem of distraction due to interaction with complex in-vehicle technologies may be more restricted to certain segments of the driving population than is currently evident in the use of mobile phones. Within the segment of willing participants, distraction is likely to vary among individuals, based most likely on differences in the ability to switch efficiently between primary and secondary tasks.

6. The results of the present study provide information concerning the relative potential for interference with driving, when the secondary tasks are undertaken in situations that have primary task demands comparable to those used in the experiment. The real-world effects of secondary task engagement on driving behavior and safety also require consideration of factors that influence drivers' willingness to engage in the secondary tasks, the resulting incidence of secondary task engagement, and the associated crash risk.

8.0 REFERENCES

Avolio, B. J., Alexander, R. A., Barrett, G. V., & Sterns, H. L. (1981). Designing a measure of visual selective attention to assess individual differences in information-processing. Applied Psychological Measurement, 5, 29-42.

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

Bendat, J. S. & Piersol, A. G. (1986). Random Data: Analysis and Measurement Procedures. New York: Wiley.

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

Eizenman, M., Jares, T., & Smiley, A. (1999). A new methodology for the analysis of eye-movements and visual scanning in drivers (Contract No. T8080-8-13248/002/22). Transport Canada.

Goodman, M. J., Tijerina, L., Bents, F. D., & Wierwille, W. W. (1999). Using cellular telephones in vehicles: safe or unsafe? Transportation Human Factors, 1, 3-42.

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: Traffic Psychology and Behavior, 6F, 23-36

Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P.A. Hancock & N. Meshkati (Eds.), Human Mental Workload (pp. 139-183). North-Holland: Elsevier Science Publishers B.V.

IIHS (2002). Status Report (Vol. 37 No. 7). Arlington, Virginia: Insurance Institute for Highway Safety.

ISO (2002). Standard 15007-1 - Road Vehicles - Measurement of driver visual behavior with respect to transport information and control systems - Part 1: Definitions and parameters. International Organization for Standardization.

Laberge-Nadeau, C., Maag, U., Bellavance, F., Desjardins, D., Messier, S., & Saida, A. (2001). Wireless Telephones and the Risk of Road Accidents (Rep. No. CRT-2001-16). Center for Research on Transportation, Universite de Montreal.

Lee, J. D., McGehee, D. V., & Brown, T. L. (2000). Prior exposure, warning algorithm parameters and driver response to imminent rear-end collisions on a high-fidelity simulator. Proceedings of the IEA 2000/HFES 2000 Congress, 3, 316-320.

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.

Moroney, W. F., Biers, D. W., Eggemeier, F. T., & Mitchell, J. A. (1992). A comparison of two scoring procedures with the NASA task load index in a simulated flight task. NAECON 92 Proceedings of the IEEE National Aerospace and Electronics Conference, Dayton OH, 2 (A93-42776 17-01), 734-740.

Muto, W. H. & Wierwille, W. W. (1982). The effect of repeated emergency response trials on performance during extended-duration simulated driving. Human Factors, 24, 693-698.

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

Ranney, T. A. (1994). Models of driving behavior: a review of their evolution. Accident Analysis & Prevention, 26, 733-750.

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

Redelmeier, D. A. & Tibshirani, R. J. (1997). Association between cellular telephone calls and motor vehicle collisions. The New England Journal of Medicine, 336, 453-458.

Smiley, A. (2000). Behavior adaptation, safety, and intelligent transportation systems. Transportation Research Record, 1724, 47-51.

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

Tijerina, L. (2000). Issues in the Evaluation of Driver Distraction Associated with In-Vehicle Information and Telecommunications Systems (Paper No. 00-1477), presented at the annual meeting of the Transportation Research Board, Washington, D.C.

Tijerina, L., Kiger, S., Rockwell, T., & Tornow, C. (1995). Workload Assessment of In-Cab Text Message system and Cellular Phone Use by Heavy Vehicle Drivers on the Road. (Rep. No. DOT HS 808 467 (7A)). Washington, D.C.: NHTSA,U.S. Department of Transportation.

Utter, D. (2001). Passenger Vehicle Driver Cell Phone Use: Results from the Fall 2000 Occupant Protection Use Survey (Rep. No. Research Note DOT HS 809 293). NHTSA, U.S. Department of Transportation.

9.0 APPENDICES

9.1. Appendix A. Test Participant Consent Form

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

Study Description: Cell phones have become widely used in recent years. In response to concerns that they may interfere with driving, manufacturers are developing hands-free phones with voice-recognition capabilities. In addition, newer in-vehicle technologies now offer a wider range of opportunities to drivers, including retrieval of email messages or Internet access. The experiment in which you will participate is one of several experiments being conducted by the National Highway Traffic Safety Administration (NHTSA) to determine how well drivers are able to perform in-vehicle tasks while driving.

Your participation in this study will consist of two sessions, each lasting approximately 4-5 hours. During each session, you will be given training and practice on different in-vehicle tasks. You will also be given instructions and training on the requirements of the driving task. You will then complete a specified number of trials while driving on the TRC test track. During part of the study you will be asked to wear a head-mounted eye tracker, which will allow us to record exactly where you are looking while driving and performing in-vehicle tasks.

An experimenter will accompany you at all times during the experiment. The experimenter will give specific instructions about the driving and when to begin performing the 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.

It is very important to 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. The ride-along experimenter will not be able to ensure safety; you as the driver are responsible for safety.

Risks: While driving for this study, you will be subject to all risks normally associated with driving on the TRC test track, plus any additional risks associated with completing in-vehicle tasks while driving. There are no known physical or psychological risks associated with participation in the study beyond these.

Benefits: The experiment will provide data on driver behavior and in-vehicle task performance that will used by researchers to provide a scientific basis for developing recommendations or standards for performing in-vehicle tasks while driving. There are no direct benefits from participation in the study.

Confidentiality: The data recorded will be analyzed along with data gathered from other test participants in this experiment. Your name will not be associated with any final report, publication, or other media that might arise from this study. However, your likeness (in video-tape or still photo formats created from the video-tape), in-vehicle audio, and engineering data from you specifically may be used for educational and research purposes. A waiver of confidentiality for permission to use the videotape and engineering data (including data or images derived from these sources) is included for you to sign as part of this form. It is not anticipated that you will be informed of the results of this study, however, you may contact the Principal Investigator (see below) at a later date to obtain a copy of the final report.

Informed Consent: By signing below, you agree that your participation in this study is voluntary and you understand and accept all terms of this agreement. You have the option of not performing any requested task at any time during the experiment without penalty. You may also discontinue your participation at any time by notifying the experimenter.

Compensation: You will not receive any additional compensation beyond the pay you normally receive for work.

Principal Investigator: Contact Dr. Thomas Ranney (TRC) or Dr. Riley Garrott (NHTSA VRTC) if you have questions or comments regarding this study. They may be reached at the address and phone number given below:

Vehicle Research and Test Center
10820 SR 347
East Liberty, OH 43319
Phone: (937) 666-4511

Disposition of Informed Consent: The VRTC will retain a signed copy of this Informed Consent form. A copy of this form will also be provided to you upon completion of your participation in the study.

INFORMED CONSENT:

I, __________________________, UNDERSTAND THE TERMS OF THIS AGREEMENT AND VOLUNTARILY CONSENT TO PARTICIPATE.

____________________________ __________
Signature Date

____________________________ __________
Witness Date

WAIVER OF CONFIDENTIALITY:

I, __________________________, grant permission, in perpetuity, to the National Highway Traffic Safety Administration (NHTSA) to use, publish, or otherwise disseminate the video-tape (including all in-vehicle audio and still photo formats derived from the videotape) and engineering data collected about me in this study for educational, outreach, and research purposes. I understand that such use may involve widespread distribution to the public and may involve dissemination of my likeness in videotape or still photo formats, but will not result in release of my name or other identifying personal information.

9.2. Appendix B. Subject Instructions

Pre-Trial One Instructions

We are now ready to begin testing. This will be the first of several trials. Each trial will consist of a number of laps on the test track.

On each lap you will be doing two or three tasks simultaneously. First, you should maintain a constant following distance behind the lead vehicle. This may require speeding up or slowing down. Second, you should perform the light-detection task continuously by activating the button on your left hand as quickly as possible when you see a red light. Third, when instructed by the experimenter, you should perform the AutoPC or other in-vehicle task. The experimenter will tell you when to start and stop these tasks. You may not complete all of these tasks. If you get stuck and do not know what to do, you may ask the experimenter. However, you should try to do as much as you can without asking for help.

As always, safe vehicle operation is your first priority. When the experimenter instructs you to begin an in-vehicle task, this will become your second priority. The light-detection task is your third priority. When there is no in-vehicle task, the light detection task becomes your second priority.

Do you have any questions?

END of Trial Instructions/ Ratings

AFTER TRIAL 1: That is the end of the first trial. Please complete the workload rating scales using the scale definitions (give to subject to read). Your ratings should refer to the combination of all tasks <u>for the trial that we just completed</u>.

- Experimenter complete performance rating below
- Experimenter provide performance feedback to subject

AFTER TRIAL 3: That is the end of the third trial. Please complete the workload rating scales using the scale definitions (give to subject to read). Your ratings should refer to the combination of all tasks <u>for the trials since your last workload rating</u>.

- Experimenter complete performance rating below
- Experimenter provide performance feedback to subject

AFTER TRIAL 5: That is the end of the last trial. Please complete the workload rating scales using the scale definitions (give to subject to read). Your ratings should refer to the combination of all tasks <u>for all trials that were completed today</u>.

- Experimenter complete performance rating below
- Experimenter provide performance feedback to subject
-

Task #	Experimenter Instructions	Email Subject	Email Text
A0	Baseline		
A1	Your next task will be described in the message entitled Circuit City. When I give you the signal to begin, please select and open this message. Then follow directions contained in the message. Remember to create a voice memo after obtaining the required information. Please maintain your chosen following distance and perform the light-detection task at all times. Once again the message you are to retrieve is Circuit City.	Circuit City	Store hours today
A2	Your next task will be described in the message entitled 1 Movies. When I give you the signal to begin, please select and open this message. Then follow directions contained in the message. Remember to create a voice memo after obtaining the required information. Please maintain your chosen following distance and perform the light-detection task at all times. Once again the message you are to retrieve is 1 Movies.	1 Movies	List any 2 movies
A3	Your next task will be described in the message entitled Delta. When I give you the signal to begin, please select and open this message. Then follow directions contained in the message. Remember to create a voice memo after obtaining the required information. Please maintain your chosen following distance and perform the light-detection task at all times. Once again the message you are to retrieve is Delta.	Delta	Departure time flight 700
A4	Your next task will be described in the message entitled UPS. When I give you the signal to begin, please select and open this message. Then follow directions contained in the message. Remember to create a voice memo after obtaining the required information. Please maintain your chosen following distance and perform the light-detection task at all times. Once again the message you are to retrieve is UPS.	UPS	Drop off hours
A5	Your next task will be to tune the radio to several different stations. When I give you the signal to begin, I will tell you the first station id number and you should tune the radio to this frequency. When you are finished, please say "done" and I will give you another frequency. I will continue in the same way for several additional radio frequencies. Please maintain your chosen following distance and perform the light-detection task at all times.		
A6	Your next task will be to dial phone numbers that are not familiar to you. When I give you the signal to begin, I will show you a card containing a phone number. When you see the card, you may begin dialing. When you are finished, please say "done" and I will show you another phone number to dial. I will continue in the same way for several additional phone numbers. Please maintain your chosen following distance and perform the light-detection task at all times.		
A7	Your next task will be described in the message entitled Shopping List. When I give you the signal to begin, please select and open this message. Then follow directions contained in the message. Remember to create a voice memo after obtaining the required information. Please maintain your chosen following distance and perform the light-detection task at all times. Once again the message you are to retrieve is Shopping List.	Shopping List	Hot dogs, buns, beer

Task #	Experimenter Instructions	Email Subject	Email Text
B0	Baseline		
B1	Your next task will be described in the message entitled Best Buy. When I give you the signal to begin, please select and open this message. Then follow directions contained in the message. Remember to create a voice memo after obtaining the required information. Please maintain your chosen following distance and perform the light-detection task at all times. Once again the message you are to retrieve is Best Buy.	Best Buy	Store hours Sunday
B2	Your next task will be described in the message entitled 2 Movies. When I give you the signal to begin, please select and open this message. Then follow directions contained in the message. Remember to create a voice memo after obtaining the required information. Please maintain your chosen following distance and perform the light-detection task at all times. Once again the message you are to retrieve is 2 Movies.	2 Movies	Matinee price
B3	Your next task will be described in the message entitled US Air. When I give you the signal to begin, please select and open this message. Then follow directions contained in the message. Remember to create a voice memo after obtaining the required information. Please maintain your chosen following distance and perform the light-detection task at all times. Once again the message you are to retrieve is US Air.	US Air	Arrival time flight 775
B4	Your next task will be described in the message entitled Fed Ex. When I give you the signal to begin, please select and open this message. Then follow directions contained in the message. Remember to create a voice memo after obtaining the required information. Please maintain your chosen following distance and perform the light-detection task at all times. Once again the message you are to retrieve is Fed Ex.	Fed Ex	Marysville last pick-up time
B5	Your next task will be to tune the radio to several different stations. When I give you the signal to begin, I will tell you the first station id number and you should tune the radio to this frequency. When you are finished, please say "done" and I will give you another frequency. I will continue in the same way for several additional radio frequencies. Please maintain your chosen following distance and perform the light-detection task at all times.		
B6	Your next task will be to dial phone numbers that are familiar to you. When I give you the signal to begin, I will show you a card containing a phone number. When you see the card, you may begin dialing. When you are finished, please say "done" and I will show you another phone number to dial. I will continue in the same way for several additional phone numbers. Please maintain your chosen following distance and perform the light-detection task at all times.		
B7	Your next task will be described in the message entitled Remember. When I give you the signal to begin, please select and open this message. Then follow directions contained in the message. Remember to create a voice memo after obtaining the required information. Please maintain your chosen following distance and perform the light-detection task at all times. Once again the message you are to retrieve is Remember.	Remember	Birthday cake, candles, gift

Task #	Experimenter Instructions	Email Subject	Email Text
C0	Baseline		
C1	Your next task will be described in the message entitled Best Buy. When I give you the signal to begin, please select and open this message. Then follow directions contained in the message. Remember to create a voice memo after obtaining the required information. Please maintain your chosen following distance and perform the light-detection task at all times. Once again the message you are to retrieve is Best Buy.	Best Buy	Store hours Satruday
C2	Your next task will be described in the message entitled 3 Movies. When I give you the signal to begin, please select and open this message. Then follow directions contained in the message. Remember to create a voice memo after obtaining the required information. Please maintain your chosen following distance and perform the light-detection task at all times. Once again the message you are to retrieve is 3 Movies.	3 Movies	First show time any movie
C3	Your next task will be described in the message entitled US Air. When I give you the signal to begin, please select and open this message. Then follow directions contained in the message. Remember to create a voice memo after obtaining the required information. Please maintain your chosen following distance and perform the light-detection task at all times. Once again the message you are to retrieve is US Air.	US Air	Arrival gate 1063
C4	Your next task will be described in the message entitled Fed Ex. When I give you the signal to begin, please select and open this message. Then follow directions contained in the message. Remember to create a voice memo after obtaining the required information. Please maintain your chosen following distance and perform the light-detection task at all times. Once again the message you are to retrieve is Fed Ex.	Fed Ex	Nearest drop off location
C5	Your next task will be to tune the radio to several different stations. When I give you the signal to begin, I will tell you the first station id number and you should tune the radio to this frequency. When you are finished, please say "done" and I will give you another frequency. I will continue in the same way for several additional radio frequencies. Please maintain your chosen following distance and perform the light-detection task at all times.		
C6	Your next task will be to dial phone numbers that are familiar to you. When I give you the signal to begin, I will show you a card containing a phone number. When you see the card, you may begin dialing. When you are finished, please say "done" and I will show you another phone number to dial. I will continue in the same way for several additional phone numbers. Please maintain your chosen following distance and perform the light-detection task at all times.		
C7	Your next task will be described in the message entitled Remember. When I give you the signal to begin, please select and open this message. Then follow directions contained in the message. Remember to create a voice memo after obtaining the required information. Please maintain your chosen following distance and perform the light-detection task at all times. Once again the message you are to retrieve is Remember.	Remember	Flowers for your mother

Task #	Experimenter Instructions	Email Subject	Email Text
D0	Baseline		
D1	Your next task will be described in the message entitled Circuit City. When I give you the signal to begin, please select and open this message. Then follow directions contained in the message. Remember to create a voice memo after obtaining the required information. Please maintain your chosen following distance and perform the light-detection task at all times. Once again the message you are to retrieve is Circuit City.	Circuit City	Store hours Sunday
D2	Your next task will be described in the message entitled 4 Movies. When I give you the signal to begin, please select and open this message. Then follow directions contained in the message. Remember to create a voice memo after obtaining the required information. Please maintain your chosen following distance and perform the light-detection task at all times. Once again the message you are to retrieve is 4 Movies.	4 Movies	First show time after 6pm
D3	Your next task will be described in the message entitled Delta. When I give you the signal to begin, please select and open this message. Then follow directions contained in the message. Remember to create a voice memo after obtaining the required information. Please maintain your chosen following distance and perform the light-detection task at all times. Once again the message you are to retrieve is Delta.	Delta	Arrival time flight 2542
D4	Your next task will be described in the message entitled UPS. When I give you the signal to begin, please select and open this message. Then follow directions contained in the message. Remember to create a voice memo after obtaining the required information. Please maintain your chosen following distance and perform the light-detection task at all times. Once again the message you are to retrieve is UPS.	UPS	Nearest staffed drop off location
D5	Your next task will be to tune the radio to several different stations. When I give you the signal to begin, I will tell you the first station id number and you should tune the radio to this frequency. When you are finished, please say "done" and I will give you another frequency. I will continue in the same way for several additional radio frequencies. Please maintain your chosen following distance and perform the light-detection task at all times.		
D6	Your next task will be to dial phone numbers that are not familiar to you. When I give you the signal to begin, I will show you a card containing a phone number. When you see the card, you may begin dialing. When you are finished, please say "done" and I will show you another phone number to dial. I will continue in the same way for several additional phone numbers. Please maintain your chosen following distance and perform the light-detection task at all times.		
D7	Your next task will be described in the message entitled Shopping List. When I give you the signal to begin, please select and open this message. Then follow directions contained in the message. Remember to create a voice memo after obtaining the required information. Please maintain your chosen following distance and perform the light-detection task at all times. Once again the message you are to retrieve is Shopping List.	Shopping List	Eggs, cheese, milk

9.3. Appendix C. NASA-TLX Materials

Subject Instructions: Ratings

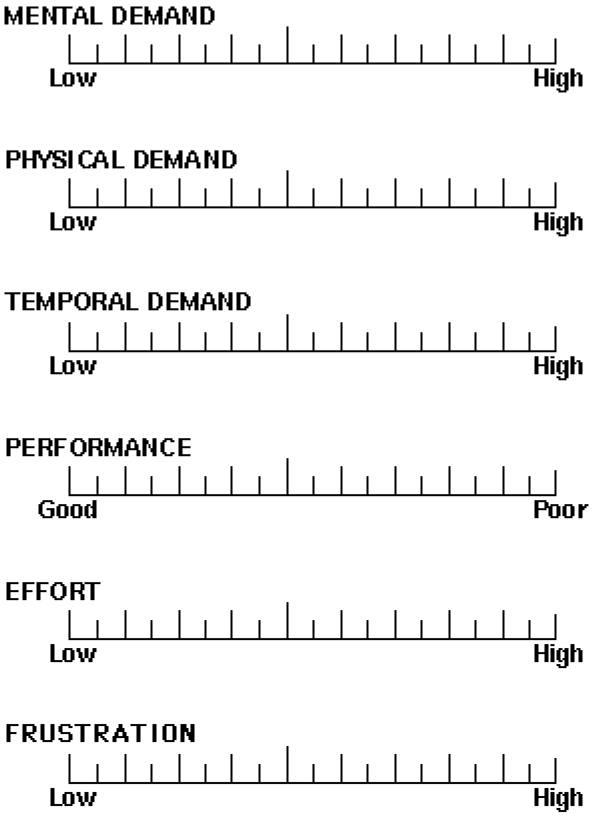
We are interested not only in assessing your performance but also the experiences you had during the different task conditions. Right now we are going to describe the technique that will be used to examine your experiences. In the most general sense we are examining the "workload" you experienced. Workload is a difficult concept to define precisely, but a simple one to understand generally. The factors that influence your experience of workload 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 workload 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. Physical components of workload are relatively easy to conceptualize and evaluate. However, the mental components of workload may be more difficult to measure.

Since workload is something experienced individually by each person, there are no effective "rulers" that can be used to estimate the workload of different activities. One way to find out about workload is to ask people to describe the feelings they experienced. Because workload may be caused by many different factors, we would like you to evaluate several of them individually rather than lumping them into a single global evaluation of overall workload. This set of six rating scales was developed for you to use in evaluating your experiences during different tasks. Please read the descriptions of the scales carefully. If you have a question about any of the scales in the table, please ask me about it. It is extremely important that they be clear to you. You may keep the descriptions with you for reference during the experiment.

After performing each task, six rating scales will be displayed. You will evaluate the task by marking each scale at the point which matches your experience. Each line has two endpoint descriptors that describe the scale. Note that "own performance" goes from "good" on the left to "bad" on the right. This order has been confusing for some people. Use the pencil to mark the scale at the desired location. Please consider your responses carefully in distinguishing among the task conditions. Consider each scale individually. 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.

RATING SCALE DEFINITIONS

Title	Endpoints	Descriptions
MENTAL DEMAND	Low/High	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?
PHYSICAL DEMAND	Low/High	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	Low/High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
EFFORT	Low/High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
PERFORMANCE	Good/Poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
FRUSTRATION LEVEL	Low/High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?


MENTAL DEMAND
Low
High
PHYSICAL DEMAND
Low
High
TEMPORAL DEMAND
Low
High
PERFORMANCE
Good
Poor
EFFORT
Low
High
FRUSTRATION
Low
High

9.4. Appendix D. Reward/Penalty Computation

Performance Ratings	After Trial 1	After Trial 3	After Trial 5	Row Total	Multiplier	Total Money
Car Following Performance					.75	
In-Vehicle Task Performance					.50	
Light-detection Performance					.25	
Total						

2 Good - few errors (errors on less than 20% of trials)
1 Acceptable - moderate number of errors (errors on less than 40% of trials)
0 Poor – many errors (errors on more than 40% of trials)

Performance Feedback:

On the most recent trial:
Your car-following performance was _______________
Your light-detection task performance was ________________
Your in-vehicle task performance was _____________

Your monetary rewards will be based on these performance ratings.

Reward computation matrix

	Multiplier		
Score	.25	.50	.75
1	.25	.50	.75
2	.50	1.00	1.50
3	.75	1.50	2.25
4	1.00	2.00	3.00
5	1.25	2.50	3.75
6	1.50	3.00	4.50

9.5. Appendix E. Pilot Study

Using the Peripheral Detection Task (PDT) raised the question of whether the demands associated with its performance would impact driving or secondary task performance. We were also interested in whether the El Mar head-mounted eye tracker would impact PDT performance or driving performance. To address these questions we conducted a pilot study in which eight licensed drivers completed four laps of the TRC 7.5-mile test track. The car-following paradigm utilized for the main experiment was used. Subjects were instructed to select the shortest following distance at which they felt comfortable following the lead vehicle. Car-following and PDT data were obtained on the eight 2.0-mile straight sections according to the following design

Table 6. Data Collection for Pilot Study

	Eye Tracker	No Eye Tracker
PDT	2	2
No PDT	2	2

We computed analyses of variance (ANOVAs) for car-following parameters, including coherence and phase shift. Model factors were PDT (yes, no), and Eye tracker (yes, no). Results are shown in the following table.

Table 7. ANOVA Results for Coherence Measures

Dependent Measure	Source	d.f.	F	P > F
Coherence	PDT	1,7	3.56	.10
	Eye tracker	1,7	2.56	.15
Phase Shift (delay)	PDT	1,7	8.84	.02
	Eye Tracker	1,7	1.32	.29

Neither the PDT nor the eye tracker affected coherence, the basic measure of car-following performance. However, the requirement to perform the PDT while driving did increase the phase shift (delay), which reflects a delay in car-following performance. Thus, while the drivers were able to track lead vehicle changes equally well with or without the requirement to perform the PDT task, their car-following performance was slower while performing the PDT. Means and standard deviations are shown in the following table.

Table 8. Mean and Standard Deviations of Car-following Delay by PDT Condition

PDT	N	Mean	SD
No	31	1.99	0.97
Yes	30	2.36	1.11

On average, the addition of the PDT increased car-following delay by 0.37 seconds. This indicates that adding the PDT to the car-following task significantly increased driver workload. This increase is intended to simulate the increased driving task demand in situations that require identification of potential hazards (e.g., pedestrians) in addition to vehicle control (car-following).

We also computed analyses of variance (ANOVAs) for two measures of PDT performance, percentage of targets detected and response time for targets detected. Eye tracker (yes, no) was the independent variable. Results are shown in the following table.

Table 9. ANOVA Results for PDT Measures

Dependent Measure	Source	d.f.	F	P > F
Target detection response time	Eye tracker	1,6	0.0	1.00
Percent of targets detected	Eye Tracker	1,6	9.77	.02

Wearing the eye tracker had no effect on the response time for PDT target detection. However, wearing the eye tracker, which had a shield over the eyes, significantly influenced the percentage of targets detected. Mean percentage of targets detected without the eye tracker was 94.3 (SD = 10.1) versus 89.7 (SD = 11.9) when the eye tracker was worn.

9.6. Appendix F. Analyses of Visual Behavior When Using AutoPC to Perform Complex Tasks While Driving

Drivers' visual behavior was recorded using an eye tracker while they drove and performed tasks using the AutoPC. The most complete subset of data was available for 10 drivers during trials involving complex task performance. There were four instances of complex tasks for each interface; consequently the analyses were based on one to four instances of complex tasks for each participant, depending on the availability of valid data.

The purpose of the analyses was to document any differences in visual behavior that may have arisen as a result of using the two types of interface (voice-based or manual) while performing tasks using the AutoPC. The analyses of drivers' visual behavior focus on three main types of questions. To address these issues the following types of analyses of visual behavior were performed as a function of interface type and driver group:

1. How long do drivers look off-road (at AutoPC) or at the roadway (forward view) while performing tasks? The analyses of Percentage of Task Time spent looking at regions of interest address these issues.
2. How often do drivers look at the device (or out to the roadway) while carrying out the tasks? To address this issue, analyses were done on the number of glances made per 10 seconds to the specified areas (e.g., AutoPC and forward view) while performing tasks while driving.
3. How long are the glances that drivers make to the device (or out to the roadway) while performing the tasks? The Analyses of Glance Durations provide this information.

In addition, a series of analyses were performed examining drivers' visual behavior while carrying out the peripheral detection task.

9.6.1. Overview of Analysis Design

There were three areas of interest for the visual behavior analyses: AutoPC, Forward View, and PDT area.

Depending on the questions being addressed by the analyses, two different approaches were taken to the data analysis. When the focus was on a specific comparison of the two interfaces, the data were analyzed using a 2 (Driver group: Group 1 vs. Group 2) X 2 (Interface type: Voice vs. Visual/Manual) Mixed ANOVA with repeated measures on the latter factor. For other analyses, where it was appropriate to include the control condition (No Secondary Task), the data were analyzed using a 2 (Driver group: Group 1 vs. Group 2) X 3 (Interface condition: None, Voice or Visual/Manual) Mixed ANOVA with repeated measures on the latter factor. Post hoc comparisons were carried out using LSD tests.

Analyses addressing drivers' visual behavior during the peripheral detection task (PDT) are included in a separate section at the end of this appendix. The PDT was included in the study as a measure of event detection.

9.6.2. Percent Time Looking

"Percent time looking" is defined as the total time a participant spent looking in a defined area (e.g., AutoPC) divided by the total task time. Percent time looking was computed for each of the following locations: (1) the AutoPC, (2) Forward roadway scene, and (3) PDT targets.

The mean percent time spent looking at the AutoPC is presented in Table 10 for each combination of Subject Group and Interface Condition. Although Group 1 spent a higher percentage of time (27.17%) looking at the AutoPC than Group 2 (18.71%), this difference was not reliable (*F*(1,8) = 2.04, p = .19). Neither the interaction between interface type and driver group (*F*(1,8) = 0.23, p = .65) nor the main effect for interface type was significant (*F*(1,8) = 0.22, p = .65).

Table 10. Mean Percent Time Looking at AutoPC as a Function of Interface Type and Driver Group.

	Interface Type		
Groups	Manual	Voice	Marginal Means
Group1	29.11	25.22	27.17
Group2	18.69	18.73	18.71
Weighted Marginal Means	22.86	21.33	

Mean percent times looking in forward view, for each condition, are presented in Table 11. The main effect for condition was significant (*F*(2,16) = 32.01, p < .001) indicating that drivers spent a greater percentage of time looking in the forward view when they were not using the AutoPC (97.6%) than when they performed a task using either the voice-based interface (68.8%) or the visual/manual interface (64.8%). Neither the main effect of driver group (*F*(1,8)= 1.95, p = .20) nor the interaction between interface type and driver group was significant (F(2,16) = 0.81, p = .46).

Table 11. Mean Percent Time Spent Looking in Forward View as a Function of Interface Type and Driver Group.

	Condition			
Group	No-Task	Manual	Voice	Marginal Means
Group1	97.63	59.99	61.91	73.17
Group2	97.64	68.07	73.37	79.70
Weighted Marginal Means	97.63	64.84	68.79	

The specific impact of Driver Group and Interface Type on the percentage of time spent in forward view was further explored in a 2 X 2 ANOVA in which the No-Task condition was eliminated. Although Group 2 drivers spent 10% more time looking in the forward view, the main effect of driver group was not significant (*F*(1,8) = 1.90, p = .21). Neither the main effect for Interface (*F*(1,8) = 0.56, p = .48) for the Interface/Driver Group interaction was (*F*(1,8) = 0.12, p = .74).

The strongest finding for this measure was that drivers spent more time looking in the forward view when not performing a secondary task. Otherwise, the results were not very informative as to the relative benefit/cost of either type of interface. Although some of the means were in the expected direction (i.e., Group 1 drivers spent more time looking at the AutoPC, and the voice-based interface afforded the drivers more time in the forward view), none of the comparisons was significant.

9.6.3. Number of Glances per 10 Seconds

A glance is defined as the time from the moment at which the gaze direction moves toward a target area (e.g., AutoPC) until it begins to move out of that area. Glance durations thus include the prior transition time (transition time to that target), consistent with ISO 15007 (2002). Each glance is separated by at least one glance to a different area. Because the different secondary tasks took varying amounts of time to complete, the analyses are based on the mean number of glances to a specified location made during a 10 second interval.

The mean number of glances per 10 seconds to the AutoPC for each combination of Interface and Subject Group are presented in Table 12. There were no appreciable differences in the number of glances made per 10 seconds to the AutoPC as a function of Driver Group or Interface Type (all *F*s < 0.6; all ps > .49).

Table 12. Mean Percent Number of Glances per 10 Seconds to AutoPC as a Function of Interface Type and Driver Group.

	Interface Type		
Group	Manual	Voice	Marginal Means
Group1	2.45	2.74	2.59
Group2	2.29	2.15	2.22
Weighted Marginal Means	2.35	2.39	

The mean numbers of glances made per 10-second interval to the forward view are presented in Table 13. The main effect of condition type was significant (*F*(2,16) = 44.03, p < .001). Participants made significantly fewer glances per 10 seconds to the forward view in the no-task condition (0.34 glances) than in the voice (3.74 glances) and manual (3.56 glances) interface conditions. The interaction between interface type and driver group was non-significant (*F*(2,16) = 2.14, p = .15) as was the main effect of driver group (*F*(1,8) = 0.004, p = .95).

Table 13. Mean Number of Glances per 10 Seconds to Forward View as a Function of Interface Type and Driver Group

	Condition			
Group	No-task	Manual	Voice	Marginal Means
Group1	0.46	2.99	4.15	2.53
Group2	0.25	3.95	3.48	2.56
Weighted Marginal Means	0.34	3.56	3.74	

To further explore whether interface type had any impact on the number of glances that drivers made to the forward view, the means were analyzed in a 2 X 2 (Driver X Interface) ANOVA.

The interaction between Interface and Driver Group was non-significant (*F*(1,8) = 2.51, p = .15). Neither the main effect of driver group (*F*(1,8) = 0.08, p = .78), nor the main effect of interface type (*F*(1,8) = 0.44, p = .52) was significant.

Although the Group 1 drivers made numerically more glances per 10 s to the AutoPC, this difference was not statistically reliable. The significant main effect indicating that overall drivers made fewer glances to the forward view when not performing an in-vehicle task is due primarily to their not having to alternate their attention between the forward view and the AutoPC.

9.6.4. Mean Glance Durations

Mean glance durations (in seconds) to the device (AutoPC) for each condition are presented in Table 14. The interaction between Interface Type and Driver Group was significant (*F*(1,8) = 8.14, p < .03). The two groups of drivers made glances of similar duration to the AutoPC when using the voice-based interface (.83-.84s); however when using the visual/manual interface, Group 1 drivers made marginally longer glances (M = 1.01 s) than Group 2 (M = .72 s) drivers (p = .07). Neither the main effect of driver group (*F*(1,8) = 2.29, p = .17), nor the main effect of interface type (*F*(1,8) = 0.40, p = .55) was significant.

Table 14. Mean Glance Durations (in Seconds) to Device as a Function of Interface Type and Driver Group.

	Interface Type		
Group	Manual	Voice	Marginal Means
Group1	1.01	0.84	0.92
Group2	0.72	0.83	0.77
Weighted Marginal Means	0.84	0.83	

The mean glance durations to forward view for each combination of Driver Group and Interface type are presented in Table 15. Drivers' forward glances were significantly longer in the No-Task condition than in either Secondary Task condition, *F*(2, 16) = 5.45, p < .02. Both the interaction between condition and driver group (*F*(2,16) = 0.95, p = .40) and the main effect of driver group were non-significant (*F*(1,8) = 1.01, p = .34).

Table 15. Mean Glance Durations (in Seconds) to Forward View as a Function of Interface Type and Driver Group.

	Condition			
Group	No-Task	Manual	Voice	Marginal Means
Group1	23.60	1.96	1.43	9.00
Group2	55.65	1.84	2.42	19.97
Weighted Marginal Means	42.83	1.89	2.02	

A separate comparison of forward-view glance durations between the two interface conditions did not reveal any reliable differences.

9.6.5. Categories of Glance Durations

Researchers have raised concerns that glances to in-vehicle devices exceeding 2 seconds are unsafe (e.g., Zwahlen et al., 1988). Wikman et al. (1998) classified glances into three categories: (1) unusually short (< .5 sec.), (2) medium (0.5-2 sec.), and (3) over-long (> 2 sec.), based on typical glance length distributions found in earlier research (i.e., Rockwell, 1998; Zwahlen et al. 1988; Wierwille et al. 1991). The distributions of glance durations to the AutoPC were further examined using these categories.

The mean percentages of AutoPC glances shorter than .5 second are presented in Table 12. Glances of this duration may be too short to acquire much visual information about the in-vehicle task and would thus represent a "checking glance". There was a significant interaction of Interface by Driver Group *F*(1,8) = 5.47, p < .05. Further analyses indicated that Group 2 drivers made a larger percentage of short glances when using a visual/manual interface opposed to the voice interface than did Group 1 drivers (p < .05). Neither the main effect of driver group (*F*(1,8) = 0.23, p=.64), nor the main effect of interface type (*F*(1,8) = 0.08, p = .79) were significant.

Table 16. Mean Percentage of Glances Shorter than .5 Seconds to AutoPC as a Function of Interface Type and Driver Group.

	Interface Type		
Group	Manual	Voice	Marginal Means
Group1	17.83	26.76	22.29
Group2	31.58	20.24	25.91
Weighted Marginal Means	26.08	22.85	

The mean percentages of mid-range (.5 s - 2 s) AutoPC glances are presented in Table 17. No differences were observed for driver groups or interfaces (Interface X Driver Group interaction, *F*(1,8) = 1.73, p = .22; Driver Group main effect, *F*(1,8) = 0.23, p = .65; Interface main effect, *F*(1,8) = 2.56, p = .15).

Table 17. Mean Percentage of Glances Between .5 and 2 Seconds to AutoPC, as a Function of Interface Type and Driver Group

	Interface Type		
Group	Manual	Voice	Weighted Marginal Means
Group1	68.37	69.66	69.02
Group2	65.41	78.60	72.01
Weighted Marginal Means	66.60	75.03	

The percentages of AutoPC glances longer than 2 seconds to the are presented in Table 18. The main effect of Interface was significant, (*F*(1,8) = 5.61, p < .05) indicating that a greater percentage of these long glances were made using the visual/manual interface (7.32%) relative to the voice interface (2.12%). Although the interaction was not significant, Group 1 drivers made a greater percentage (over 13%) of these long glances when using the visual/manual interface compared with the voice interface or either condition for the Group 2 drivers (*F*(1,8) = 2.70, p = .14). There was no effect of driver group (*F*(1,8) = 2.25, p = .17).

Table 18. Mean Percentage of Glances Longer than 2 Seconds to AutoPC, as a Function of Interface Type and Driver Group.

	Interface Type		
Group	Manual	Voice	Marginal Means
Group1	13.78	3.57	8.68
Group2	3.00	1.16	2.08
Weighted Marginal Means	7.32	2.12	

9.6.6. Cumulative Distribution of Drivers Making Glances of Varying Durations

For safety reasons, extreme glance durations are of particular concern. Consequently, we examined the cumulative distribution of the longest glances made by the drivers to the AutoPC.

Figure 18 shows the cumulative distribution of the percentage of drivers who had AutoPC glances of various durations when using the visual/manual interface. Glances longer than 2.5 s occurred in 50% of Group 1 drivers but in none of the Group 2 drivers. Glances over 3 s were made by 25% of Group 1 drivers but not by any Group 2 drivers. These results suggest that, when using the visual/manual interface, more Group 1 drivers made more dangerously long glances than Group 2 drivers.

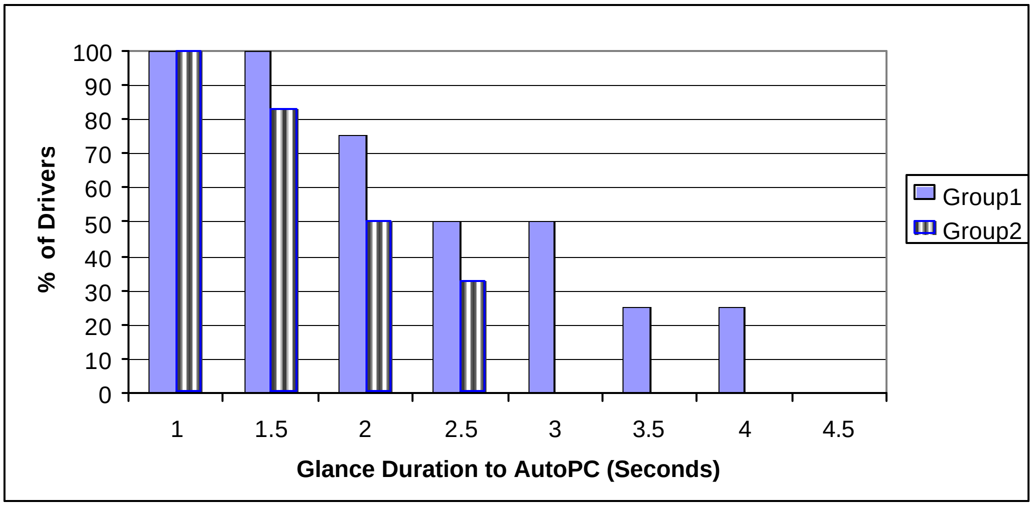


Figure 18. Visual/manual interface: The cumulative distribution of maximum glance durations to the AutoPC for Group 1 and Group 2 Drivers.

Figure 19 shows the cumulative distribution of the percentage of AutoPC glances by duration, when using the voice-based interface. In this condition, glances longer than 2.5 s occurred in 25% of Group 1 drivers compared to 17% in Group 2 drivers. Glances over 3 s occurred in 17%

of Group 2 drivers but did not occur in any of the Group 1 drivers. These results suggest that more Group 2 drivers made dangerously long glances than Group 1 drivers when using a voice-based interface.

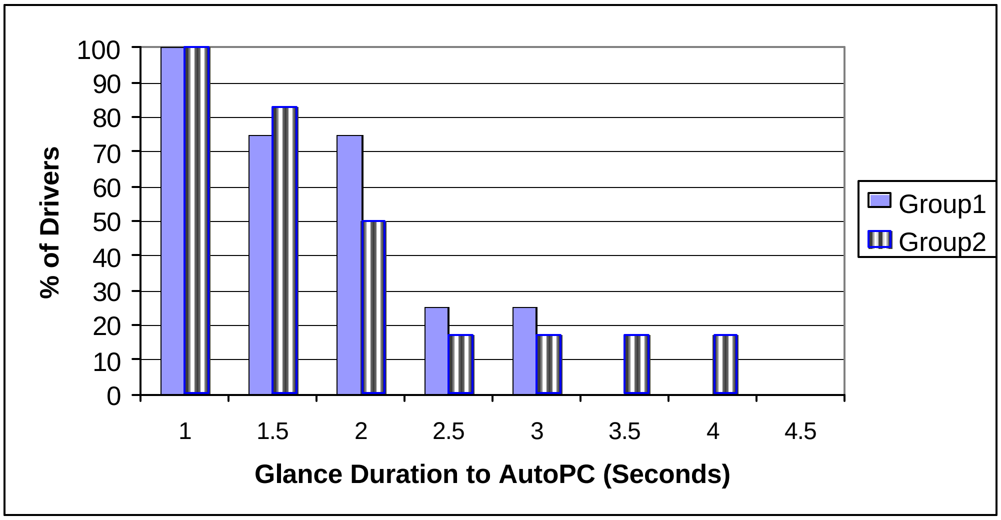


Figure 19. Voice interface: The cumulative distribution of maximum glance durations to the AutoPC for Group 1 and Group 2 drivers.

9.6.7. Distributions of Glance Durations to AutoPC

Figure 20 shows the distribution of AutoPC glance durations in the visual/manual interface condition for both driver groups. Group 2 drivers made proportionately more short glances (i.e.,up to .5 s) than did Group 1 drivers (31.6% vs. 17.8%). Conversely, Group 2 drivers made proportionately fewer long glances (exceeding 2 s) than did Group 1 drivers (3.0% vs. 13.8%). The two groups were comparable in the percentage of midrange (.51-2 s) glances.

Figure 21 shows the distribution of AutoPC glance durations in the voice interface condition for both driver groups. The mean percentage of short glances (up to .5 s) made by Group 1 drivers was 26.76% versus 20.24% for Group 2 drivers. The percentage of midrange glances (between .51 and 2 s) made by Group 2 drivers (M = 78.60%) was greater than that for Group 1 drivers (M = 69.66%). Long glances, exceeding 2 s, were made by both groups of drivers (Group 1 = 3.57%; Group 2 = 1.16%). However, they constituted a lower percentage of the overall glances in the voice interface condition.

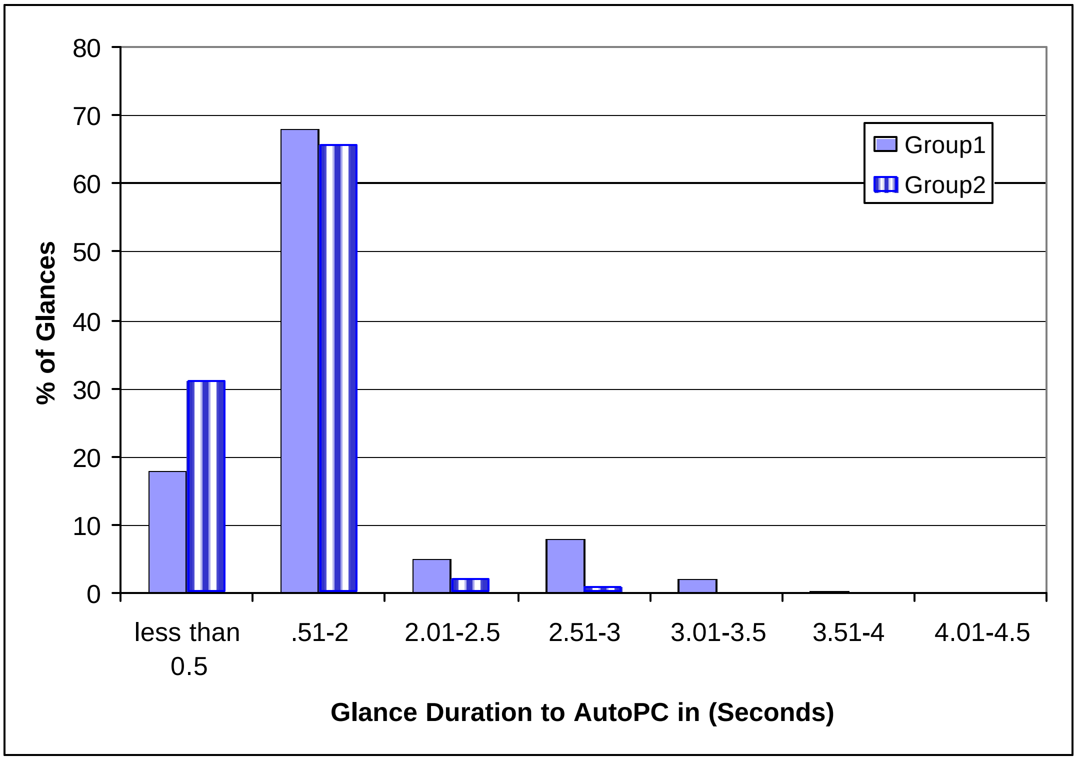


Figure 20. Distribution of glance durations to the AutoPC: Visual/manual interface

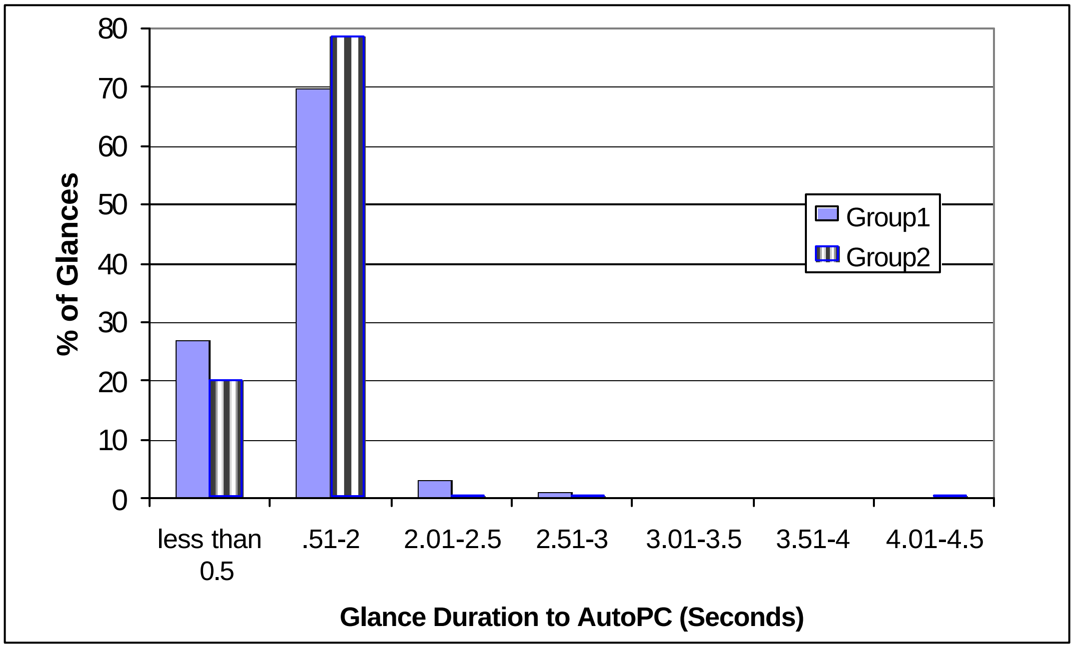


Figure 21. Distribution of glance lengths to the AutoPC: Voice interface

9.6.8. Summary of Glance Duration Analyses

The longest mean glance durations to the AutoPC were found for Group 1 drivers using the visual/manual interface. Overall, these analyses did not reveal any benefit in terms of longer glance durations (and as a result more time) to the forward view for the voice-based compared the visual/manual interface.

Some of the most important safety relevant data comes from the analyses of the distribution of glance durations. A greater percentage of these long glances (> 2 s) were made using the visual/manual interface (7.32%) than when using the voice interface (2.12%). For Group 1 drivers, over 13%, of their glances fell into the long glance category.

The examination of the distribution of maximum glance durations to the AutoPC revealed that long glance durations were a problem for both interfaces and both groups of drivers. When using the visual/manual interface, more Group 1 drivers made dangerously long glances than Group 2 drivers. However, when the voice-based interface was used, long glances were also observed, although not to such a great degree.

9.6.9. Analyses of Visual Behavior and the Peripheral Detection Task (PDT)

While the participants drove and performed the tasks using the AutoPC, they were also required to perform the Peripheral Detection Task (PDT). This task involved responding to a visual target (illuminated LED reflected in the windshield) by pressing a finger switch (a full description of the task is provided in the Section 3 of the Main Report).

The PDT has been used in the past as a measure of driver workload. It can also be considered an event detection task. To the extent that a particular interface allowed drivers to spend more time looking out, PDT performance would be expected to improve. The mean percentage of task time spent looking at PDT is presented in Table 19. There was a significant main effect of driver group (*F*(1,8) = 6.20, p < .04). Group 1 drivers spent a higher percentage of their task time looking at the PDT than Group 2 drivers. Neither the interaction between Interface type and Driver Group (*F*(2,16) = 1.37, p = .28) nor the main effect of condition was significant (*F*(2,16) = 1.45, p = .26).

Table 19. Mean Percentage Time Spent Looking at PDT as a Function of Interface Type and Driver Group

	Condition			
Group	No-task	Manual	Voice	Weighted Marginal Means
Group1	0.72	1.79	3.40	1.97
Group2	1.18	0.20	1.10	0.83
Weighted Marginal Means	0.99	0.84	2.02	

The specific impact of interface type on time spent looking at the PDT was further examined. Again there was a significant main effect of driver group (*F*(1,8) = 7.29, p < .03) reflecting that Group 1 drivers spent a higher percentage of time looking at the PDT than Group 2 drivers. Drivers also spent a higher percentage of time looking at the PDT while using the voice-based interface (2.02%) compared with the visual/manual interface (0.84%; *F*(1,8) = 5.50, p < .05). The interaction was not significant (p = .52).

Table 20 presents the mean number of PDT glances per 10 seconds. Group 1 participants made more glances to the PDT per 10 seconds (0.43) than Group 2 participants (0.16; *F*(1,8) = 5.80, p

$< .05$). It is possible that Group 1 drivers may have been more distracted by the PDT task or felt that they had to monitor it more closely than the Group 2 subjects. Another explanation for this result might be that Group 2 drivers were better able to detect the peripheral stimuli without looking directly at the PDT. This latter explanation is consistent with the PDT performance data for this group of 10 participants performing the complex tasks. Group 2 subjects detected a larger proportion of targets (M = .88) than Group 1 subjects (M = .69). Group 2 drivers were also significantly faster at detecting PDT targets than Group 1 drivers (M = .63 s vs. M = .81 s). The PDT detection and latency results for this subgroup of subjects reflect the findings for the full set of subjects.

There was a main effect of interface condition ($F(2,16) = 5.50$, $p < .02$) indicating that more glances were made to the PDT when using the voice-based interface (.48 glances/10 sec) than when using the visual/manual interface (0.22 glances/10 sec; $p = .05$) or no in-vehicle task at all (0.12 glances/10 sec; $p < .02$). These results are consistent with the finding that drivers detected marginally more ($p = .06$) targets while performing secondary tasks using the voice interface (M = .80) than they did while using the visual/manual interface (M = .69). However, drivers were most likely to detect targets (M = .92) when not performing any secondary task. The interaction between interface type and driver group was not significant ($F(2,16) = 2.89$, $p = .09$).

Table 20. Mean Number of Glances per 10 Seconds to the PDT Area as a Function of Interface Type and Driver Group.

	Condition			
Group	No-task	Manual	Voice	Marginal Means
Group1	0.06	0.44	0.78	0.43
Group2	0.15	0.07	0.27	0.16
Weighted Marginal Means	0.12	0.22	0.48	

The mean glance durations to the PDT were analyzed to determine how long drivers looked at the PDT while driving (Table 21). There was a significant effect of condition, $F(2,16) = 13.10$, $p < .001$). Drivers made significantly longer glances to the PDT when performing tasks using the voice-based interface (M = 0.39 s) than when using the visual/manual interface (M = 0.26 s; $p = .05$). Furthermore, glances to the PDT in the no-task condition (M = 0.08 s) were significantly shorter than in the voice ($p < .01$) or visual/manual interface ($p < .01$) conditions. Neither the interaction between interface type and driver group ($F(2,16) = 0.42$, $p = .67$) nor the main effect of driver group was significant ($F(1,8) = 0.83$, $p = .39$).

Table 21. Mean Glance Durations to PDT as a Function of Interface Type and Driver Group.

	Condition			
Group	No-task	Manual	Voice	Marginal Means
Group1	0.11	0.35	0.41	0.29
Group2	0.05	0.21.	0.38	0.21
Weighted Marginal Means	0.08	0.26	0.39	

The two groups of drivers differed in the amount of time spent looking at the PDT and in the number of glances made (per 10-second interval) to the PDT. Group 1 drivers spent a greater percentage of task time looking at the PDT and made more glances (per 10-second interval) to the PDT. However, their greater amount of time spent looking at the PDT was not associated with better performance. Group 2 drivers detected proportionately more PDT targets (M = 0.88) than Group 1 drivers (M = 0.69) and they were significantly faster at doing so (M RT: Group 1 = .081 s, Group 2 = 0.63 s).

Compared with when they used the visual/manual interface, drivers using the voice-based interface, spent a greater percentage of their time looking at PDT, made more glances (per 10-second interval) to the PDT and made longer glances to the PDT. These results are consistent with the finding that drivers detected more targets while performing secondary tasks using the voice interface (M = 0.80) than they did while using the visual/manual interface (M = 0.69). However, drivers were most likely to detect targets (M = 0.92) when not performing a secondary task.

9.6.10. Discussion of Visual Behavior Results

Most of the useful information concerning drivers' visual behavior and the AutoPC comes from the analyses of glance durations. Use of the visual/manual interface was associated with longer mean glance durations to the AutoPC for the Group 1 drivers and also with a greater percentage of glances exceeding 2 s for all drivers. These findings suggest a visual/manual interface may be more distracting than a voice-based interface, at least under the conditions tested in this study.

However, the examination of the distribution of maximum glance durations raised concerns about both interface types. When using the visual/manual interface, more Group 1 drivers made dangerously long glances away from the road. However, when the voice-based interface was used long glances were observed in the Group 2 drivers as well. The activities associated with these long glance durations to the AutoPC are reading a message, visually scanning message headers, and correcting mistakes.

Zwahlen et al. (1988), and others, have raised concerns that off road glances exceeding 2 s are unsafe. Several groups investigating possible safety criteria measures (e.g., European Commission, 1998; AAM, 2002) have incorporated this idea into their work. Although these groups differ in their approach and have suggested various combinations of glance duration and total task time in their work, there is consistency in their concern about the safety of glances exceeding 2 seconds.

The data from the PDT analyses suggest that there may be some benefits associated with the voice-base interface. Compared with when they used the visual/manual interface, drivers using the voice-based interface spent a greater proportion of their time looking at the PDT, made more glances (per 10 s interval) to the PDT, and made longer glances to the PDT. This increased inspection of the PDT was reflected in their performance: greater and faster target detection with the voice-based interface. These results suggest that the voice-based interface may have safety benefits in that it leads to better event detection.

Overall the analyses of visual behavior suggest that there are safety concerns associated with the use of both interfaces, although the voice-based interface appears to have some benefits.

9.7. References

The Alliance of Automobile Manufacturers (2002). *Statement of Principles on Human Machine Interface (HMI) for In-Vehicle Information and Communication Systems.* http://www.autoalliance.org/DF-0402-protected.pdf

European Commission. (1998). Commission Recommendation of 21 December 1999 on safe and efficient in-vehicle information and communication systems: A European statement of principles on human machine interface. *Official Journal of the European Communities* (OJ L 19, 2000).

ISO (2002). Standard 15007-1 - Road Vehicles - Measurement of driver visual behavior with respect to transport information and control systems - Part 1: Definitions and parameters. International Organization for Standardization.

Rockwell, T. H. (1998). Spare visual capacity in driving-revisited. In A.G. Gale, H.M. Freeman, C.M. Haslegrave, P. Smith & S.P. Taylor (Eds.) Vision in Vehicles II (pp. 317-324). Amsterdam: Elsevier.

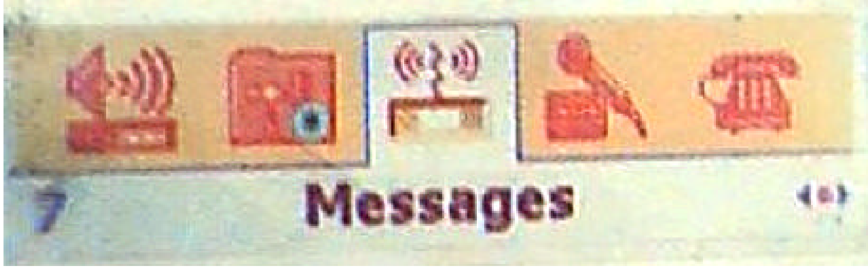
Wierwille, W.W., Hulse, M.C., Fischer, T.J., & Dingus, T.A. (1991). Visual adaptation of the driver to high demand driving situations while driving with an in-car navigation system. In A.G. Gale, C.M. Haslegrave, I. Moorhead, & S.P. Taylor. (eds.) Vision in Vehicles III (pp.79-89) Amsterdam: Elsevier.

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

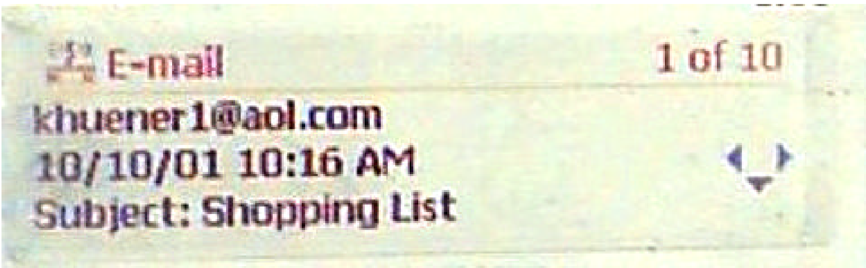
Zwahlen, H.T., Adams, C.C. Jr., & Debald, D.P. (1988). Safety aspects of CRT touch panel controls in automobiles. In A.G. Gale, H.M. Freeman, C.M. Haslegrave, P. Smith & S.P. Taylor (Eds.) Vision in Vehicles II (pp. 335-344). Amsterdam: Elsevier.

9.8. Appendix G. Anatomy of a Complex Secondary Task

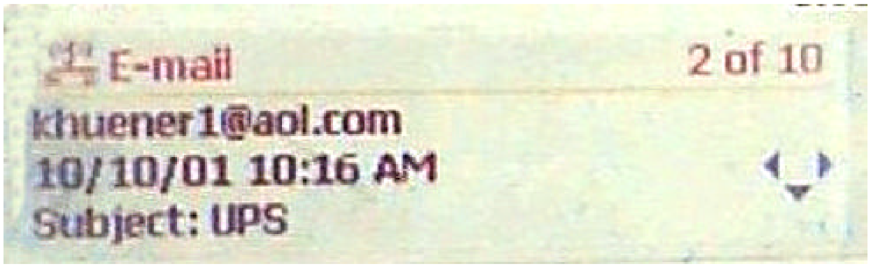
(Note: Image colors have been reversed in order to improve legibility in the printed copy.)



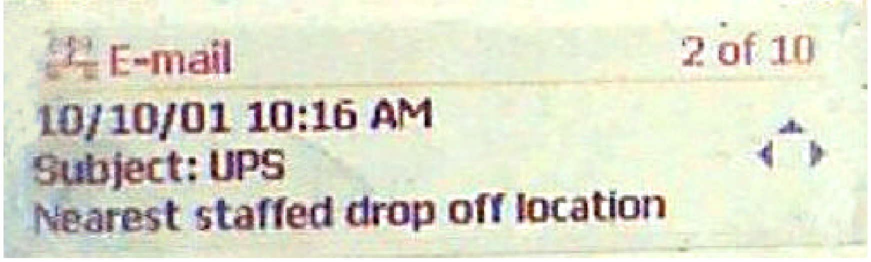
Locate “Messages” Program


E-mail 1 of 10
khuener1@aol.com
10/10/01 10:16 AM
Subject: Shopping List

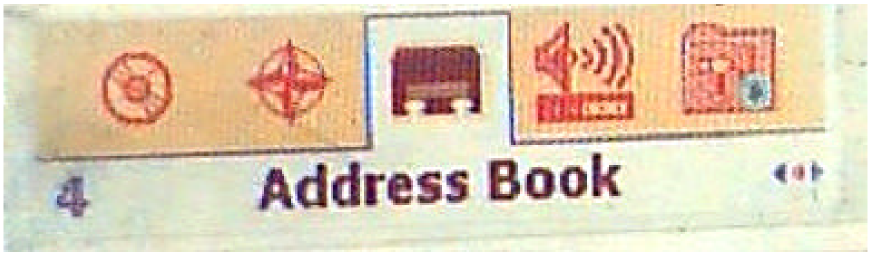
Activate “Messages” Program


E-mail 2 of 10
khuener1@aol.com
10/10/01 10:16 AM
Subject: UPS

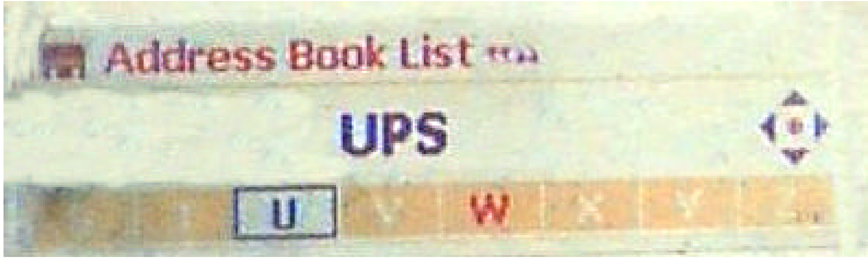
Scroll to specified message


E-mail 2 of 10
10/10/01 10:16 AM
Subject: UPS
Nearest staffed drop off location

Select, Read/Listen to message



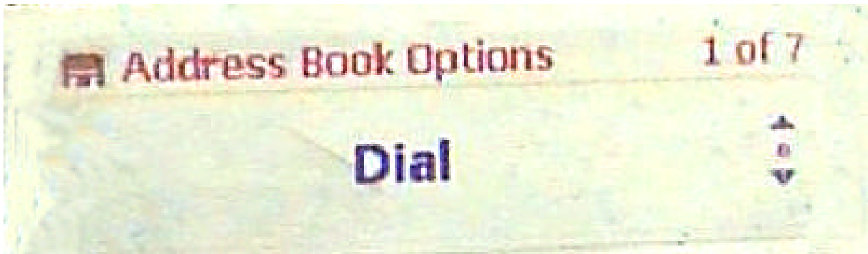
Activate “Address Book” program



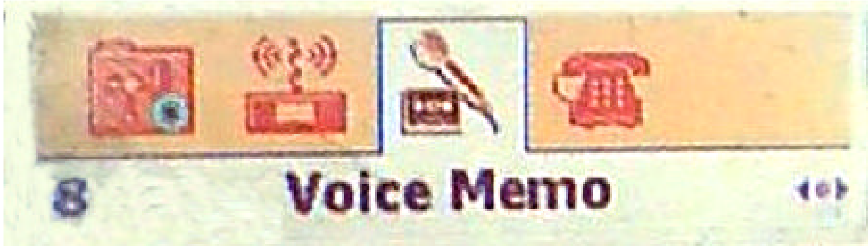
Scroll to specified contact

Work: 8007425877

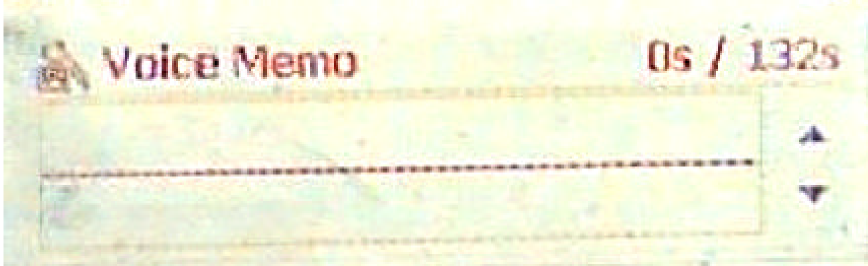
Select contact telephone number



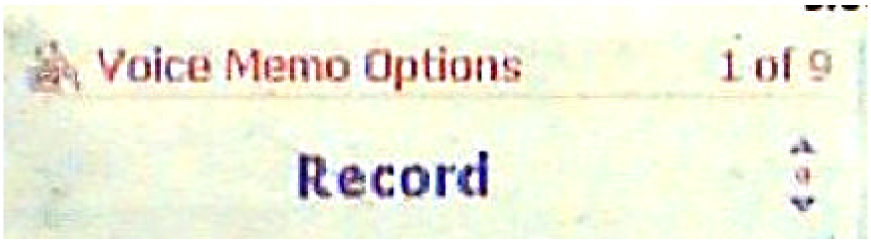
Issue “Dial” command



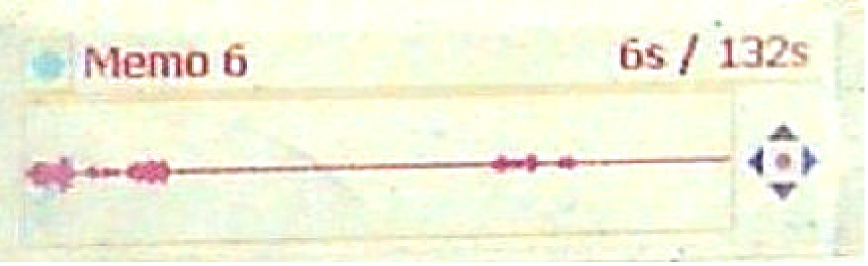
Locate “Voice Memo” program



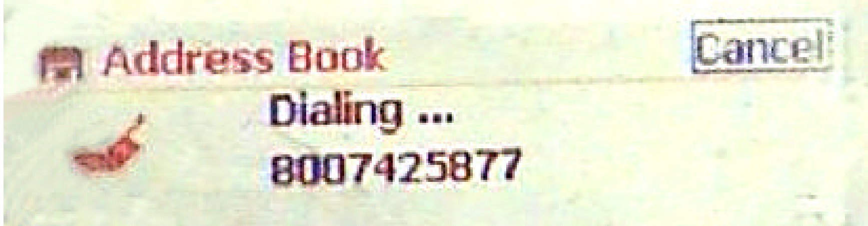
Open “Voice Memo” program



Select “Record” from menu



Record voice memo



Dialing