Development of Human Factors Guidelines for Advanced Traveler Information Systems (ATIS) and Commercial Vehicle Operations (CVO): Driver Memory for In-Vehicle Visual and Auditory Messages

PUBLICATION NO. FHWA-RD-96-148

DECEMBER 1999

U.S. Department of Transportation
Federal Highway Administration

Research, Development, and Technology
Turner-Fairbank Highway Research Center
6300 Georgetown Pike
McLean, VA 22101-2296

FOREWORD

This report is one of a series of reports produced as part of a contract designed to develop precise, detailed human factors design guidelines for Advanced Traveler Information Systems (ATIS) and Commercial Vehicle Operations (CVO). During the analytic phase of the project, research issues were identified and rated by 8 human factors experts along 14 separate criteria. The goal of the experimental phase was to examine the highest rated research issues that can be addressed within the scope of the project. The 14 experiments produced in that phase reflect the results of those ratings.

This study is part of the empirical phase of this ATIS/CVO guidelines development effort and is one of a series of investigations designed to provide supporting rationale for the in-vehicle design guidelines. The research reported in this document investigated the effects of message format and modality on comprehension and retention of in-vehicle messages for younger and older drivers.

Copies of this report can be obtained through the Research and Technology Report Center, 9701 Philadelphia Court, Unit Q, Lanham, Maryland 20706, telephone: (301) 577-0818, fax: (301) 577-1421, or the National Technical Information Service (NTIS), 5285 Port Royal Road, Springfield, Virginia 22161, telephone: (703) 605-6000, fax: (703) 605-6900.

Michael F. Trentacoste
Director, Office of Safety
Research and Development

NOTICE

Technical Report Documentation Page

1. Report No. FHWA-RD-96-148	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle: DEVELOPMENT OF HUMAN FACTORS GUIDELINES FOR ADVANCED TRAVELER INFORMATION SYSTEMS (ATIS) AND COMMERCIAL VEHICLE OPERATIONS (CVO): DRIVER MEMORY FOR IN-VEHICLE VISUAL AND AUDITORY MESSAGES.		5. Report Date: December 1999 6. Performing Organization Code
7. Author(s): Kantowitz, B. H., Hanowski, R. J., & Garness, S. A.		8. Performing Organization Report No.
9. Performing Organization Name and Address Battelle Seattle Research Center 4000 NE 41st Street, NE Seattle, WA 98105		10. Work Unit No. (TRAIS): 3B2C1012 3B2C 11. Contract or Grant No. DTFH61-92-C-00102
12. Sponsoring Agency Name and Address Office of Safety and Traffic Operations R&D Federal Highway Administration 6300 Georgetown Pike McLean, Virginia 22101-2296		13. Type of Report and Period Covered Technical Report 4/95 - 4/99 14. Sponsoring Agency Code
15. Supplementary Notes: Contracting Officer’s Technical Representative (COTR): M. Joseph Moyer, HRDS		

16. Abstract:
Three experiments were conducted in a driving simulator to evaluate effects of in-vehicle message modality and message format on comprehension and memory for younger and older drivers. Visual icons and text messages were effective in terms of high comprehension and high memory retention over a 50-sec delay period for both older and younger drivers. Auditory icons (earcons) were unsuitable for older drivers, although younger drivers performed well using symbolic and speech in-vehicle messages. For visual only, auditory only, and auditory plus visual messages both textual and symbolic message formats were equally effective. No adverse effects of in-vehicle message presentation were found for lateral and longitudinal vehicle control. Implications of these findings for development of ATIS guidelines were discussed.

17. Key Words: In-vehicle information system, IVIS Advanced Traveler Information System (ATIS) memory older driver		18. Distribution Statement: No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161	
19. Security Classif. (of this report) None	20. Security Classif. (of this page) None	21. No. of Pages 154	22. Price

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

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yards	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
NOTE: Volumes greater than 1000 l shall be shown in m^3.				
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE		
EF	Fahrenheit temperature	5(F-32)/9 or (F-32)/1.8	Celcius temperature	EC
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.71	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE		
EC	Celcius temperature	1.8C+32	Fahrenheit temperature	EF
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.

(Revised September 1993)

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF FIGURES (continued)

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF TABLES (continued)

Table **Page**

LIST OF TABLES (continued)

Table **Page**

LIST OF ABBREVIATIONS

AHS	Automated Highway System
ANOVA	Analysis of Variance
ATIS	Advanced Traveler Information Systems
BAS	Battelle Automobile Simulator
CVO	Commercial Vehicle Operations
ITS	Intelligent Transportation Systems
IVIS	In-Vehicle Information Systems
LCD	Liquid Crystal Display
MUTCD	Manual on Uniform Traffic Control Devices
STI	Systems Technology, Inc.
STISIM	Systems Technology, Inc. Simulator
SDL	Scenario Definition Language
VCR	Video Cassette Recorder

EXECUTIVE SUMMARY

This report describes three experiments that investigated the ability of younger and older drivers to comprehend and remember messages displayed on an in-vehicle information system (IVIS). The primary independent variables studied were message modality (visual, auditory, and visual plus auditory messages) and message format (symbolic and lexical). Three major research questions were asked:

- How do in-vehicle message format and modality affect comprehension?
- How do in-vehicle message format and modality affect memory retention?
- What is the impact of driver age on comprehension and memory?

Both message comprehension and memory are important to the designer of an advanced traveler information system (ATIS). It is obvious that messages that are not comprehended cannot guide the driver. However, even a comprehended message will not guide the driver if it cannot be remembered long enough for the driver to act on it. Hence, system designers need advice on what message formats and modalities will best suit the younger and older driver in terms of comprehension and memory.

All experiments were conducted in a low-medium fidelity driving simulator. Procedures were similar for all three experiments. First, drivers learned a list of visual symbols (icons) or auditory symbols (earcons). These symbols were presented on a screen or through loudspeakers inside the vehicle. Then messages were presented while the vehicle was in motion. Drivers were asked to recognize the messages either immediately after presentation (comprehension) or after a 50-s delay (memory). Response latencies and confidence were also scored for each in-vehicle message. Measures of vehicle control were recorded from the simulator.

The first experiment, called experiment 5A in the project work plan, examined only visual messages. Nine younger (ages 18-22) and nine older (ages 65-80) drivers were tested. Although the work plan called for this memory experiment to take place in a laboratory test booth, the early completion of new simulator facilities allowed us to move the experiment into the simulator. The simulator controlled vehicle speed with cruise control while the driver was responsible for steering. Two message formats were tested: text and icons. The visual icons were largely selected from existing symbols familiar to traffic engineers, with a few new symbols created for this experiment. Comprehension of symbols was high for older (90 percent) and younger (99 percent) drivers. The higher performance of the younger drivers was primarily associated with the less familiar symbols. Younger drivers suffered little memory loss, while older drivers suffered slight (about 5 percent) memory loss. Message presentation did not degrade vehicle lateral control. These results indicate that visual messages, either text or icons, are eminently suitable for in-vehicle display for both younger and older drivers.

The second experiment (experiment 5B) used only auditory messages and was otherwise quite similar to the first experiment. Twelve younger (ages 18-30) and six older (ages 65-80) drivers were tested. A dramatic difference was found in the utility of symbolic messages (earcons) according to driver age. While all the younger drivers were able to learn the set of earcons with almost perfect recall scores in the initial phase of the experiment, none of the older drivers were able to learn them even after considerable practice. This important finding strongly suggests that earcons are not suitable for in-vehicle displays used by older drivers. For younger drivers, comprehension and memory scores were similar for earcons and speech messages.

The third experiment (experiment 5C) used visual, auditory, and simultaneous visual plus auditory messages. Since auditory messages may have a built-in alerting advantage when drivers are looking at the road rather than at the IVIS display, we included three types of auditory alerts that preceded message presentation: tone alert, speech alert, and a no-alert control condition. Thirty-six younger drivers (ages 18-30) were tested. Unlike the first two experiments, drivers had full control of vehicle speed and lane position. Contrary to our expectations, comprehension and memory were not improved by using visual plus auditory messages; using only a single message modality worked just as well. Comprehension and memory were very high (above 96 percent) and none of the alert conditions offered an advantage or disadvantage. Driving performance immediately preceding message presentation was compared with performance while a message was being displayed. In-vehicle message presentation did not impair driving performance. Indeed, there was some evidence that processing in-vehicle messages tends to suppress lateral and longitudinal control actions by the driver, resulting in less variable positioning of the vehicle on the roadway when the display was active.

CHAPTER 1. EXPERIMENT 5A

INTRODUCTION

The importance of human factors research in the development of Intelligent Transportation System (ITS) technologies has been outlined by the ITS America, Safety and Human Factors Committee (ITS America, 1995). Specific research areas recommended include investigation into enroute information, route guidance, and traveler service information. Each of these research areas involves the design of transportation technologies and the driver's accessibility to in-vehicle information.

The need for human factors research in the development of new technologies has also been outlined (e.g., Norman, 1988). Often, when human factors issues are not addressed early in the design process, the final design is associated with less-than-optimal use. An example of such a design is the video cassette recorder (VCR) and the difficulties many experience in set-up and programming. (Have you ever been to someone's house and noticed their VCR flashing "12:00"?). Sometimes, inattention to the effects of human factors in product design may be benign. Difficulty in programming a VCR may, at worst, result in a missed taping of a television program. However, with transportation-related applications, the result of poorly designed systems may lead to unsafe driving. Therefore, the careful application of human factors principles and guidelines is required in ITS design (Dingus & Hulse, 1993).

Advanced Traveler Information Systems (ATIS) is an ITS technology that provides drivers with enroute information, including driver advisories and in-vehicle signing. Driver advisories include real-time information that assists drivers in making navigation decisions. As the name implies, in-vehicle signing provides drivers with road sign information directly inside the vehicle (e.g., on a dash mounted display). In-vehicle information organization and display is one specific ATIS-related research issue outlined by ITS America (1995). One design issue concerns optimally presenting information inside the vehicle in a manner where it is quickly understood and not distracting to the driver. In-vehicle information that is difficult to comprehend and/or diverts the driver's attention from the primary task of operating the vehicle may result in unsafe driving.

The research presented here is the first in a series of experiments to investigate in-vehicle information organization and display. This series of experiments will be used in the development of guidelines for ATIS and Commercial Vehicle Operation (CVO) technologies. Specifically, this research examines performance differences associated with focusing all ATIS information through either single or multiple display channels (Broadbent, 1971), and investigates how to display multiple ATIS messages so that drivers can identify relevant information and react appropriately. The present experiment was limited to visual messages. Follow-up experiments will investigate auditory messages and visual/auditory message combinations.

The focus of the experiment presented here was to investigate drivers' memory for traffic and traveler-related messages that were displayed on an in-vehicle ATIS, or more specifically, an In-Vehicle Information System (IVIS). Rapid access and ease in processing in-vehicle information are important considerations for system effectiveness and, perhaps more importantly, driver safety. Messages that are difficult to understand may lead to extended glance duration and attention diversion from the primary task of operating the vehicle. Therefore, it is important to consider different formats that in-vehicle messages might take, and investigate their possible effects on driver performance and behavior.

In the present research, three questions were examined: (1) How does in-vehicle visual message format affect comprehension? (2) How does message format affect memory retention? and (3) What impact does driver age have on memory for in-vehicle visual messages? Regarding the first question, two formats that are typical of static traffic signs are symbols and text. Published research is limited with respect to comprehension comparisons between symbol and text traffic sign messages. Hawkins, Womack, and Mounce (1993) used a multiple-choice questionnaire and examined comprehension for 10 symbols and 7 text traffic warning signs. The results indicated similar comprehension for the two message types. Correct comprehension for the symbol signs varied from approximately 32 percent to 87 percent, with a mean of 61 percent. Text message sign comprehension ranged from 29 percent to 89 percent, with a mean of 61 percent. Based on these findings, we might expect similar results for the in-vehicle environment; comprehension for text and symbol messages should be similar.

Regarding the second research question, the effect of message format on memory retention, past research has examined retention for various recall intervals (Brown, 1959; Peterson & Peterson, 1959). Results suggest that the accuracy of recall decreases as the recall interval increases. From this, we might expect driver's memory for in-vehicle messages to be better with shorter recall intervals. This issue may be particularly important when outlining guidelines for message timing and presentation. Providing information to the driver that allows sufficient time to react and make required adjustments (e.g., change lanes) is a desired goal that must be weighed against the driver's ability to retain that information in memory. For example, consider the potential dilemma in the presentation timing of directional information (e.g., "turn left on Cavers Street"). Presenting this information too late (e.g., at the intersection) risks the driver missing the turn. Presenting the information too early (e.g., five blocks before the intersection) risks the driver forgetting what to do. The optimal timing of information presentation, and the consequences of early and late presentations, needs careful consideration.

For the third research question of interest, the impact of age on memory, Dewar, Kline, and Swanson (1994) investigated symbolic traffic sign comprehension as a function of age for 85 of the symbols in the U.S. Manual on Uniform Traffic Control Devices (Federal Highway Administration, 1988). The results indicated that older drivers had a poorer understanding than did younger drivers on 39 percent of the symbols examined. There were no age differences with the remaining 61 percent of the symbols. From this result, we might anticipate a similar finding for in-vehicle messages; younger drivers may have superior understanding of some messages, though there may be no age difference for other messages.

To summarize, the purpose of this experiment was to investigate drivers' understanding and recall for traffic- and traveler-related messages that were displayed on an IVIS. Rapid access and ease in processing in-vehicle information is an important consideration for system effectiveness and, perhaps more importantly, driver safety. Messages, either icons or text, that are difficult to understand may lead to extended glance duration and attention diversion from the primary task of operating the vehicle. Therefore, it is important to consider different formats that in-vehicle messages might take, and investigate their possible effects on driver performance and behavior.

METHOD

SUBJECTS

Twenty-six subjects agreed to participate in this experiment. Data from eight of them were not used due to either experimenter error, simulator failure, or participant withdrawal. There were three older drivers who, after the practice sessions, were either not able to perform the task or were not comfortable with the experimental procedure and withdrew from the study. In addition, the data from four older drivers and one younger driver weres not used due to simulator and/or experimenter error. The 18 drivers who did complete the study were grouped on the basis of age; drivers under 30 years made up the "younger drivers" group (N = 9), while those over 64 comprised the "older drivers" group (N = 9). Five of the younger drivers and four of the older drivers were female. The age range for younger drivers was 18 to 22 years and the range for older drivers was 65 to 80 years. All subjects had a valid driver's license, drove at least twice per week, and reported not being prone to motion sickness. Younger drivers were recruited from the University of Washington, while older drivers were recruited from local church, volunteer, and retirement groups. Each driver was paid $5.00 per hour, for approximately 3 hours of research time.

To determine subjects' driving experience, data were collected on the following: age (younger, M = 19.8; older, M = 71.3), years as a licensed driver (younger, M = 3.89; older, M = 49.1), and estimated total miles driven annually (younger, M = 8599; older, M = 7500).

APPARATUS

Driver behavior was investigated using the Battelle Automobile Simulator (BAS) shown in figure 1. The major components of simulator include: (1) the automobile test buck, (2) message and question displays, and (3) the simulation software.

Figure 1. The Battelle Automobile Simulator.

Automobile Test Buck

The automobile test buck was built by Walter Dorwin Teague Associates, Incorporated. The buck was constructed using a 1986 Merkur XR4Ti automobile. The original side and top body work, from 12 inches in front of the firewall to 20 inches behind the driver's seat, have been maintained to conserve the feel of a real automobile. The floor pan has been replaced by a wood base to provide easy movement of the buck. The dash of the automobile has been modified to allow multiple configurations, including combinations of active matrix liquid crystal display (LCD) touch-screens and electroluminescent displays, and a completely analog instrument panel. A small fan was also included in the instrument panel to provide air circulation to the driver. The steering column is that of the Merkur with no modifications. The steering wheel has been modified to include up to 12 push-button switches (6 on each side of the wheel) placed at approximately 130 and 240 degrees. These switches connect to the digital input port of a Computer Dynamics CIO-DIO24 card. The steering shaft is also connected to a torque motor, which can be adjusted to produce accurate roadway feedback to the driver. Interior lights are located in the center of the headliner near the front windshield and can be aimed by the driver as needed. The rear of the vehicle is open to allow access to the rear speakers. Both doors are operational with side-view mirrors. The buck also has adjustable driver and passenger seats.

The front "windshield" is completely enclosed. The left side of the windshield houses a 20-inch multisync color monitor providing a simulated roadway display for the various driving scenarios. The monitor is covered with a black wooden hood, and the right side of the windshield is covered with a black piece of plastic to reduce the ambient background lighting. The front windshield enclosure can be removed completely for use with the three screen option, allowing free vision in all directions.

Message and Question Displays

The message display is a Planar Systems, Incorporated, EL640.350-DA Series Multicolor EL Display. The viewing area is 121.9 mm x 179 mm. The center of the display is situated 330.2 mm to the right and 88.9 mm above the center of the steering wheel.

Questions are displayed on a TekVisions, Incorporated, 238.8-mm diagonal Rainbow Visions Active Matrix Color LCD display. This display is offset 228.6 mm below and 57.2 mm to the right of the EL display and is centered on the transmission channel of the vehicle. The touch-screen uses resistive technology with a serial controller. Both monitors are driven by a Colorgraphic Communications Super Warp Accelerator. This graphic card is a dual VGA Video Adapter based on Tseng Labs ET4000/W32 video accelerator chip. This card allows resolutions of up to 1280 x 1024 with 16.8 million colors (640 x 480). The displays are driven by a 486-based computer, which is interlinked with the STI computer using a second CIO-DIO24 digital input/output card.

Simulation Software

A closed-loop, low-fidelity driving simulator that was developed by Systems Technology, Inc. (STI) (V. 8.01) was used, consisting of software and commercially available IBM PC compatible hardware components for producing visual scenes and auditory displays relevant to driving. Driver relevant vehicle dynamics are specified by STI. Using a simple scenario definition language (SDL), the researcher can specify scenario attributes that relate to driver psychomotor, divided attention, and cognitive behavior.

The STI simulator (STISIM) is fully interactive and includes the following features: 5-speed automatic transmission, variable vehicle dynamics, simulated road noise (engine and drive train), tire squeal to signal loss of control on high-speed turns, and wire-framed rendering of displayed objects. Major components of the system include the following: a 586-based computer that controls the simulation, a 20-inch multisynchronous monitor for the simulation display, a 14-inch EGA monitor for the experimenter, a sound blaster card, and steering, brake, and accelerator potentiometers and cables. Major capabilities of the simulator include variable length and radius curves, both expected and unexpected obstacles, a random access sound file, and a secondary visual detection task integrated into the system. Though the STI software allows for full driver interaction, for the purposes of this experiment, the driver was only able to steer the vehicle and had no control over the accelerator or brakes. In this respect, this study can be thought of as simulating cruise control or an Automated Highway System (AHS).

EXPERIMENTAL DESIGN

Overview of Independent Variables

For the simulator portion of this experiment, a repeated measures design with three within-subjects variables (message type, question delay, message repetition) and two between-subjects variables (age and gender) was used. As subjects drove through the simulation, traveler-related messages were presented on an IVIS. There were six types of in-vehicle messages: (1) very low comprehension symbol, (2) medium low comprehension symbol, (3) medium high comprehension symbol, (4) very high comprehension symbol, (5) short text, and (6) long text. Subjects were queried as to the meaning of a message after it was presented. Questions pertaining to a message were presented either immediately after the message was presented or after a 50-s delay. That is, there were two levels of the question delay variable: (1) 0-s delay and (2) 50-s delay. Over the course of the simulation, all messages were repeated twice. As such, there were two levels of the repetition variable. As outlined in the *Subjects* section, drivers were grouped into one of two age groups: (1) younger drivers and (2) older drivers. In terms of gender, both males and females were included in the subject sample. A summary of the independent variables is presented in table 1.

Table 1. Summary of the independent variables.

INDEPENDENT VARIABLE	LEVELS
Message Type	(1) Very Low Comprehension Symbol
	(2) Medium Low Comprehension Symbol
	(3) Medium High Comprehension Symbol
	(4) Very High Comprehension Symbol
	(5) Short Text
	(6) Long Text
Question Delay	(1) 0-second Delay
	(2) 50-second Delay
Message Repetition	(1) First Repetition
	(2) Second Repetition
Age	(1) Younger (18 - 22 yrs)
	(2) Older (65 - 80 yrs)
Gender	(1) Male
	(2) Female

The following section provides details on the two primary independent variables, message type and question delay. Following this is a description of the dependent variables collected in this experiment.

Primary Independent Variables

A repeated measures design was used that had two primary independent variables: (1) message type and (2) question delay. Message type, a within-subjects variable, had six levels: (1) very low comprehension symbol, (2) medium low comprehension symbol, (3) medium high comprehension symbol, (4) very high comprehension symbol, (5) short text, and (6) long text. The results from Dewar et al. (1994), and Saunby, Farber, and DeMello (1988), were used to define symbol comprehension. In the Dewar et al. study, 480 drivers aged 18 to 70+ were tested on traffic sign comprehension and familiarity. Their task was to provide a written response for the meaning of different traffic signs. A measure of comprehension was calculated from these responses. A similar procedure was used in the Saunby et al. study. For the present study, symbols labeled very low comprehension were those with Dewar et al. and Saunby et al. comprehension ratings ranging from 10 percent to 11 percent. Medium low comprehension symbols ranged from 34.8 percent to 68.1 percent. Medium high comprehension symbols ranged from 77.1 percent to 85 percent, while very high comprehension symbols ranged from 91.9 percent to 99.8 percent. Of the 22 symbols that fell into these 4 categories, 3 were very low, 3 were medium low, 5 were medium high, and 11 were very high. The set of 22 symbol messages used is included in appendix B. Eight additional symbols were presented during the simulation that were not rated on comprehension in either the Dewar et al. or Saunby et al. studies. These symbols were not included in the Analysis of Variance (ANOVA), but were analyzed in post hoc multiple comparisons.

As outlined in table 2, a total of 40 text messages were examined; 20 each of the long and short types. For every long text message, there was a corresponding short text message. For example, one long text message read, "School crossing ahead, children present between 8 AM - 4 PM." The corresponding short text message read, "School crossing." Long text messages had a mean of 5.3 words, while short text messages had a mean of 2.5 words.

Table 2. Twenty long and 20 corresponding short text messages.

LONG TEXT MESSAGES	WORD COUNT	SHORT TEXT MESSAGES	WORD COUNT
Pedestrians Not Allowed To Cross Here Due To Heavy Traffic	10	No Pedestrian Crossing	3
School Crossing Ahead Children Present Between 8AM - 4PM	8	School Crossing	2
Greyhound Bus Stop At Next Exit	6	Bus Stop	2
Bicycles Not Allowed On This Road	6	No Bicycles Allowed	3
Right Front Tire Has Low Pressure	6	Tire Pressure Low	3
Amtrak Train Station At Next Exit	6	Train Station	2

Table 2. Twenty long and 20 corresponding short text messages (continued).

LONG TEXT MESSAGES	WORD COUNT	SHORT TEXT MESSAGES	WORD COUNT
Caution Oil Pressure Extremely Low	5	Oil Pressure Low	3
Kentucky Fried Chicken Next Exit	5	Food Next Exit	3
Police Car Approaching From Behind	5	Police Car Approaching	3
Stray Cattle On Road Ahead	5	Cattle Crossing	2
Cross Country Skiing Next Exit	5	Winter Recreation	2
Hitchhiking Is Not Allowed Here	5	No Hitchhiking	2
Stray Deer On Road Ahead	5	Deer Crossing	2
Axle Weight Limit 5 Tons	5	No Heavy Trucks	3
Residential Area Speed Bumps Ahead	5	Speed Bumps Ahead	3
Water Recreation Next Exit	4	Water Recreation	2
Texaco Station Next Exit	4	Gas Next Exit	3
Caution Fallen Rocks Ahead	4	Fallen Rocks	2
Ambulance Approaching From Behind	4	Ambulance Approaching	2
Pedestrian Crossing Ahead	3	Pedestrian Crossing	2
	Mean = 5.3		**Mean = 2.5**

Question delay, a within-subjects independent variable, refers to the time lag between the offset of a message and the appearance of a related question. Recall that each message was presented on the IVIS for 8 s, after which the screen blanked. After the screen blanked a question pertaining to that message was presented. *When* this question was presented, in relation to the offset of the message, defines question delay. Two levels of question delay were investigated: (1) 0-s delay and (2) 50-s delay. A Greco-Latin Square design was used to balance the messages and question delay types.

Primary Dependent Variables

As outlined in table 3, data were collected during three primary phases of the experiment: (1) Subject Recruitment Phase, (2) Test Phase, and (3) Post-Test Phase. The dependent measures collected during each of these three phases are detailed in the following table.

Table 3. Outline of dependent variables and the three phases of data collection.

SUBJECT RECRUITMENT PHASE	TEST PHASE	POST-TEST PHASE
Subject Suitability	Comfort With Computers	Symbol Comprehension
Demographic Information	Measures of In-Vehicle System Message Recognition Accuracy C Recognition Accuracy C Recognition Latency	Comfort With Computers
Driving Experience	Measures of Self-Confidence of In-Vehicle System Message Recognition Response C Self-Confidence in Recognition Response C Latency of Self-Confidence Response	
Technology Use Experience	Measures of Simulated Driving Performance C Mean Lane Position C Standard Deviation Lane Position C Crash Occurrence	

Subject Recruitment Phase

There were two missions of the Subject Recruitment Phase. The first was to determine a subject's suitability for this experiment. This was assessed via the "Subject Selection Phone Questionnaire" (appendix A). To be suitable for this study, subjects were required to have an active driver's license, drive at least twice per week, and report to rarely have had past difficulties with motion sickness.

After it had been determined that a subject qualified to participate, they were then administered the "Driver Demographic Questionnaire". This questionnaire elicited responses pertaining to a wide range of demographic characteristics (e.g., age, marital status), driving experience (e.g., number of miles driven annually), and experience with technology (e.g., rated comfort using new technology).

Test Phase

The Test Phase consisted of data collection in the Battelle Human Performance Laboratory prior to and during simulator testing. Prior to beginning the simulation portion of the study, drivers were administered a questionnaire entitled, "Experiment 5A, Pre-Study Questionnaire: How Comfortable Are You With Computers?" (appendix A). As the name suggests, the purpose of this questionnaire was to assess the subjects' self ratings of comfort when working with computers.

After this brief questionnaire had been completed, the simulation portion of the testing phase began. The following data were recorded:

- **Accuracy of Recognition Question Response:** Measured in percent correct; accuracy of recognition to in-vehicle system message questions.
- **Latency of Recognition Question Response:** Measured in milliseconds; response time to in-vehicle system message recognition.
- **Self-Confidence in Recognition Question Response:** Measured on a scale from 0 (low) to 100 (high); rated self-confidence in responses to in-vehicle system message recognition questions.
- **Latency of Self-Confidence in Question Response:** Measured in milliseconds; time to respond to self-confidence questions.
- **Mean Lane Position:** Measured in feet; mean distance from the center of the subject's lane.
- **Standard Deviation Lane Position:** Measured in feet; standard deviation distance from center of the subject's lane.
- **Crash Occurrence:** Measured in percent; incidents of crashes.

Post-Test Phase

At the conclusion of the simulator portion of the testing phase, data were collected on: (1) comfort working with computers and (2) symbol comprehension. The "Experiment 5A, Post-Study Questionnaire: How Comfortable Are You With Computers?" was administered after the symbol comprehension test. Identical to the comfort with computers questionnaire administered during the Testing Phase, the purpose of this questionnaire was to assess how comfortable subjects were with computers, comparing before and after driving the simulator. To assess symbol comprehension, a symbol recall test was administered. This test served to gather data on drivers' comprehension of 20 different symbols.

PROCEDURES

The initial screening for participant suitability was done by telephone. The "Subject Selection Phone Questionnaire" and the "Driver Demographic Questionnaire" were administered at this time. The purpose of the screening procedure was to ensure a homogeneous population in terms of the age groups, driving knowledge, and experience. Another goal of this screening was to rule out subjects who might be prone to motion sickness and have difficulty driving the simulator. Prospective subjects who either did not have an active driver's license, drove less than twice per week, or reported experiencing motion sickness "often" were eliminated from the subject pool. Those who met the outlined criteria were administered a series of demographic characteristics questions and scheduled for a laboratory testing time. The demographic questionnaire was given during the telephone interview to reduce the time required of subjects for the testing session.

At the testing site, subjects filled out a written consent form. Subjects began the experiment by listening to instructions about the purpose of the study and their task. Subjects were told that the data from their participation would allow researchers to better understand how people use an ATIS. Subjects were then administered the pre-test comfort with computers questionnaire (appendix A).

Prior to beginning the data collection portion of the simulated driving, subjects were given a practice session. In the practice session, subjects were given instruction about, and were able to use, the touch-screen and operate the vehicle. The length of the practice session was approximately 5 minutes. The experimenter was present in the passenger seat during the practice session. The experimenter's purpose for being in the vehicle was to: (1) provide instruction about the touch-screen and vehicle operation, (2) outline the task, and (3) monitor and assist the subject in completing the task correctly. Subjects who could correctly perform the task began the testing scenarios. Subjects who had difficulties completing the task (e.g., were not comfortable with touch-screen operation) were administered a second practice session, also lasting 5 minutes. If a subject continued to have problems with the task, a third 5-minute practice session was allowed. Subjects who could not perform the task after three practice sessions were withdrawn from the experiment. There were three older drivers who, after the practice sessions, were either not able to perform the task or were not comfortable with the experimental procedure and withdrew from the study. In addition, the data from four older drivers and one younger driver were not used due to simulator and/or experimenter error.

After the practice session, each subject "drove" three simulated scenarios, lasting 34 minutes in length. Drivers were given a brief break between each scenario. The simulator was programmed to maintain a constant acceleration (automatic cruise control) and, therefore, in terms of operating the vehicle, subjects were only required to steer. The driver's task was threefold: (1) to safely operate the vehicle, (2) to view the traffic and traveler-related messages displayed on the top screen, and (3) to respond to the bottom screen questions that pertained to the top screen messages. Periodically, questions would appear on the bottom screen that pertained to the driving scene (e.g., "was the cross traffic that you just passed on your right or left side?"). The purpose of these "distraction" questions was to help keep the drivers focused on the driving events, and watching the road, rather than only watching and waiting for messages. A total of 16 distraction questions were administered over the course of the simulated drive.

When a traffic or traveler message had reached the in-vehicle display, a recorded voice instructed the driver to look at the top screen. Each message remained on the screen for 8 s, after which the screen blanked. A tone informed drivers when a question appeared on the bottom touch-screen. After reading the question, drivers would select (touch) one of two response boxes (i.e., forced-choice). For example, for the "school crossing" message, the question read, "What type of crossing is ahead?" The response choices were "pedestrian" or "school." After answering the question, a follow-up question immediately appeared that queried the driver on the confidence that he/she had in the previous response. A horizontal scale was presented on the bottom screen that ranged from 0 (Very Unsure) to 100 (Very Sure). By touching a point on the scale, drivers could indicate their confidence. Drivers were allowed 15 s to answer both questions, after which the screen blanked.

During the course of the simulated drive, the six message types outlined (i.e., very low comprehension symbol, medium low comprehension symbol, medium high comprehension symbol, very high comprehension symbol, short text, and long text) would appear and, either immediately or 50 s after leaving the screen, drivers were asked the above described forced-choice questions.

At the conclusion of the three scenarios, drivers began the Post-Test Phase of the experiment. One very low comprehension symbol, 1 medium low comprehension symbol, 2 medium high comprehension symbols, 3 very high comprehension symbols, and 13 novel symbols of previously untested comprehension were presented. Half of the symbols presented during the post-test, including three of the previously untested comprehension symbols, had been previously presented during the simulation.

The 13 previously untested comprehension symbols were selected as reasonable symbol possibilities that might appear on an IVIS. This category of symbols was labeled "previously untested" because they were not previously rated on comprehension in either the Dewar et al. (1994) or Saunby et al. (1988) studies. Also, all of the previously untested comprehension symbols are not listed in the 1988 published version of the Manual on Uniform Traffic Control Devices for Streets and Highways (MUTCD) (Federal Highway Administration, 1988). Appendix B includes the entire set of 20 symbols presented.

The procedure for the Post-Test Phase was as follows: each of the symbols, along with a label of their meaning, was displayed one at a time, each for 8 s, to subjects on the in-vehicle screen. After all 20 symbols had been presented, in a different random order for each subject, each symbol was again presented. The subject's task was to write down the name of the icon on a response sheet. After writing down a response, the subject touched the touch-screen and another symbol was presented. Unlike the simulator portion of the Test Phase, in which the subjects had a *recognition task*, the Post-Test Phase involved a *recall task*. Once all 20 responses had been given, the subject was then re-presented the "How Comfortable Are You With Computers?" questionnaire. Once completed, subjects were debriefed and paid for their time.

RESULTS

As outlined in table 3, this experiment can be divided into three phases of data collection. These consist of: (1) the Subject Recruitment Phase, (2) the Test Phase and, (3) the Post-Test Phase. Each of these three data collection phases are outlined in turn. For details of all analyses conducted for experiment 5A, please refer to appendix C for the relevant ANOVA tables.

SUBJECT RECRUITMENT PHASE

Recall that one of the primary functions of the Subject Recruitment Phase was to screen prospective subjects for experiment suitability. Subjects who met the scoring criteria indicated on the "Subject Selection Phone Questionnaire," provided in appendix A, were scheduled for participation. As such, this subject selection questionnaire can be thought of as a "go, no-go gate" in which all subjects who eventually participated met the outlined scoring criteria.

A second function of the Subject Recruitment Phase was to administer the "Driver Demographic Questionnaire." The purpose of this particular questionnaire was to collect demographic data and information on participant driving and technology experience. A sample of the results from the "Driver Demographic Questionnaire" is presented in table 4.

Table 4. Sample of the primary results from the "Driver Demographic Questionnaire" (as administered for experiment 5A).

DRIVER AGE GROUP	DEMOGRAPHIC			DRIVING EXPERIENCE			TECHNOLOGY EXPERIENCE
	AGE	OWN AN AUTOMOBILE	YEARS LIVING IN SEATTLE	YEARS AS A LICENSED DRIVER	YEARS DRIVING IN SEATTLE	AVERAGE MILES DRIVEN ANNUALLY	COMFORT USING NEW TECHNOLOGY (0 = LOW, 100 = HIGH)
YOUNGER	19.8	56%	10.7	3.89	2.94	8599	78.1
OLDER	71.3	100%	33.0	49.1	30.0	7500	69.3

TEST PHASE

The Test Phase of this experiment consisted of data collection prior to and during simulator testing. Subjective data were collected prior to simulator testing, while both subjective and objective data were collected during simulator testing.

As outlined in the *Procedures*, prior to simulator testing subjects were administered the "Experiment 5A, Pre-Study Questionnaire: How Comfortable Are You With Computers?" This questionnaire was also administered after the simulator testing in the Post-Test Phase. The results of the pre-test/post-test comparison are outlined later in the Post-Test Phase results section.

During the simulator portion of the experiment, several subjective and objective dependent measures were collected. These included three primary categories of measures: (1) measures of in-vehicle system message recognition, specifically accuracy of question response and latency of question response; (2) measures of self-confidence of in-vehicle system message recognition, specifically self-confidence of recognition question response and latency of recognition question response; and (3) measures of simulated driving performance, specifically mean lane position, standard deviation lane position, and crash occurrence. Each of these three categories of measures is delineated below.

Measures of In-Vehicle System Message Recognition

Two measures of in-vehicle system message recognition were investigated. These consisted of: (1) accuracy of in-vehicle system message recognition question response and (2) latency of in-vehicle system message recognition. Note that, where applicable, Greenhouse-Geisser probability values (Greenhouse andGeisser, 1959) are provided. The relevant ANOVA tables for experiment 5A can be found in appendix C. Note that missing data comprised less than 3 percent of the entire data set. Missing data were due to subjects' failure to respond to the in-vehicle questions. As two repetitions of each message type were presented, data from a completed repetition of a given message type were substituted for the missing repetition values of the same type. This operation served to fill the missing values and complete the data set.

Accuracy of In-Vehicle System Message Recognition Question Response

Figure 2 shows that younger drivers were more accurate when answering the in-vehicle message recognition questions than were older drivers, $F(1, 14) = 5.39$, $p < 0.04$. Notice that both the younger and older drivers scored quite well, percentage-wise, with the younger drivers achieving a mean percent correct score of 94.7 percent and the older drivers achieving 90 percent. However, the relatively small difference between the younger and older populations was sufficiently reliable to reach statistical significance.

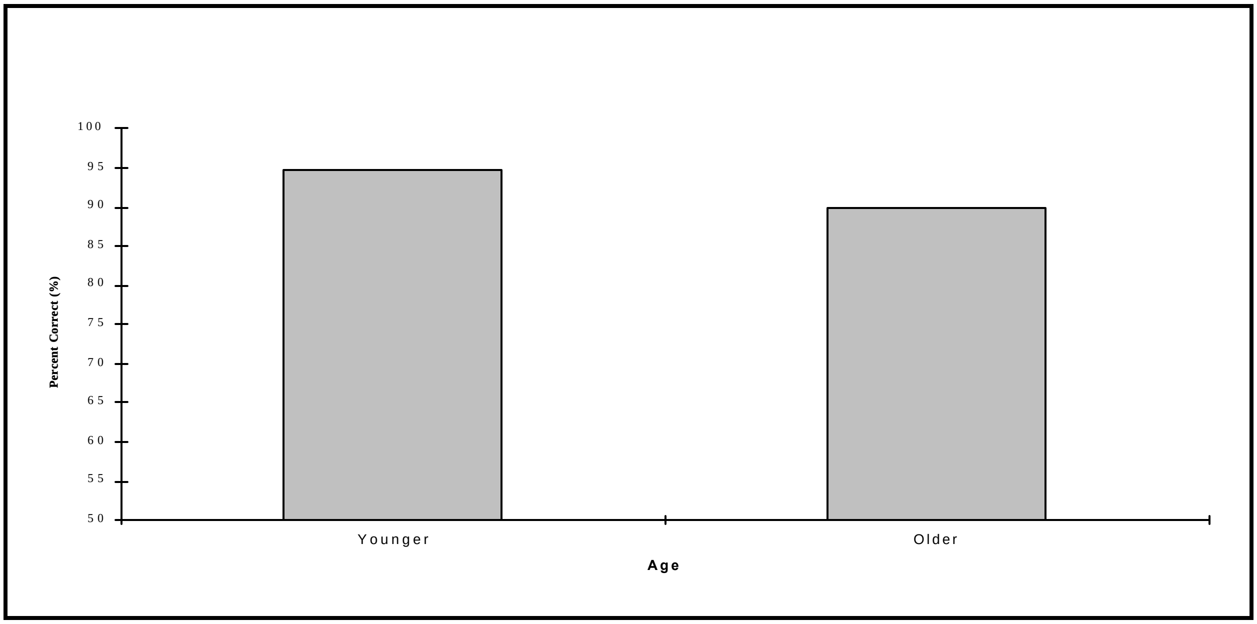


Figure 2. Percent correct as a function of driver age.

Figure 3 shows that drivers were more accurate when responding to questions in the 0-s delay condition than the 50-s delay condition, $F(1, 14) = 30.7$, $p < 0.001$. Questions administered immediately after the message left the IVIS were answered with greater accuracy (M = 94.6 percent) than those with a 50-s delay (M = 90.1percent).

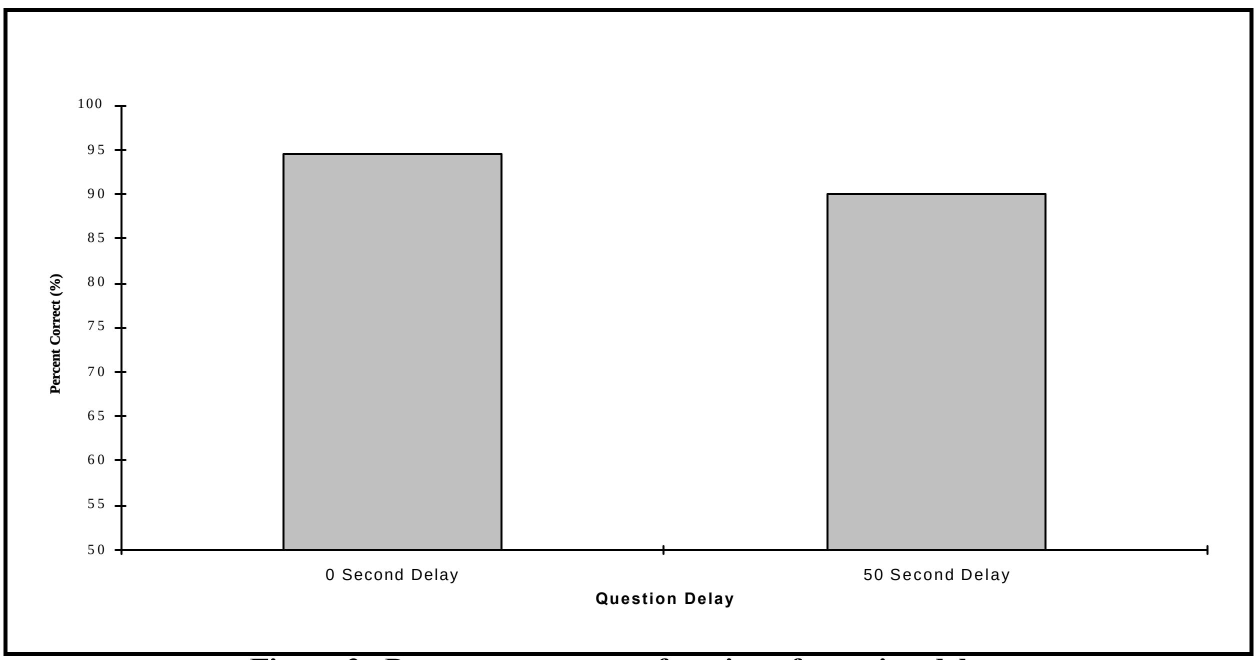


Figure 3. Percent correct as a function of question delay.

Figure 4 shows that response accuracy was affected by the different message types, $F(5, 70) = 3.24$, $p < 0.04$. Though the corrected Huynh-Feldt probability (Huynh & Feldt, 1970) indicated significance, the Greenhouse-Geisser value was marginal ($p = 0.05$). The distraction type was not included in the ANOVA, because distraction events only had questions in the 0-s delay condition. However, the distraction type was examined in a post hoc analysis (Tukey, a = 0.05). Drivers had significantly lower response accuracy scores for the distraction questions (M = 79.7) compared with all other message types, except for the very low comprehension symbol type. Note that in the experimental procedure, drivers were cued to the in-vehicle message presentation and not cued to the distraction events. Because of this confound, it is impossible to determine if the poor recognition of the distraction events was due to the event's location (i.e., outside of the vehicle) or to the cue received with the in-vehicle messages. Despite this confound, the converging results from the other dependent measures (e.g., self-confidence discussed later) suggest that the lack of an auditory cue for the distraction events may have had little or no influence on their recognition. Nonetheless, future research is planned to examine this issue more carefully where cuing is tied to both in-vehicle and out-of-vehicle messages or events.

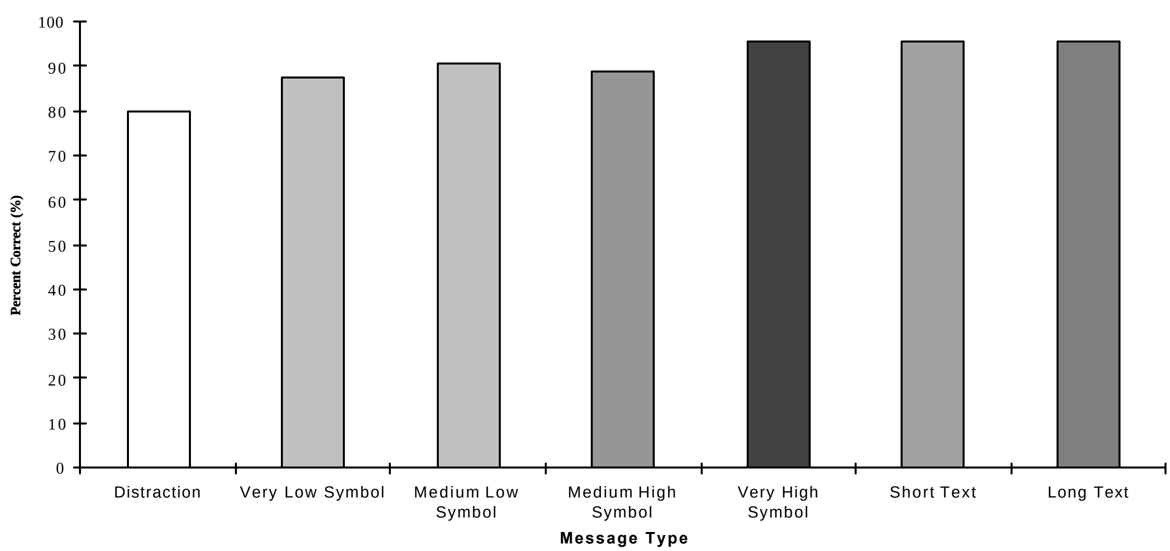


Figure 4. Percent correct as a function of message type.

The Tukey post hoc multiple comparison (a = 0.05) also indicated that the very low comprehension symbol (M = 87.5) was significantly less than the very high comprehension symbol (M = 95.6), the short text (M = 95.7), and the long text (M = 95.8) messages.
Though not illustrated in figure 4, the percent correct for the "previously untested comprehension" symbols was 95 percent. A Tukey post hoc multiple comparison (a = 0.05) indicated that the previously untested comprehension symbol was significantly higher than the distraction type, the medium high comprehension symbol type, and the very low comprehension symbol type.

Figure 5 illustrates that younger drivers had similar percent correct scores for both the 0-s and 50-s question delay conditions. In addition, older drivers had a mean percent correct score in the 0-s delay condition that was similar to that of younger drivers. However, older drivers had significantly lower percent correct scores in the 50-s delay condition compared with: (1) their scores in the 0-s delay condition, and (2) the younger driver scores in both delay conditions, $F(1, 14) = 11.0$, $p < 0.006$. A Tukey post hoc comparison of the means in the figure confirmed this result ($p < 0.01$).

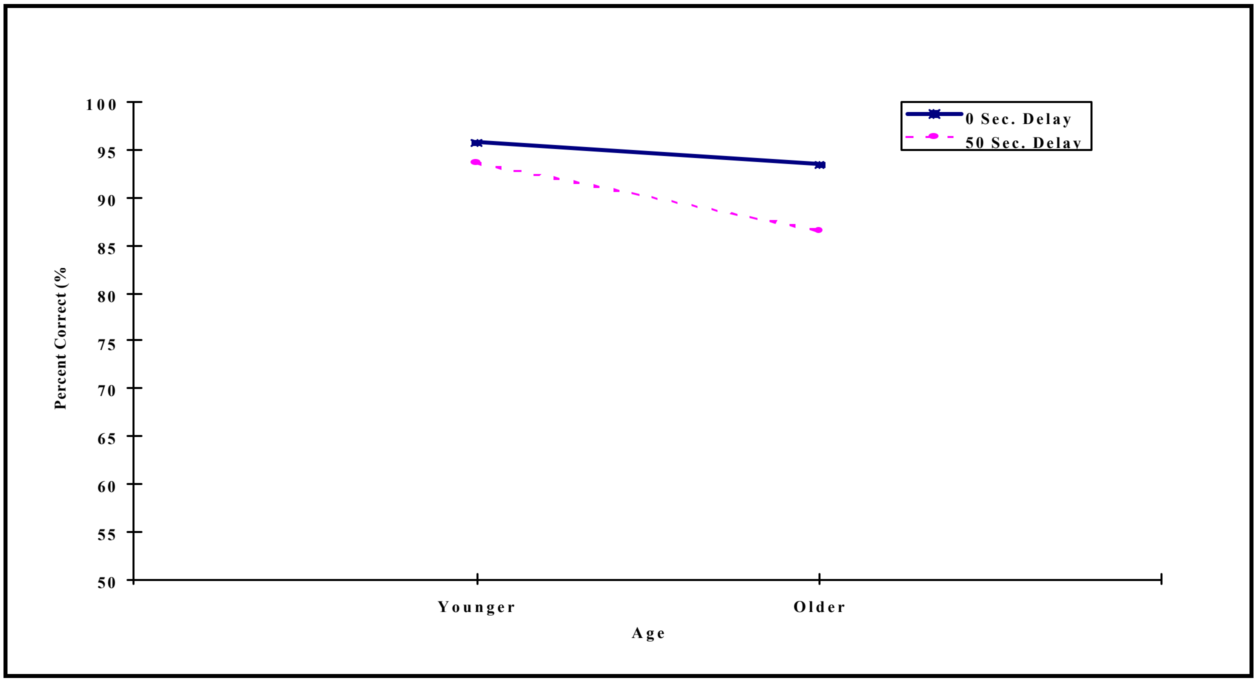


Figure 5. Percent correct as a function of driver age and question delay.

The interaction between message type, question delay, and repetition proved to be significant, $F(5, 70) = 5.92$, $p < 0.004$. This three-way interaction is shown plotted in figure 6. The inconsistent pattern of the function lines makes the meaning of this interaction difficult to interpret.

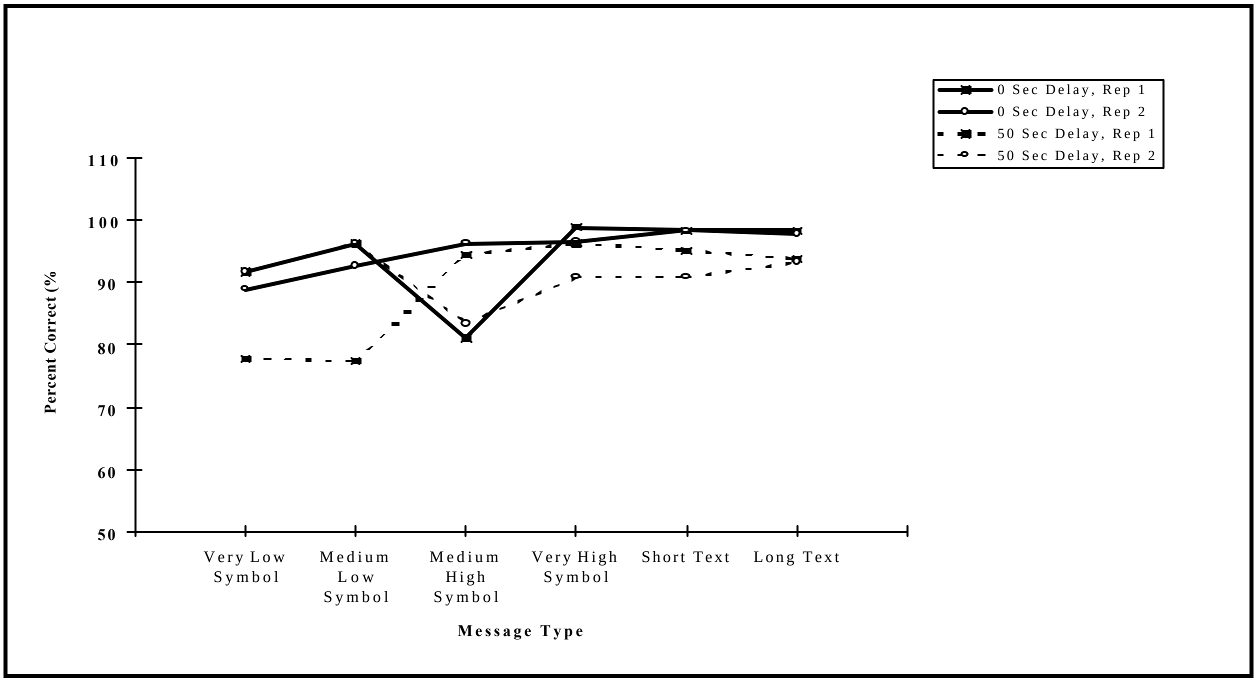


Figure 6. Percent correct as a function of message type, question delay, and repetition.

Latency of In-Vehicle System Message Recognition Question Response

As illustrated in figure 7, latency values were greater in repetition 1 compared with repetition 2, $F(1, 14) = 11.6$, $p < 0.04$. Mean response latency in repetition 1 was 3.57 s, while for repetition 2 it was 3.26 s.

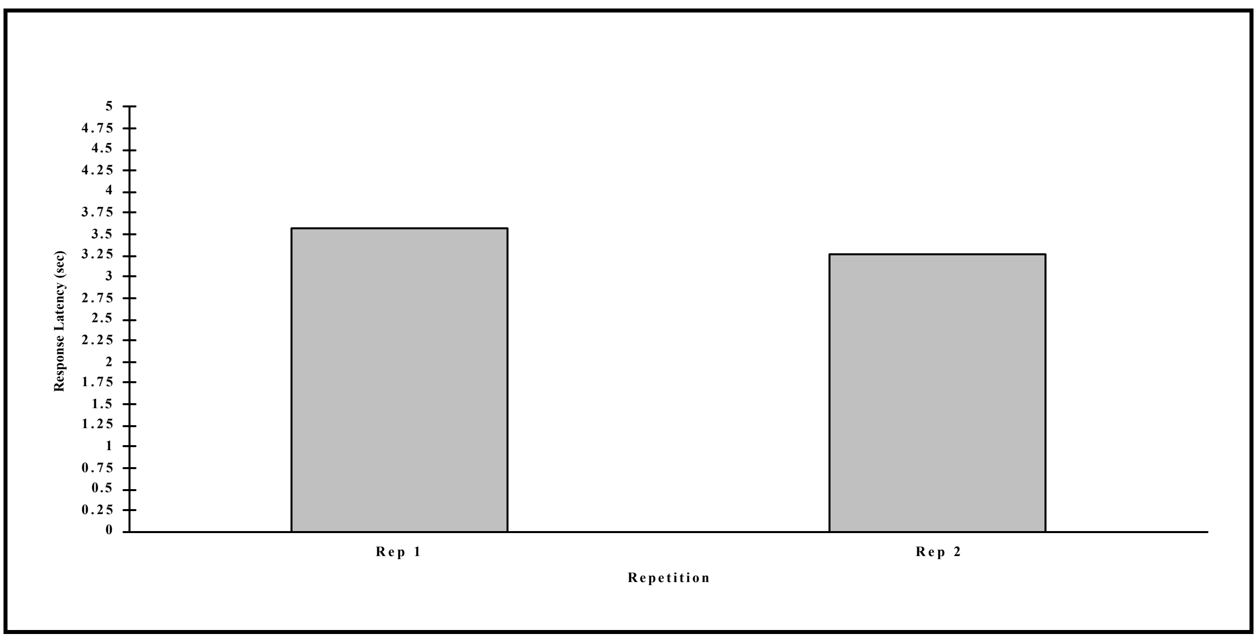


Figure 7. Response latency as a function of repetition.

Figure 8 shows that drivers had significantly lower response latencies in the 0-s delay condition than in the 50-s delay condition, $F(1, 14) = 114$, $p < 0.0001$. As the plot indicates, drivers were faster to respond in the 0-s delay condition (M = 2.9 s) compared with the 50-s delay condition (M = 3.93 s).

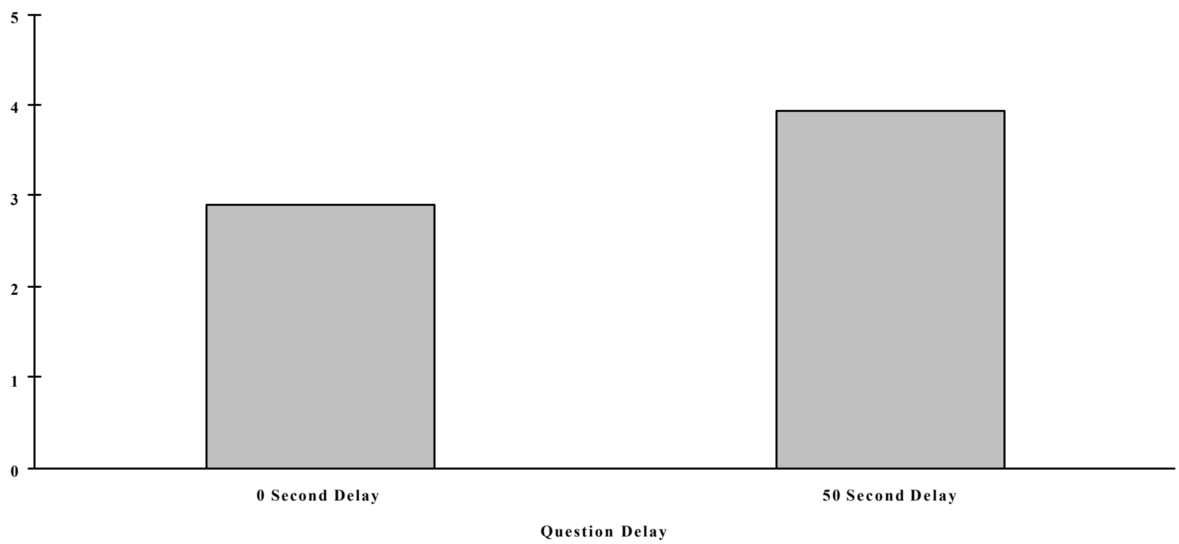


Figure 8. Response latency as a function of question delay.

Figure 9 shows latency to respond to the message recognition questions as a function of message type. Response latency was affected by the different message types, $F(5, 70) = 10.7$, $p < 0.001$. A Tukey post hoc multiple comparison test indicated that the very high comprehension symbol was responded to with a significantly lower response latency than the medium low comprehension symbol ($p < 0.01$) and the very low comprehension symbol ($p < 0.05$). The long text also proved to be responded to significantly faster than the medium low comprehension symbol ($p < 0.01$). Note that the distraction questions (i.e., questions pertaining to the driving scene and not to the in-vehicle messages) were only administered immediately after the event occurred in the driving scene and, as such, there was no 50-s delay condition. Therefore, the distraction type was not included in the repeated measures model. However, for comparison, it is worth noting that the mean response latency value for the distraction questions was 4.24 s. A Tukey post hoc multiple comparison test (a = 0.05) indicated that drivers were slower to respond to the distraction type compared with all other types, except for the medium low comprehension symbol. Mean latency for the previously untested comprehension symbol type was 3.38 s. A Tukey post hoc comparison (a = 0.05) found that response latency for the previously untested comprehension type was significantly less than the distraction type and the medium low comprehension symbol type.

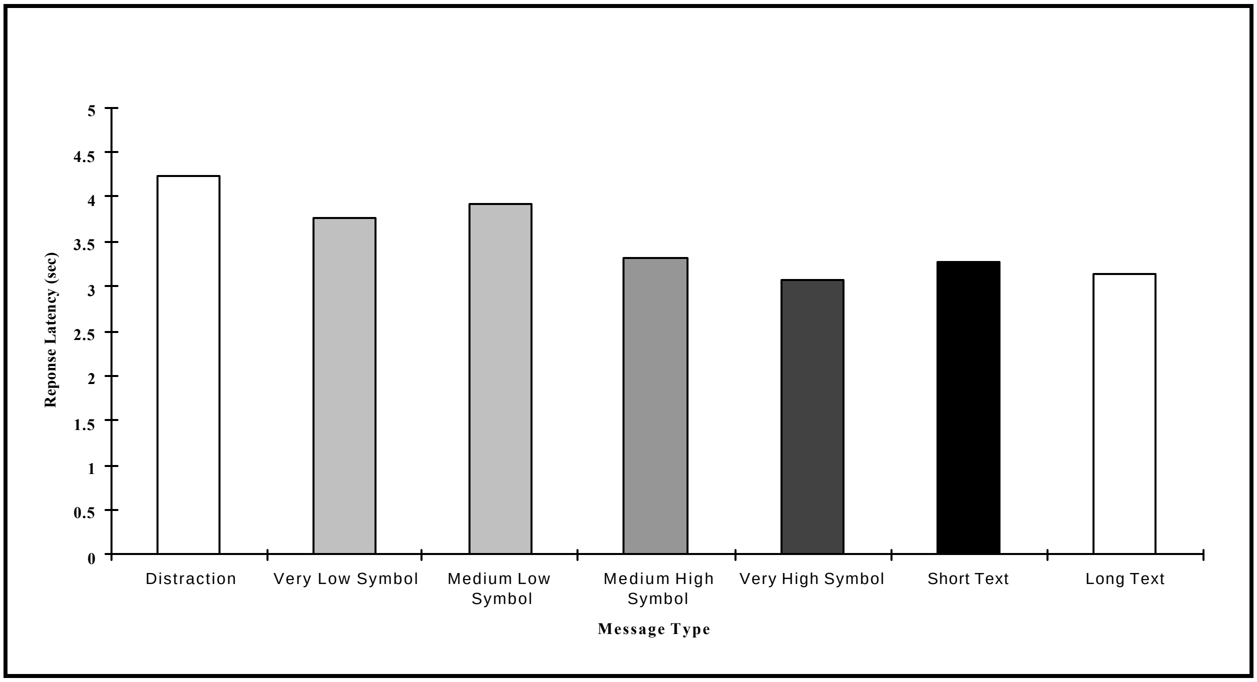


Figure 9. Response latency as a function of message type.

The results outlined in figure 9 generally support those found with the percent correct dependent measure (figure 4): for message type, as percent correct scores increased, response latency decreased. A regression analysis was conducted to support this finding, where percent correct scores (x) were used to predict response latency scores (y). The resulting regression equation was $y = -0.007x + 9.46$ with an R square of 0.75.

Figure 10 shows that response latency varied as a function of driver age and message type, $F(5, 70) = 3.63$, $p < 0.03$. Though the functions for the high symbols and text messages are nearly parallel, the interaction appears to be due to the intersecting function lines across the low symbol message types. A Tukey post hoc comparison (a = 0.05) was performed to support this finding. The very low comprehension symbol type for the older drivers and the medium low comprehension symbol type for the younger drivers were significantly higher than several of the other plotted means. However, aside from these two data points, none of the other Message Type x Age combinations significantly differed.

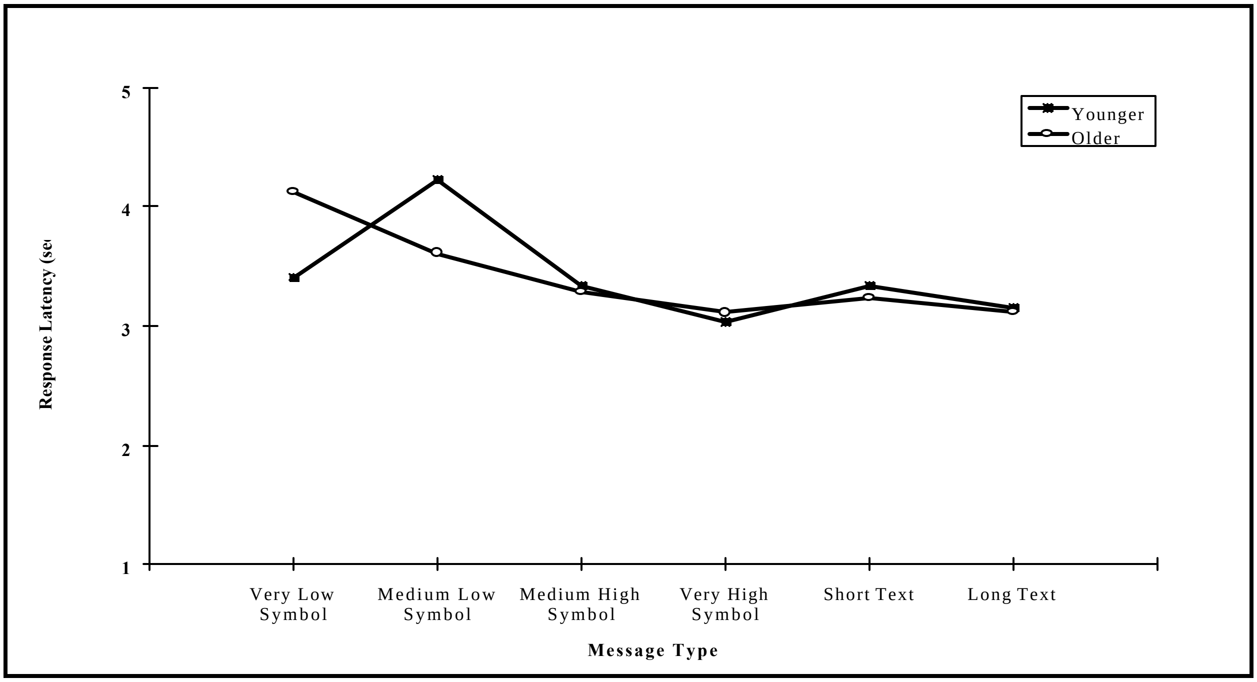


Figure 10. Response latency as a function of message type and driver age.

Figure 11 shows mean response latency for the various message types across repetition. Though the mean values appear stable for repetition 2 across message type (M = 3.26 s), response latency values for the very low comprehension symbol and medium low comprehension symbol types for repetition 1 are both more than 4 s. These higher response latency values for the low comprehension conditions are represented in the significant Message Type x Repetition interaction, $F(5, 70) = 4.38$, $p < 0.02$. That is, the interaction appears to be localized in the very low comprehension symbol and medium low comprehension symbol types. A Tukey post hoc (a = 0.05) confirmed this result, where the response latencies for the low comprehension symbol types for repetition 1 were significantly higher than the response latencies for several of the other message types for both repetitions. That is, except for the response latencies for the very low comprehension and medium low comprehension symbol types in repetition 1, none of the other message types, across either repetition, reached statistical significance.

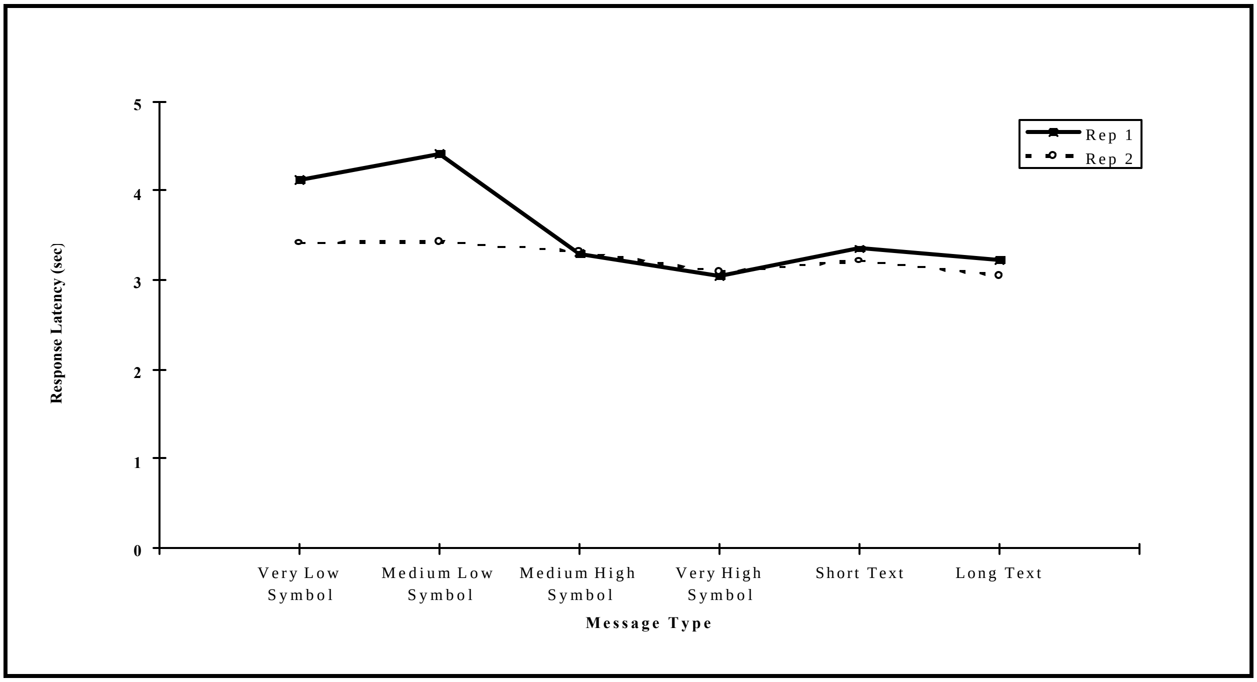


Figure 11. Response latency as a function of message type and repetition.

Figure 12 shows response latency as a function of message type, question delay, and repetition. Though the Repetition x Question Delay x Message Type interaction proved to be significant, $F(5, 70) = 8.25$, $p < 0.0001$, the inconsistent patterns of the function lines make it difficult to nterprete this result.

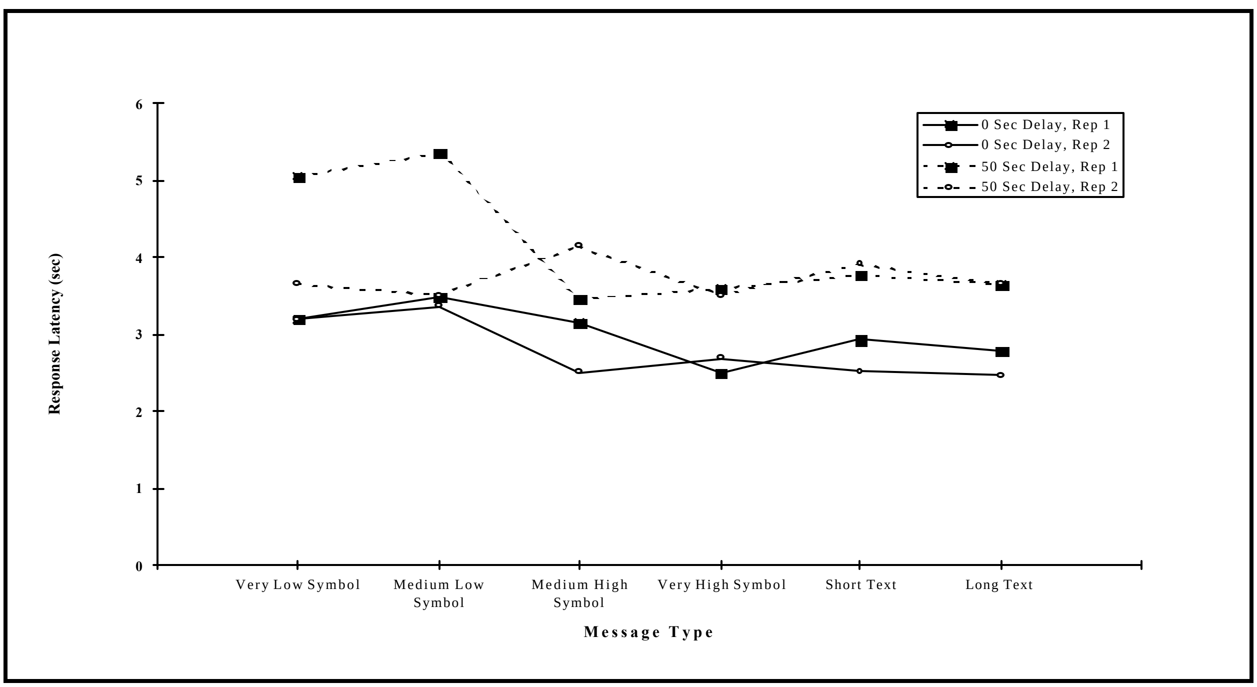


Figure 12. Response latency as a function of message type, question delay, and repetition.

Measures of Self-Confidence of In-Vehicle System Message Recognition

Two measures of self-confidence, with respect to in-vehicle system message recognition responses, were investigated. These consisted of: (1) rated self-confidence of in-vehicle system message recognition question response and (2) latency to provide the rated self-confidence of in-vehicle system message recognition response. As in reporting the previous results, where applicable, the Greenhouse-Geisser probability values are provided. The relevant ANOVA tables for experiment 5A can be found in appendix C.

Self-Confidence in Recognition Response

Figure 13 illustrates two significant effects. First, with regard to question delay, drivers had higher self-confidence ratings in the 0-s delay condition compared with the 50-s delay condition, $F(1, 14) = 8.26$, $p<0.02$. The mean self-confidence rating in the 0-s delay condition was 82.4, while the self-confidence rating in the 50-s delay condition was 77.5.

The second significant result illustrated in figure 13 shows that rated self-confidence for the 50-s delay condition for repetition 1 was lower than the 0-s delay condition for repetition 1. However, self-confidence ratings for both delay conditions in repetition 2 were similar. This Question Delay x Repetition interaction proved to be significant, $F(1, 14) = 6.4$, $p<0.03$. A Tukey post hoc test of the means support this result whereby drivers self-confidence was significantly lower in the 50-s delay/repetition 1 condition than in the 0-s delay/repetition 1 condition ($p<0.01$). No statistically significant differences were found between 50-s delay/repetition 2 and 0-s delay/repetition 2 ($p>0.05$). Confidence ratings in the 50-s delay condition, therefore, increased over time to where there were no statistically significant differences between the two delay conditions in repetition 2.

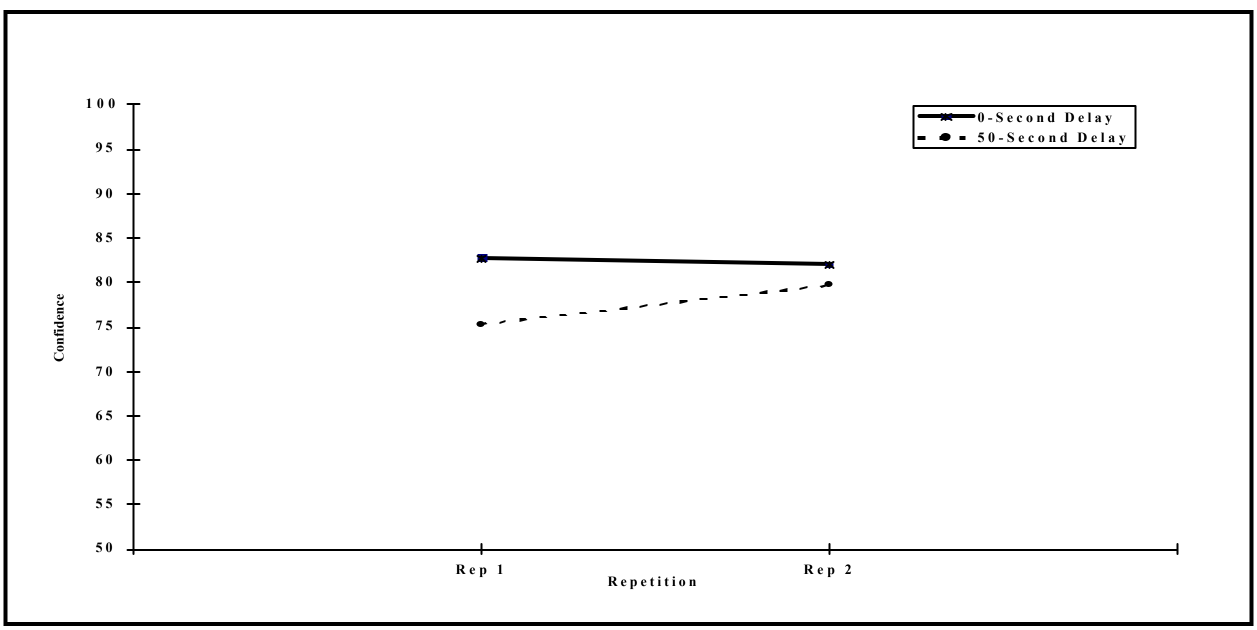


Figure 13. Confidence in responses as a function of question delay and repetition.

Figure 14 shows drivers' rated self-confidence in their recognition question response as a function of message type. Response latency was affected by the different message types, $F(5, 70) = 6.11$, $p<0.001$. As previously noted, the distraction type was not included in the repeated measures ANOVA, but was examined in a post hoc analysis (Tukey, a = 0.05). Drivers' rated self-confidence was significantly lower for the distraction type (M = 67.5) than for all other message types, except for very low comprehension symbols in which no statistical difference was found.

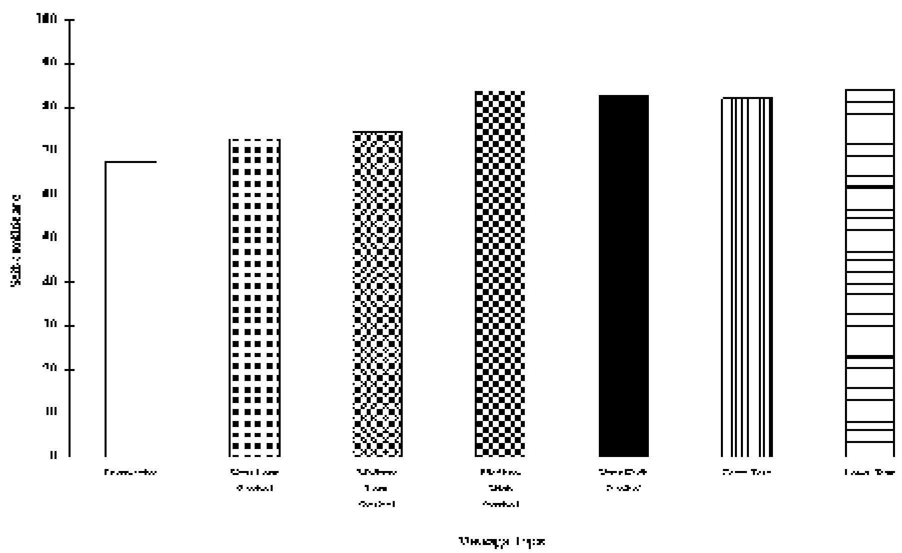


Figure 14. Confidence in responses as a function of message type.

With regard to the possible impact of the auditory cue confound, where the in-vehicle messages were cued but the distraction events were not, the significantly lower self-confidence scores for the distraction type suggest that this confound had little influence on the results. If the auditory cue was a confound that impacted the results, we might expect response latency to be improved with message types that were coupled with an auditory cue (which it was). However, there would be no *a priori* basis for expecting that an auditory cue would increase drivers' rated self-confidence. As such, the reason suggested for the low percent correct scores, high response latencies, and low self-confidence ratings for the distraction type is that these events occurred out of the vehicle and not because they lacked an auditory cue.

A Tukey post hoc multiple comparison (a = 0.05) also indicated that self-confidence ratings for the very low comprehension symbol (M = 72.8) were significantly less than all other message types, except medium low comprehension symbols. Self-confidence ratings for the medium low comprehension symbols were significantly lower than long text, medium high comprehension symbol, and very high comprehension symbol messages. Though not shown in figure 14, mean self-confidence ratings for the previously untested comprehension message type was 84.2. A Tukey post hoc multiple comparison (a = 0.05) indicated that the previously untested comprehension symbols were rated with significantly higher self-confidence than the distraction type, the very low comprehension symbols, and the medium low comprehension symbols.

These results fit nicely with those reported on message type for percent correct and response latency; recall that high percent correct scores were associated with low response latency scores. The findings using self-confidence ratings as the dependent variables have a similar association; high percent correct scores are associated with high ratings of driver self-confidence. A regression analysis was calculated to support this finding, where percent correct (x) was used to predict self-confidence (y). The resulting regression equation was y = 0.93x - 6.2 with an R square of 0.71.

In addition to the apparent relationship between self-confidence ratings and percent correct, there seems to be a relationship between self-confidence and response latency. Comparing figures 14 and 9, low response latencies appear to be associated with high ratings of self-confidence. The resulting regression equation supporting this finding, where response latency (x) was used to predict self-confidence ratings (y), was y = -14.1x + 128.1 with an R square of 0.93.

<u>Latency of Self-Confidence Response</u>

Figure 15 shows latency to provide a self-confidence response for younger and older drivers. Younger drivers responded significantly faster (M = 0.83 s) compared with older drivers (M = 1.6 s), $F(1, 14) = 11.0$, $p<0.006$.

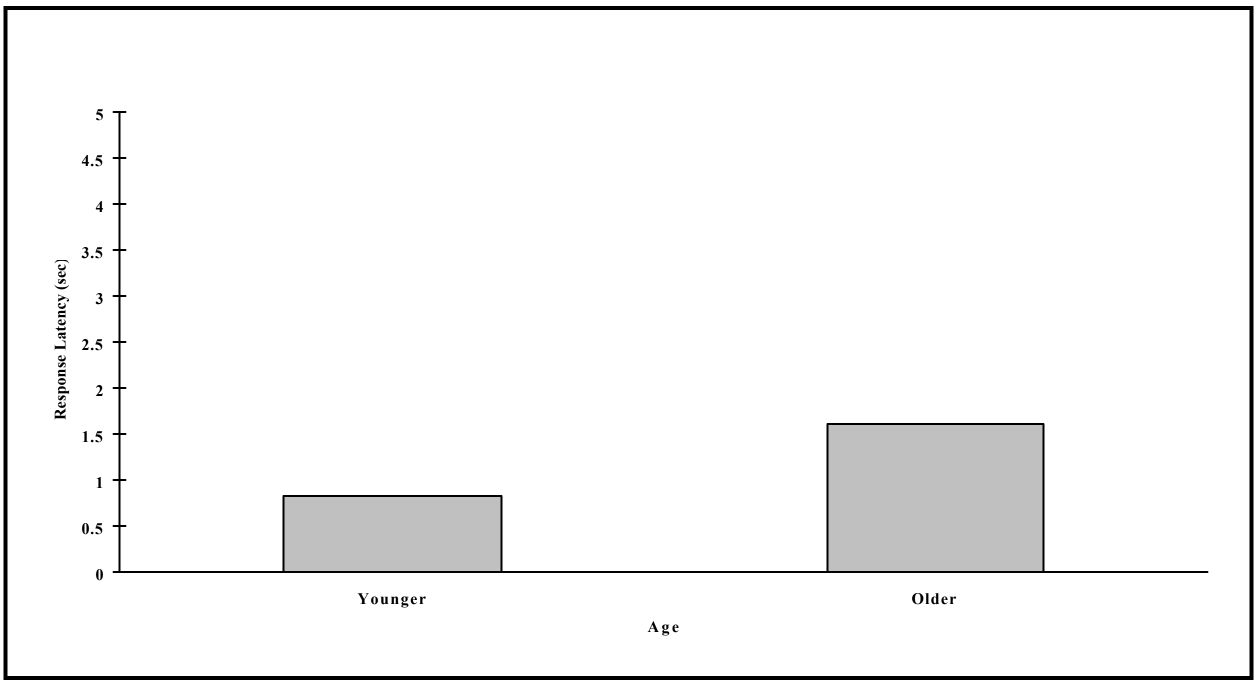


Figure 15. Latency to respond to confidence queries as a function of age.

Figure 16 shows response latency for the self-confidence queries as a function of repetition. As can be seen, repetition had an impact on confidence latency times; drivers were quicker to respond in repetition 2 than in repetition 1, $F(1, 14)$ 11.4, $p<0.006$.

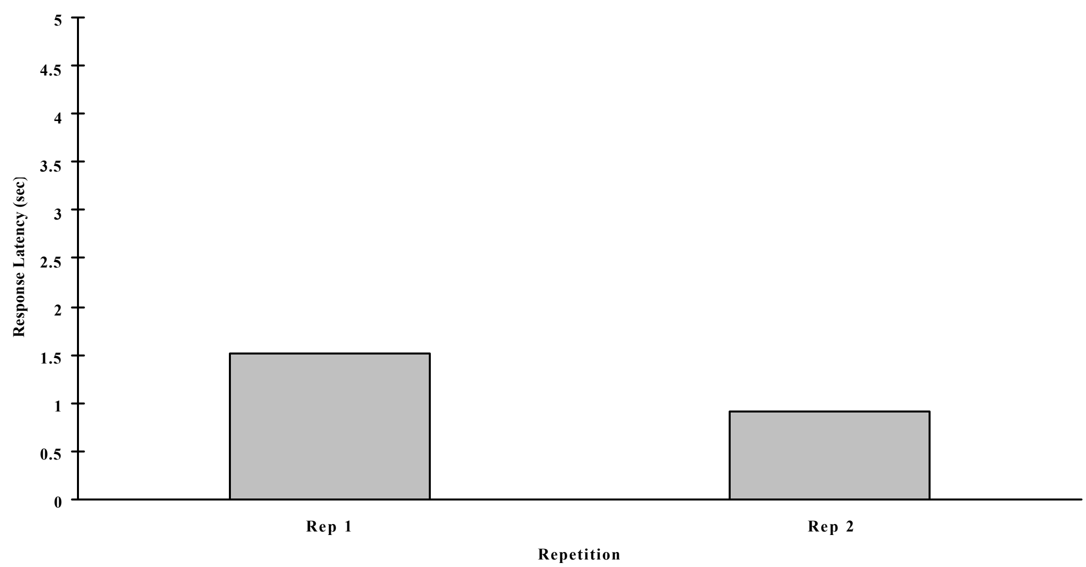


Figure 16. Latency to respond to confidence queries as a function of repetition.

Word Count Difference Between Short and Long Text Messages

As previously outlined in the *Experimental Design* section, a total of 40 text messages were examined in this experiment. Twenty of these text messages were "short" and 20 were "long." For every long text message, there was a corresponding short text message. This information is re-presented in table 5. Long text messages had a mean of 5.3 words, while short text messages had a mean of 2.5 words.

Table 5. Word count for the 20 long and 20 corresponding short text messages.

LONG TEXT MESSAGES	WORD COUNT	SHORT TEXT MESSAGES	WORD COUNT
Pedestrians Not Allowed To Cross Here Due To Heavy Traffic	10	No Pedestrian Crossing	3
School Crossing Ahead Children Present Between 8AM - 4PM	8	School Crossing	2
Greyhound Bus Stop At Next Exit	6	Bus Stop	2
Bicycles Not Allowed On This Road	6	No Bicycles Allowed	3
Right Front Tire Has Low Pressure	6	Tire Pressure Low	3
Amtrak Train Station At Next Exit	6	Train Station	2
Caution Oil Pressure Extremely Low	5	Oil Pressure Low	3
Kentucky Fried Chicken Next Exit	5	Food Next Exit	3

Table 5. Word count for the 20 long and 20 corresponding short text messages (continued).

Police Car Approaching From Behind	5	Police Car Approaching	3
Stray Cattle On Road Ahead	5	Cattle Crossing	2
Cross Country Skiing Next Exit	5	Winter Recreation	2
Hitchhiking Is Not Allowed Here	5	No Hitchhiking	2
Stray Deer On Road Ahead	5	Deer Crossing	2
Axle Weight Limit 5 Tons	5	No Heavy Trucks	3
Residential Area Speed Bumps Ahead	5	Speed Bumps Ahead	3
Water Recreation Next Exit	4	Water Recreation	2
Texaco Station Next Exit	4	Gas Next Exit	3
Caution Fallen Rocks Ahead	4	Fallen Rocks	2
Ambulance Approaching From Behind	4	Ambulance Approaching	2
Pedestrian Crossing Ahead	3	Pedestrian Crossing	2
	Mean = 5.3		**Mean = 2.5**

An analysis was conducted to determine if the difference in the number of words from the various long and short pairings influenced the results. It is possible that longer word count differences may have influenced the results differently than did shorter word count differences. For example, the first word pairing in table 5 has a word count difference of 7 words (i.e., “Pedestrians Not Allowed To Cross Here Due To Heavy Traffic” [10 words] - “No Pedestrian Crossing” [3 words] = 7 words). On the other hand, the last word pairing in the table has a word count difference of only one word (i.e., “Pedestrian Crossing Ahead” [three words] - “Pedestrian Crossing” [one word] = one word). Would subjects score differently on text message pairs that had large word count differences (e.g., seven words) vs. small word count differences (e.g., one)? Since no effort was made to control the word count difference across pairings, it was necessary to explore this further.

Figure 17 illustrates a plot of word count differences, and uses percent correct as the dependent measure. Using a least squares method, a regression analysis was conducted on the function shown in the figure where the delta word count (x) was used to predict percent correct (y). The resulting regression equation was: $y = 0.713x + 98.2$

The R square for this equation was 0.502. However, this conclusion is made with caution considering the small sample size of the groupings (i.e., n = 1 for the six- and seven-word-count groups).

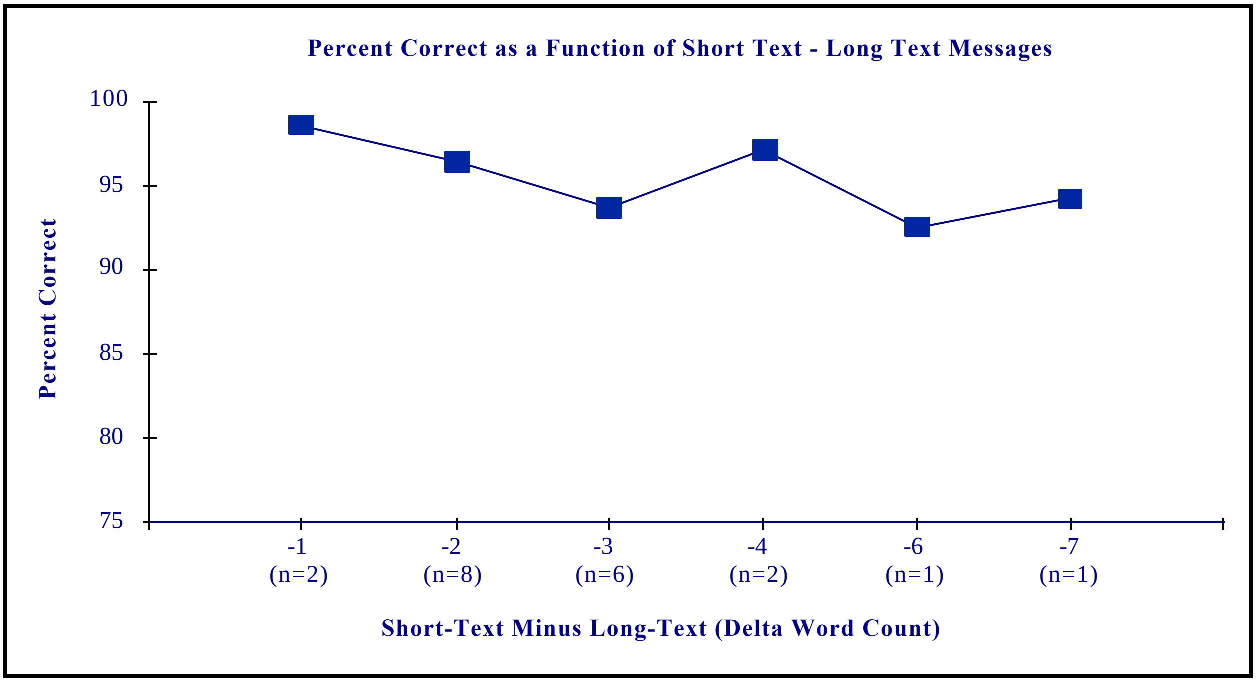


Figure 17. Percent correct as a function of short text - long text messages.

Measures of Simulated Driving Performance

As outlined in the *Procedures*, at different times during the simulated drive subjects were presented with in-vehicle system messages (symbol and text). The messages remained on the in-vehicle system screen for 8 s. After 8 s, the message disappeared. In the 0-s delay condition, a question pertaining to the preceding message was immediately shown on the bottom touch-screen. For the 50-s delay condition, a question was administered 50 s after the message disappeared. When the question was administered, the subject's task was to reach over to the touch-screen and answer the question, along with a self-confidence follow-up question. Up until a question appeared on the touch-screen, safely operating the vehicle was the subject's primary task. However, during the time that a question was on the screen, answering the question became the primary task. In terms of simulated driving performance, it is acknowledged that responding to questions while driving likely invalidated the performance data during this time. As such, a method was used to reduce the data set so that it included only those segments of data where no questions or other messages immediately preceded or followed (i.e., for 8 seconds) a given message. Therefore, the reduced data set included three data collection intervals or windows:

(1) **Pre-message window:** The 8-s period that immediately preceded period 2, where the subject was only steering the vehicle and not attending to messages or questions.

(2) **Message Window:** The 8-s period where the subject was steering the vehicle while a message was displayed.

(3) **Post-message Window:** The 8-s period that immediately followed period 2, where the subject was only steering the vehicle and not attending to messages or questions.

Thus, each record of data was for 24 s in which the subject was only driving for the first 8 s (i.e., seconds 1 - 8), driving while a message was displayed for seconds 9 - 16, and only driving for seconds 17 - 24. Note that this contingency was not met equally for all subjects and, as such, subjects had different amounts of usable data. As an example, let us assume that a given question, administered after a 50-s interval, is “Question A.” The message that Question A refers to is “Message A.” After question A, suppose that a second message is presented called “Message B.” So in this example scenario, we have message A, followed by a 50-s delay where no message or question is presented, followed by Question A. As it turns out in this scenario, Message B is presented 20 s following the *start* of Question A. Fifty seconds following Message B is “Question B.” Subject 1, in this example, may have been slow to answer Question A such that it remained on the screen during window 1 of Message B (perhaps Subject 1 responded to question A after question A had been on the screen for 14 s, leaving only a 6-s pre-message window). Subject 2, on the other hand, may have answered question A after 7 s, allowing for no questions to be present during message B’s window 1. In this example, Subject 2' s data would have been usable, but Subject 1's would not.

The goal of the analyses on the simulated driving performance data was to make comparisons between window 1, window 2, and window 3. Since the messages were displayed for 8 s, we wanted an 8-s time frame for the period before the message was displayed and an 8-s time frame for the period after the message was displayed. During both the before and after windows, no messages or questions could be present on any of the other screens. That is, we wanted before and after time frames where the subject was only driving. Setting up this contingency meant losing much of the driving performance data set. For example, all the instances of 0-s delay were not included if another question was present on the screen within the 8-s “post-message” time frame (i.e., window 3). To reduce the data set, mean values were then calculated that summarized the data for each of these three conditions. For example, mean lane position values were calculated for window 1, window 2, and window 3.

Three measures of simulated driving performance were investigated. These consisted of: (1) mean lane position, (2) standard deviation lane position, and (3) crash occurrence. The relevant ANOVA tables for experiment 5A can be found in appendix C.

<u>Mean Lane Position</u>

Mean lane position values were calculated for the three periods outlined above. Figure 18 illustrates the STISIM lane position values. The entire roadway, from road edge to road edge, is 24 ft. The subject’s vehicle is 6 ft wide. The center of the roadway is referenced as the “0 ft” point, while the center of the subject’s lane is at -6 ft. The lane edge on the right side of the subject’s lane is -12 ft.

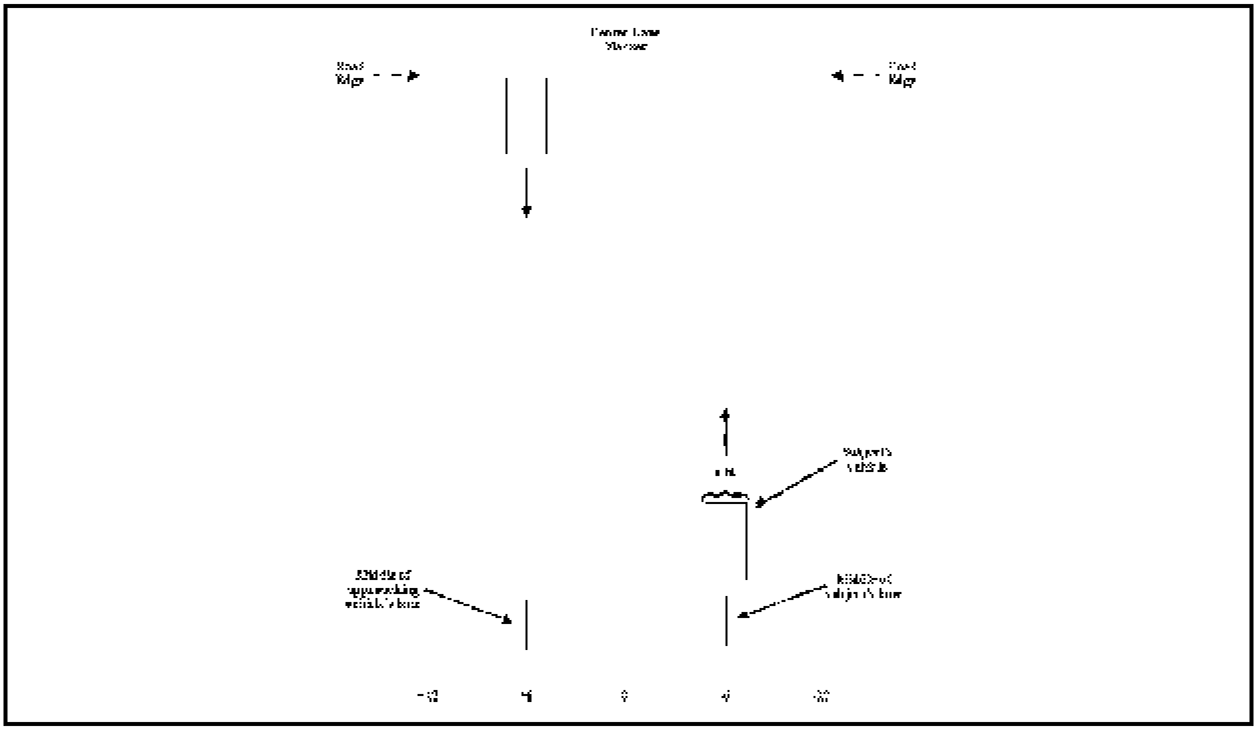


Figure 18. Description of STISIM lane position.

Figure 19 shows mean lane position across repetition. As can be seen, mean lane position in repetition 1 was *further* from the lane center (M = -5.06) than was mean lane position in repetition 2 (M = -5.77). This difference was sufficiently reliable to reach statistical significance, $F(1, 14) = 85.9$, $p<0.001$.

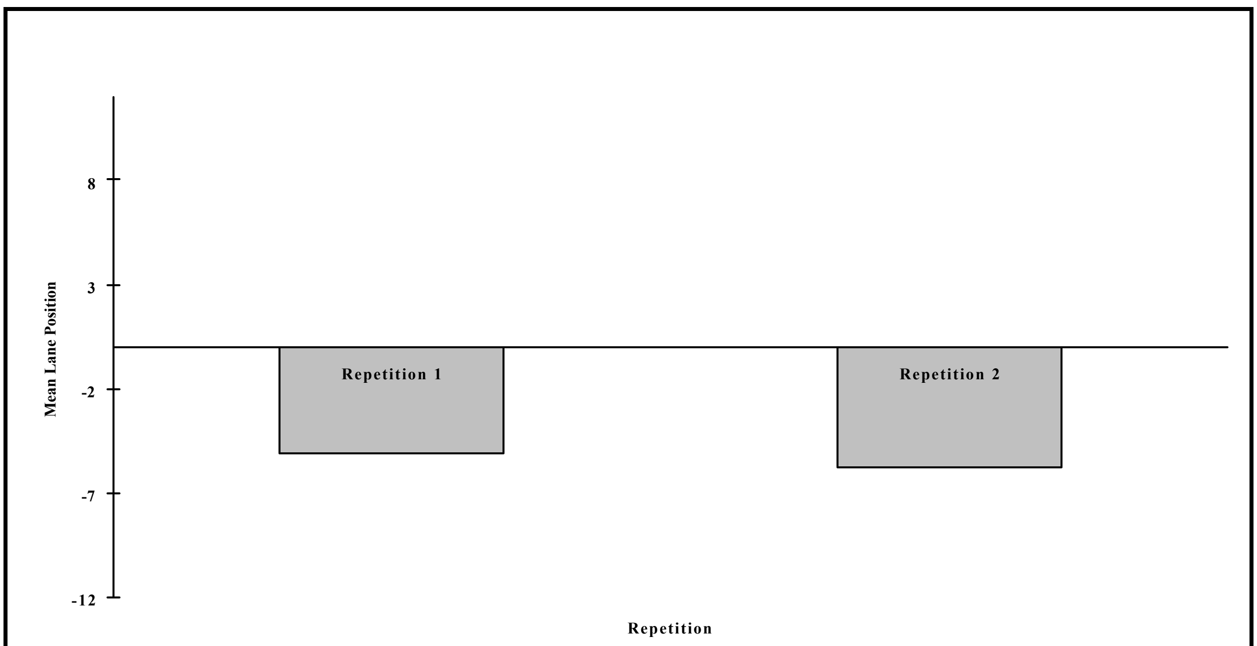


Figure 19. Mean lane position as a function of repetition.

Figure 20 shows the interaction of message type and data collection interval, $F(10, 140) = 3.27$, $p<0.01$. There appears to be no consistent pattern of the function lines in this figure, making the significance of this interaction difficult to interpret.

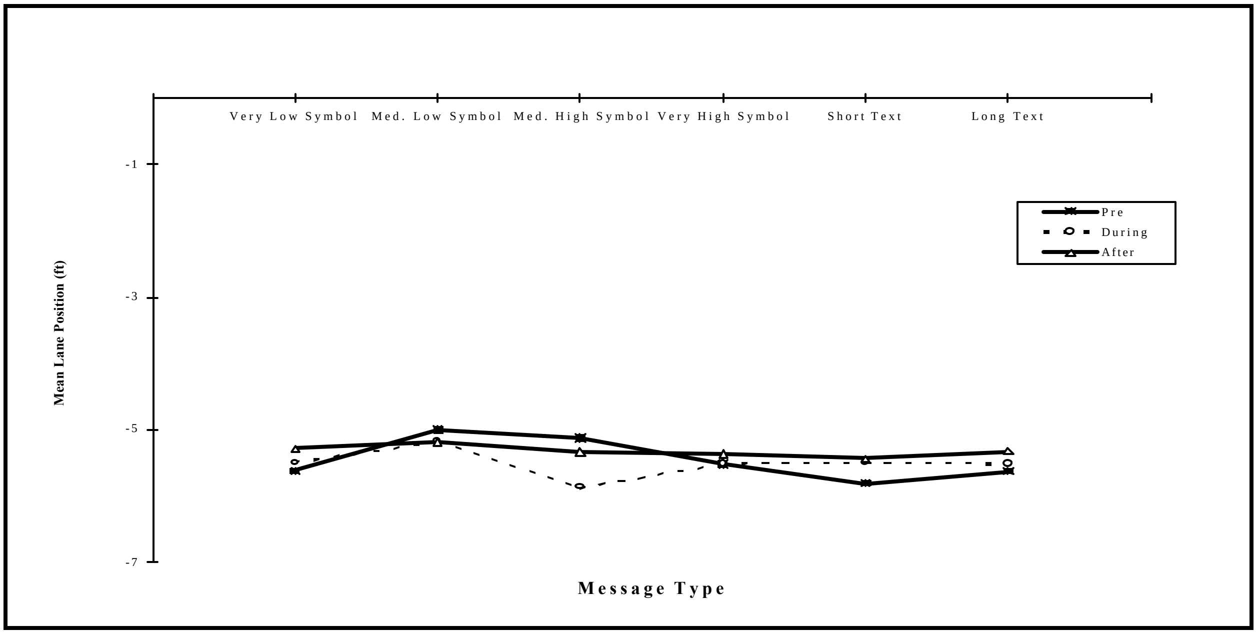


Figure 20. Mean lane position as a function of message type and data collection interval.

The results presented in figures 19 and 20 were the only ones to reach statistical significance. However, a further finding that was not significant, but is worth noting, is shown in figure 20a. Figure 20a shows mean lane position as a function of the data collection interval, or window. As can be seen, lane keeping was similar between the three windows, $p>0.05$. This result is particularly encouraging for IVIS designers since it suggests drivers are able to look at messages on an IVIS without altering their lane keeping.

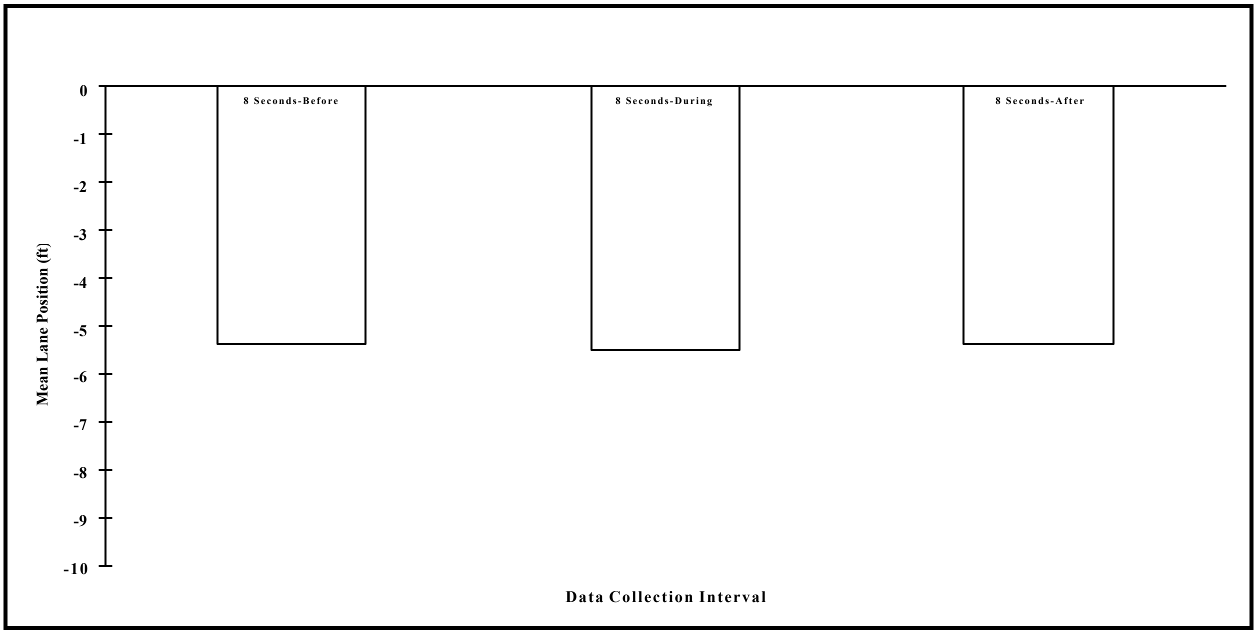


Figure 20a. Mean lane position as a function of the data collection interval.

Standard Deviation Lane Position

Mean standard deviation lane position values were also calculated for each of the three periods. None of the results reached statistical significance. Figure 20b shows that the standard deviation lane position did not vary by data collection interval, $p>0.05$.

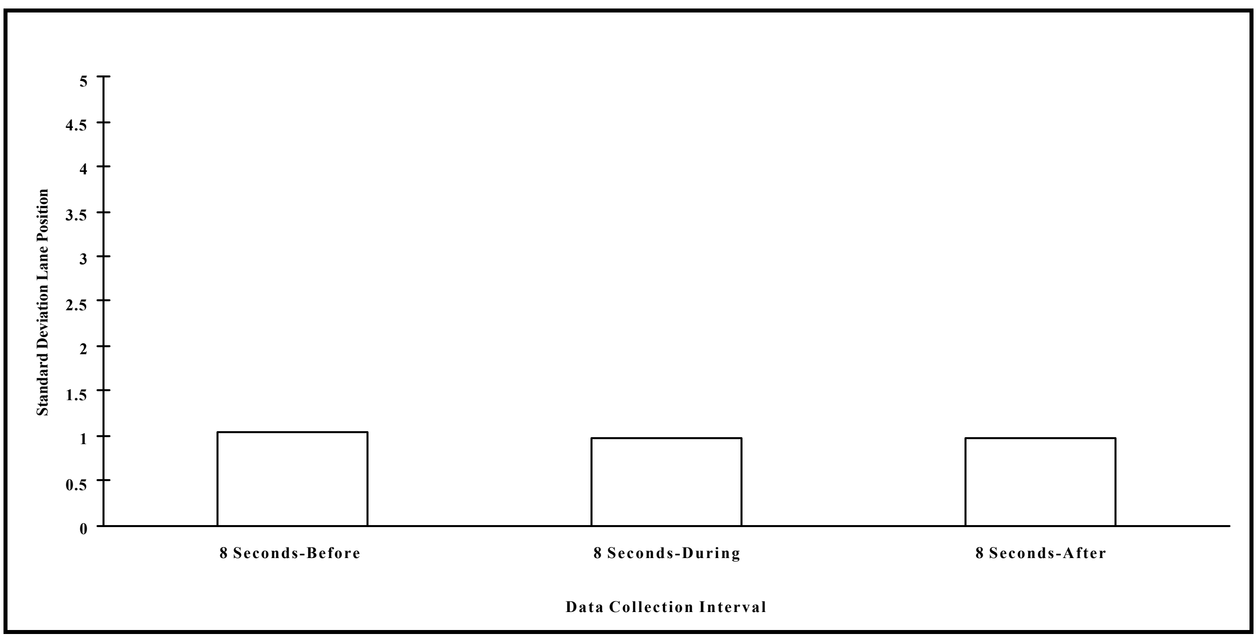


Figure 20b. Standard deviation lane position as a function of the data collection interval.

Crash Occurrence

Figure 21 shows percent crashes as a function of driver age. As can be seen, older drivers had significantly higher incidence of crashes than did younger drivers, $F(1, 14) = 6.66$, $p<0.03$.

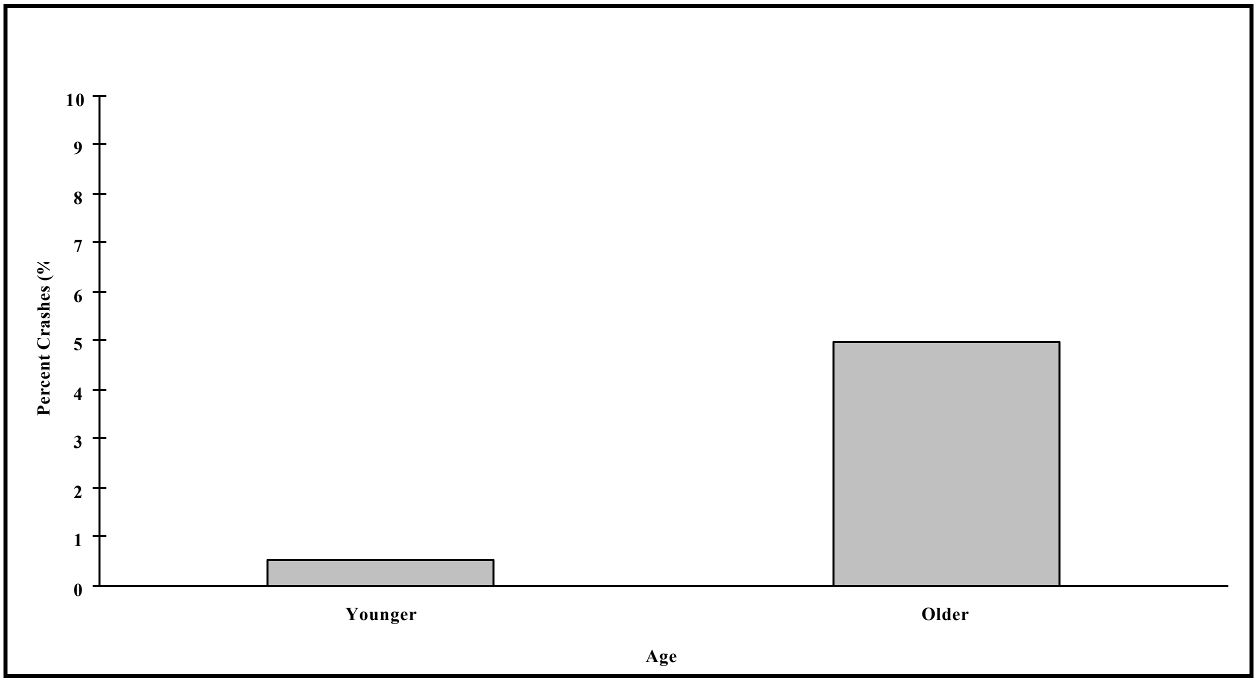


Figure 21. Percent crashes as a function of driver age.

Figure 22 shows the significant three-way interaction of Repetition x Message Type x Data Collection Interval, $F(10, 140) = 3.49$, $p<0.03$. The seeming randomness of the function lines makes this three-way interaction difficult to interpret. Note that no two-way interactions were statistically significant.

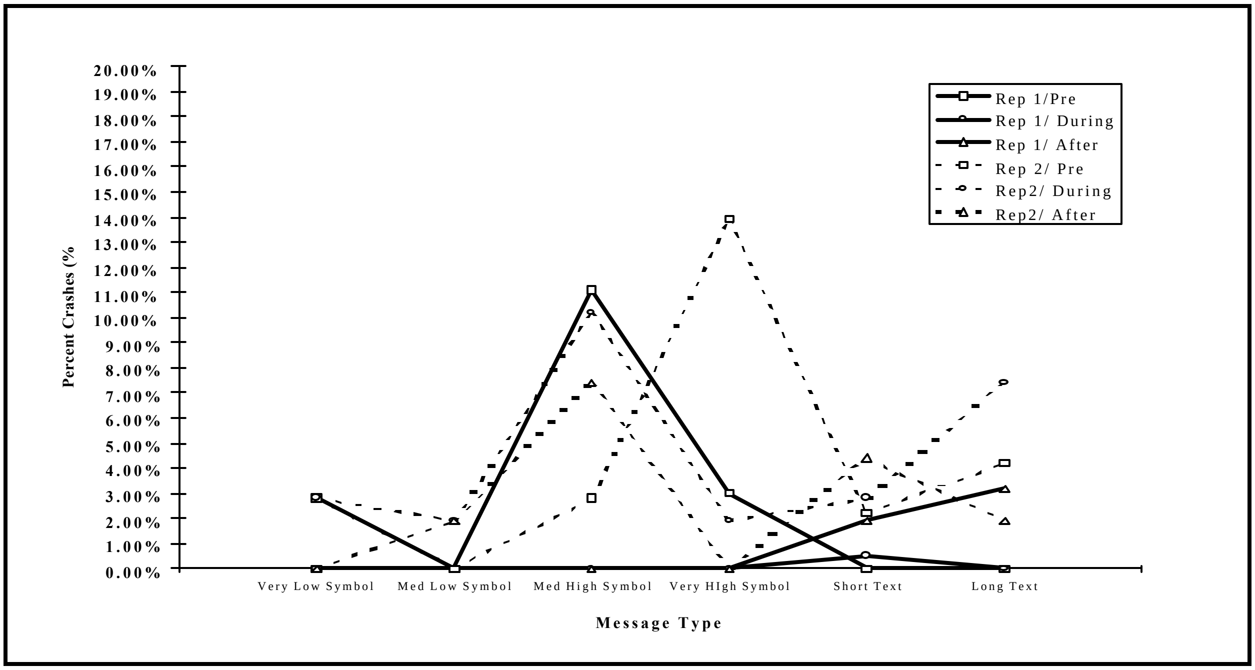


Figure 22. Percent crashes as a function of message type, repetition, and data collection interval.

Figure 22a shows percent crash occurrence as a function of the data collection interval. Differences between the three levels of data collection interval failed to reach statistical significance, $p>0.05$.

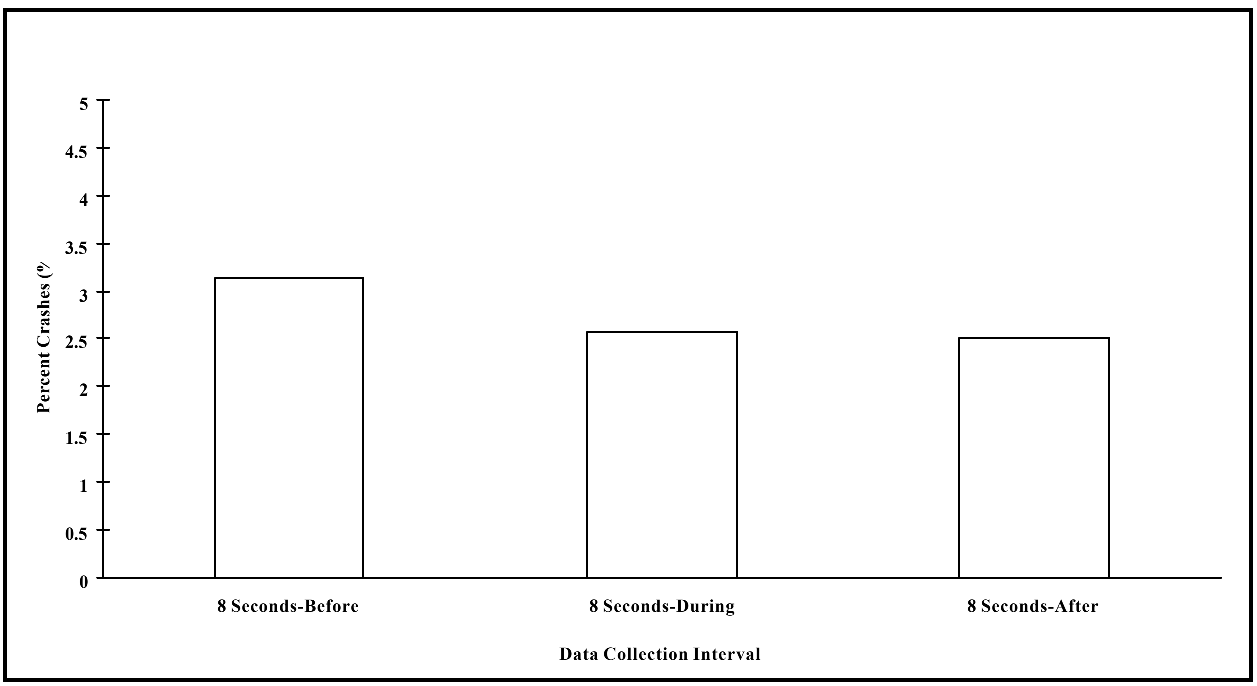


Figure 22a. Percent crashes as a function of data collection interval.

Pot-Test Phase
Two sets of data were collected in the Post-Test Phase. The purpose of these data sets was to assess subjects': (1) comfort with computers and (2) comprehension with a set of symbols.

Comfort With Computers

Recall that the "Experiment 5A, Post-Study Questionnaire: How Comfortable Are You With Computers?" was administered both before and after the simulation portion of the experiment. Subjects responded to the questions using a rating scale from 0 ("does not apply") to 100 ("strongly applies"). Table 6 outlines the mean results of each of the questions, broken up by driver age.

Table 6. Results of the pre-test and post-test "Experiment 5A, Post-Study Questionnaire: How Comfortable Are You With Computers?"

QUESTION	SCALED RATING (0 = DOES NOT APPLY, 100 = STRONGLY APPLIES)			
	YOUNGER DRIVERS		OLDER DRIVERS	
	PRE-TEST	POST-TEST	PRE-TEST	POST-TEST
1. I am sure I could do work with computers.	85.0	86.1	65.6	75.6
2. I would like working with computers.	73.9	73.3	68.9	65.0
3. I would feel comfortable working with computers.	81.7	77.2	65.6	68.3
4. Working with a computer would make me very nervous.	17.2	15.0	38.9	22.8
5. I do as little work with computers as possible.	16.7	15.6	41.1	32.2
6. I think using a computer would be very hard for me.	16.7	16.7	33.3	21.1

Each of the six questions was analyzed using a 2 (Age) x 2 (Gender) x 2 (Testing Time) repeated measures experimental design. Testing time refers to administration of the questions either: (1) pre-simulator or (2) post-simulator.

As illustrated in figure 23, the first question, "I am sure I could do work with computers," is shown as a function of driver age. Mean response for younger drivers was 85.6, while for older drivers it was 71.1. This difference was statistically significant, $F(1, 14) = 22.4$, $p<0.001$.

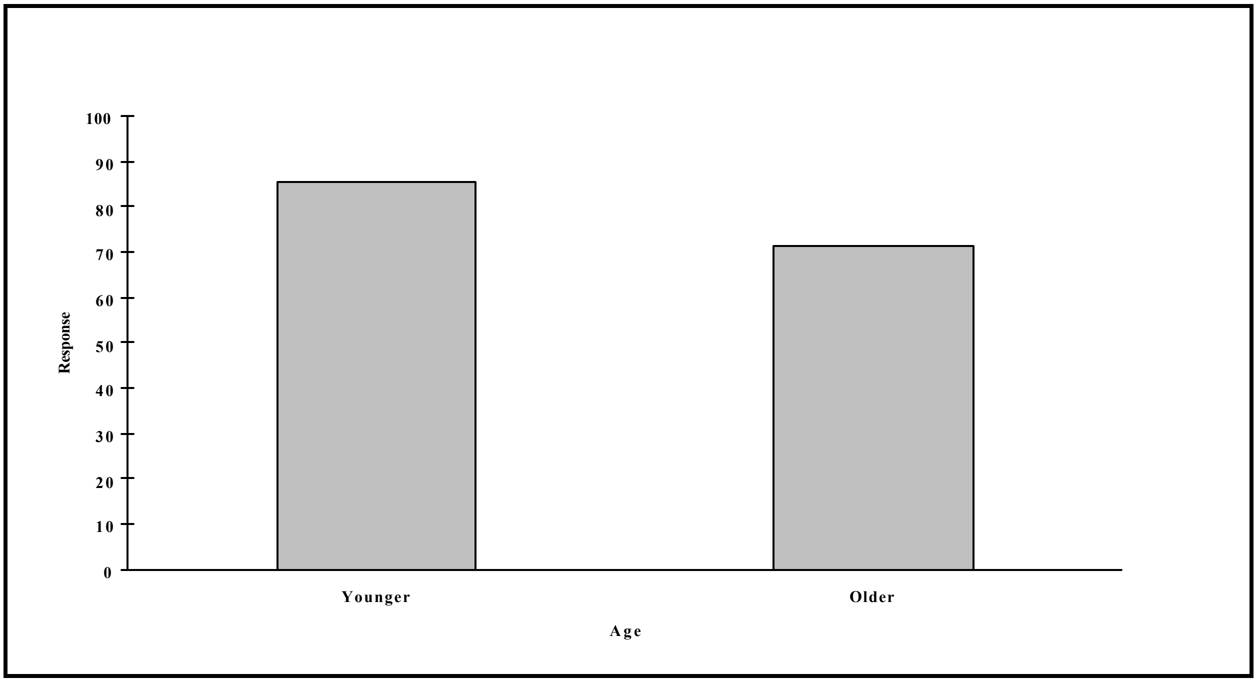


Figure 23. Comfort with computers pre- and post-test questionnaire, question 1.

Figure 24 shows question 2, "I would like working with computers," as a function of testing time (pre-test vs. post-test) and gender. Female drivers had larger responses for the post-test compared with the pre-test. However, male drivers had pre-test responses that were larger than the post-test responses. This Testing Time x Gender interaction was significant, $F(1, 14) = 5.33$, $p<0.04$.

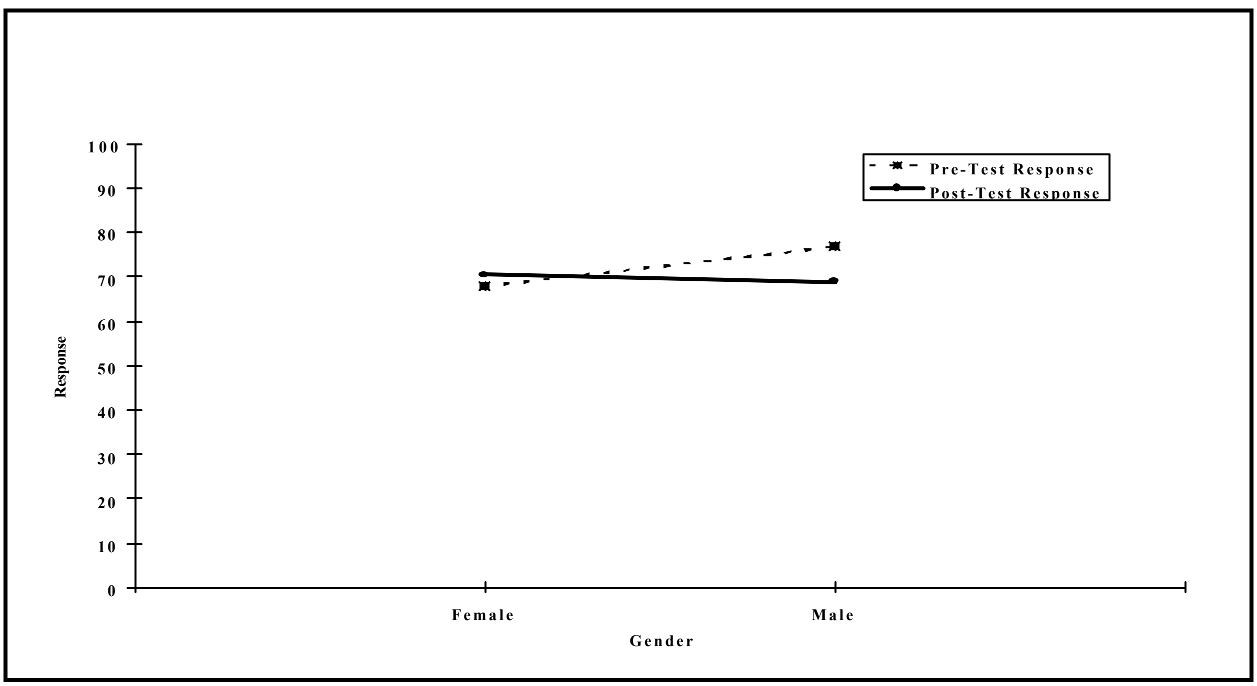


Figure 24. Comfort with computers pre- and post-test questionnaire, question 2.

Figure 25 shows question 3, "I would feel comfortable working with computers," as a function of driver age. Younger female drivers had higher response scores than did younger males. However, older males had higher response scores than younger males. This Age x Gender interaction was significant, $F(1, 14) = 5.09$, $p<0.05$.

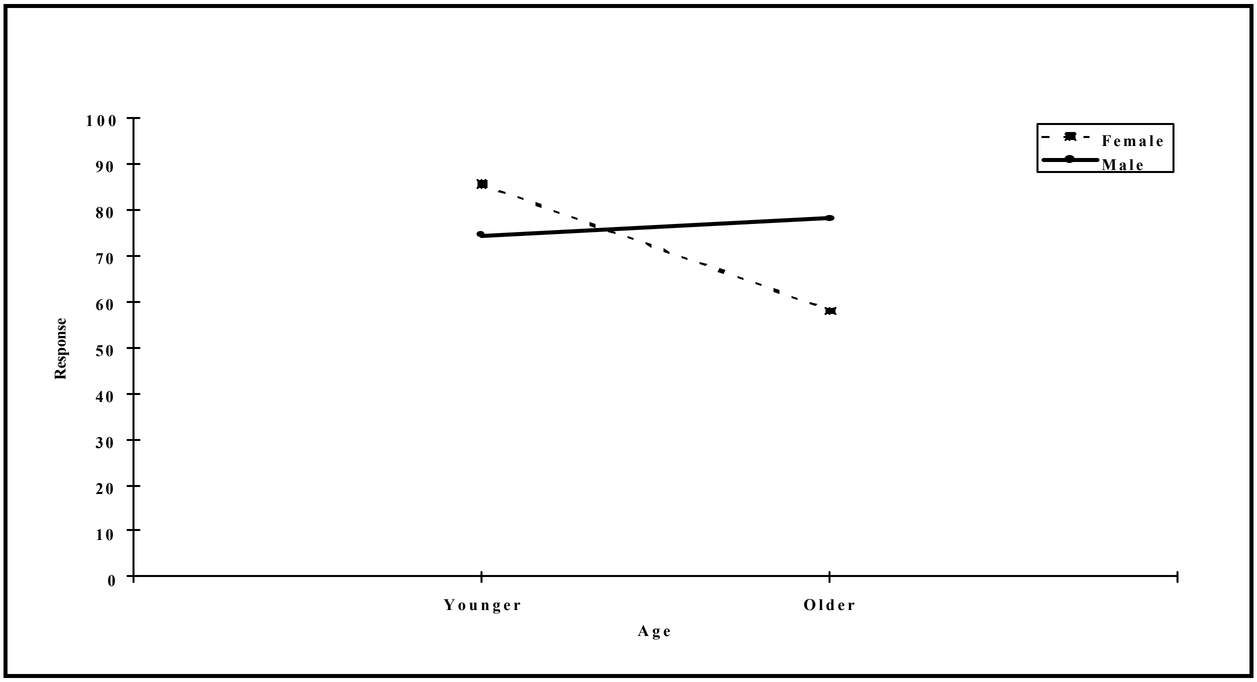


Figure 25. Comfort with computers pre- and post-test questionnaire, question 3.

Finally, figure 26 shows the main effect of age for question 4, "Working with computers would make me very nervous." Not surprisingly, older drivers rated this question significantly higher (M = 30.6) than did younger drivers (M = 16.1). The difference in responses to this question for younger and older drivers proved to be significant, $F(1, 14) = 8.11$, $p<0.02$.

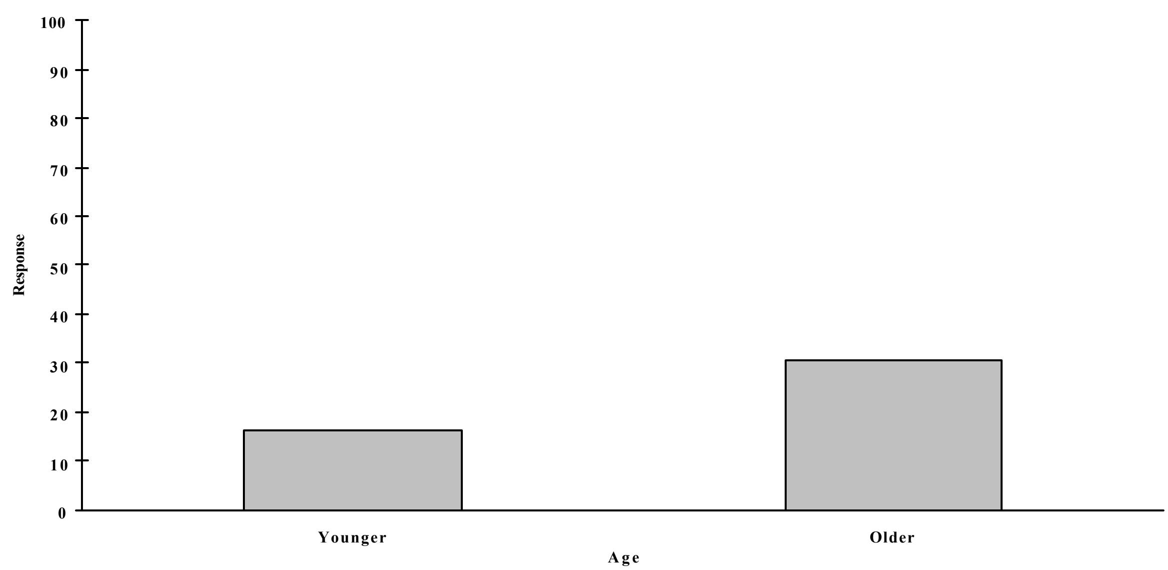


Figure 26. Comfort with computers pre- and post-test questionnaire, question 4.

Symbol Familiarity

Also during the Post-Test Phase, a test to assess subjects' comprehension with a set of symbols was administered. Recall from the *Procedures* section that during this test, drivers were shown 3 very high comprehension symbols, 2 medium high comprehension symbols, 1 medium low comprehension symbol, 1 very low comprehension symbol, and 13 previously untested comprehension symbols. All of the very high, medium high, medium low, and very low comprehension symbols shown in the Post-Test were previously presented during the simulator phase while three of the previously untested comprehension symbols were shown in the simulator phase. The 13 previously untested comprehension symbols, outlined in table 7, were selected as reasonable symbol possibilities that might appear on an IVIS. The reason that this category of symbols was labeled "previously untested" is that they were not previously rated on comprehension in either the Dewar et al. (1994) or Saunby et al. (1988) studies. Note also that all of the previously untested comprehension symbols are not listed in the 1988 published version of the MUTCD.

Table 7. Outline of previously untested comprehension symbols used in the post-test. (Each of the symbols can be found in appendix B).

PREVIOUSLY UNTESTED COMPREHENSION SYMBOL TITLE	PREVIOUSLY PRESENTED IN SIMULATOR PHASE
Ambulance Approaching	Yes
Car Crash Ahead	No
Car Fire Ahead	No
Check Speed	No
Congestion Ahead	No
Construction Equipment Ahead	No
Disabled Vehicle Ahead	No
Low Tire Pressure	Yes
Rain Ahead	No
Slow Trucks/ Steep Incline Ahead	No
Snow Ahead	No
Snow Plow Ahead	No
Speed Bumps Ahead	Yes

The procedure for this portion of the post-test can be summarized in the following four points: (1) each symbol was displayed for 8 s; (2) after all 20 symbols had been presented, in a different random order for each subject, each symbol was again presented; (3) the subject's task was to write down the name of the symbol on a response sheet (i.e., this was a *recall test*); and (4) when the subject had completed writing down a response, he/she touched the touch-screen and the next symbol was presented.

Responses were scored as either "correct" (score = 1) or "incorrect" (score = 0). Correct scores were judged as those where the wording of the response reflected the meaning of the symbol. For example, correct responses for the symbol "Check Speed" included: (1) "Check Speed," (2) "Check Your Speed," (3) "Watch Speed," and (4) "Watch Your Speed." Incorrect responses for "Check Speed" included: (1) "Speed," and (2) "Low Gas."

Three sets of analyses were conducted that examined the correctness and latency of the subjects' responses. The first analysis examined the set of symbols individually. The second analysis grouped the symbols by category (i.e., very high comprehension symbol, medium high comprehension symbol, medium low comprehension symbol, very low comprehension symbol, and previously untested comprehension symbol). The third analysis divided the previously untested comprehension symbol category into: (1) previously presented/ previously untested comprehension symbols and (2) not previously presented/ previously untested comprehension symbols. The "previous presentation" refers to presentation during the simulator phase.

The first analysis used a 2 (Age) x 20 (Symbol) experimental design. A Levene Test For Equality of Variances indicated significance across both the main effects of age and symbol type, and the Age x Symbol Group interaction (all *p*'s<0.001). This result indicated a lack of homogeneity of variance, and required a more conservative ANOVA. As such, a Brown-Forsythe ANOVA, where the variances are not assumed to be equal, was conducted.

Figure 27 shows that younger drivers had higher percent correct scores than did older drivers. Younger drivers achieved a 98.9 percent level while older drivers achieved a 90.0 percent level. This difference in percent correct scores as a function of driver age was significant, $F(1, 84) = 14.8$, $p<0.0003$.

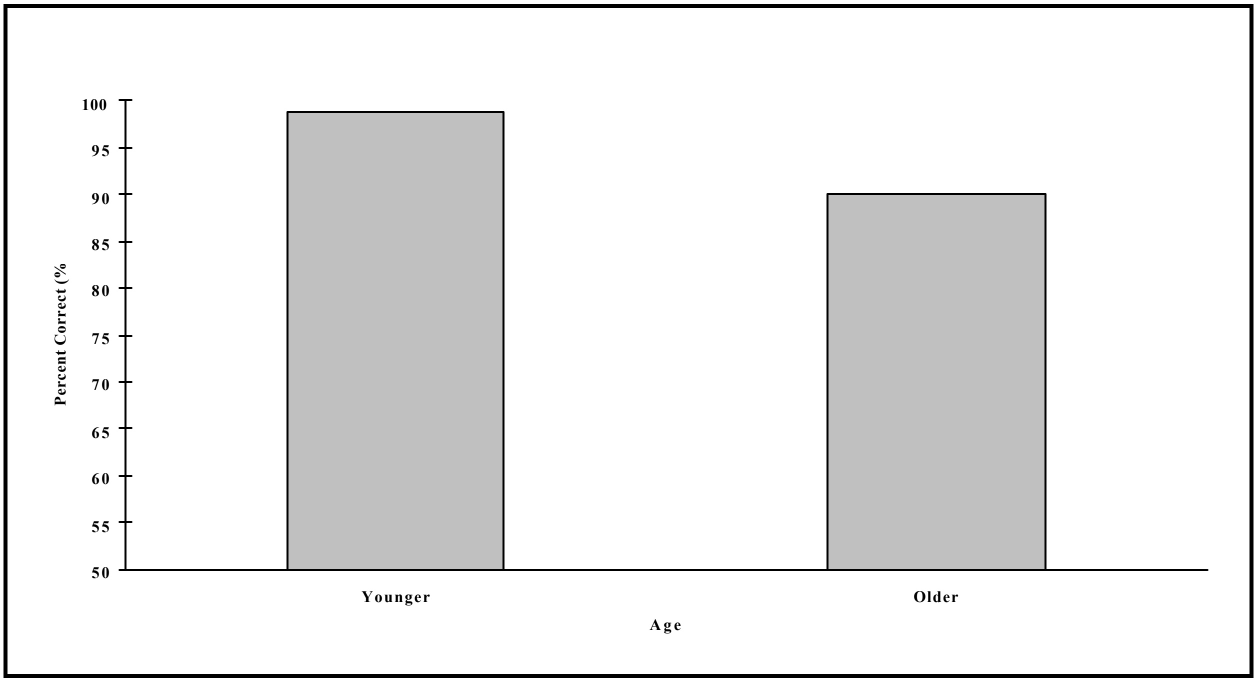


Figure 27. Percent correct on symbol comprehension post-test, and age.

Figure 28 shows percent correct for each of the symbol types. After accounting for heterogeneity of variance, the differences between the symbols did not prove to be statistically significant, $p>0.05$.

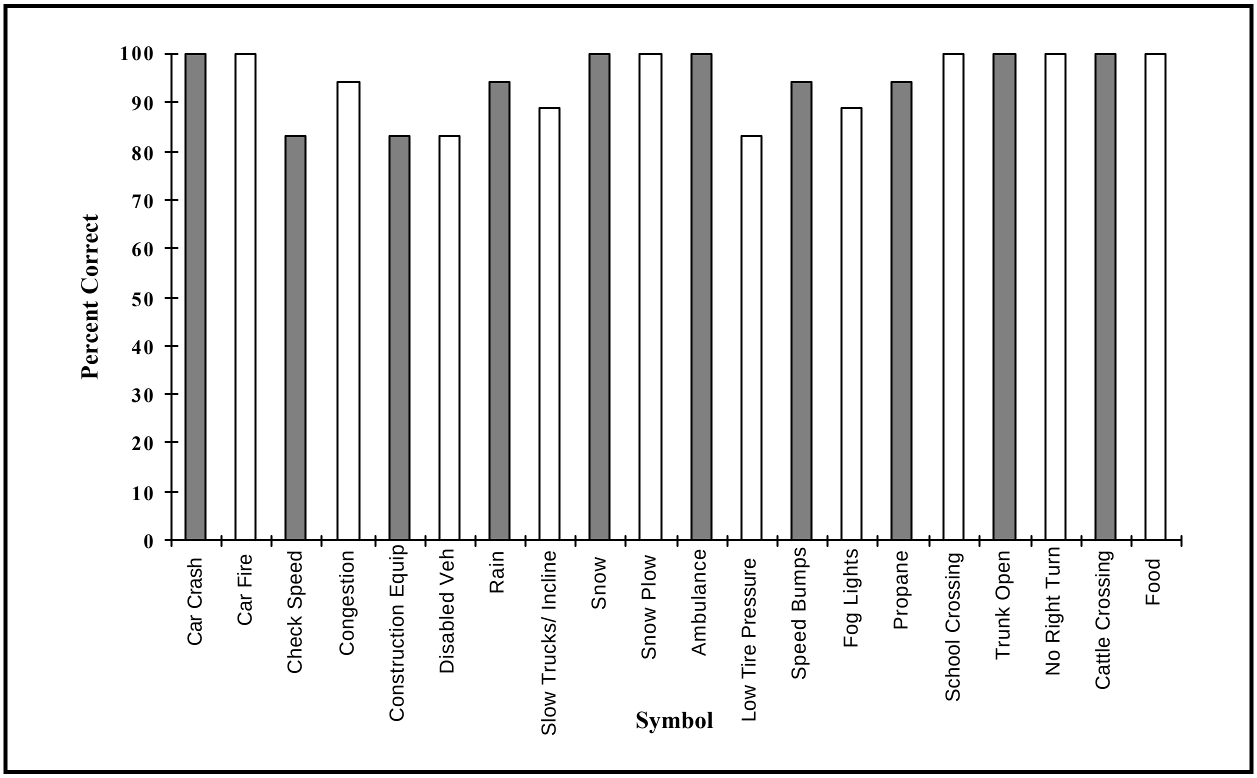


Figure 28. Percent correct on symbol comprehension post-test, and symbol type.

Using the same design model of 2 (Age) x 20 (Symbol), a second analysis was conducted that examined response latency as the dependent variable. Response latency refers to the time for drivers to examine a symbol, write down a response on the answer sheet, and touch the screen to move to the next symbol. The time began when the driver touched the screen, which presented a symbol. After providing a written response on the answer sheet provided, the drivers would again touch the screen to move to the next symbol. This second "touch" of the touch-screen ended that symbol's presentation and also provided a measure of response latency.

As in the previous analysis, and for all analyses conducted in the Post-Test phase, the Levene Test For Equality of Variances indicated heterogeneity of variance. As such, results presented are from the more conservative Brown-Forsythe ANOVA. For response latency, figure 29 shows that younger drivers were faster to respond than were older drivers, $F(1, 95) = 10.9$, $p<0.002$.

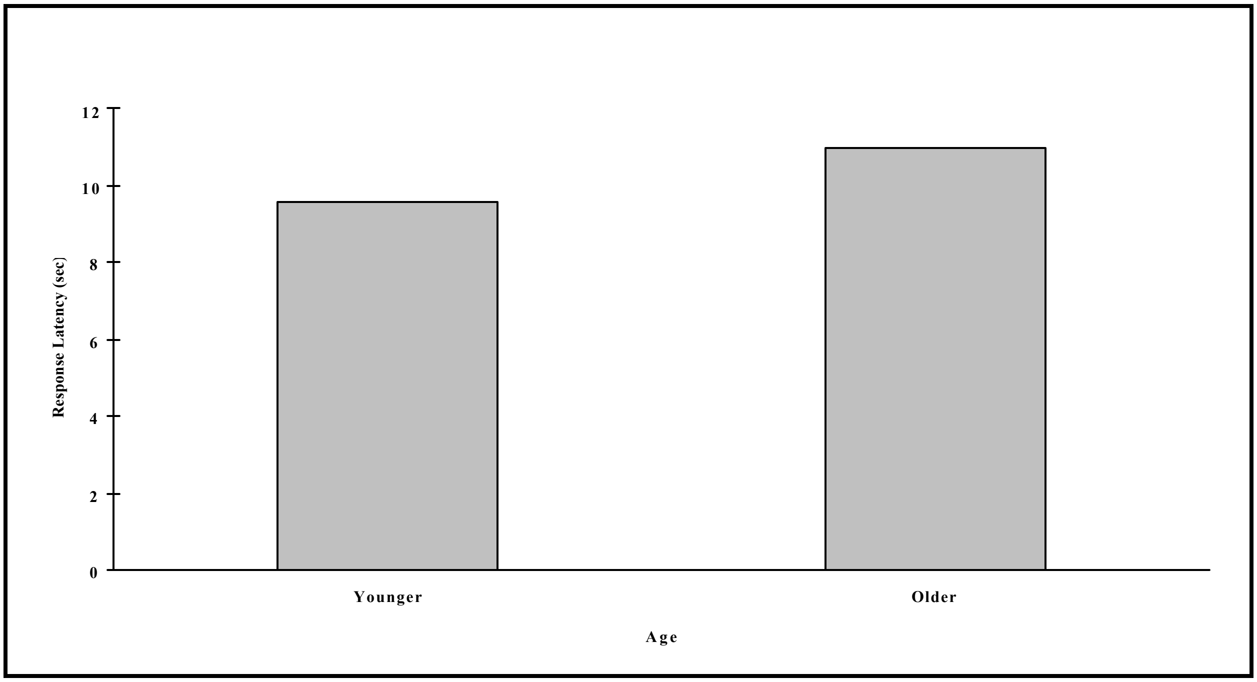


Figure 29. Response latency to symbol comprehension post-test and age.

Figure 30 shows response latency for each of the symbols. The differences between the symbols proved to be significant, $F(19, 95) = 5.07$, $p<0.0001$. A follow-up analysis examined the means in a Tukey post hoc test. These results indicated that two of the not previously presented/ previously untested comprehension symbols significantly differed from most of the other symbols. As can be seen in figure 30, the symbols that required the most response time were “Construction Equipment Ahead” and “Slow Trucks/ Steep Incline Ahead.” Other than these two symbols, no other significant Tukey results were found (all p’s>0.05).

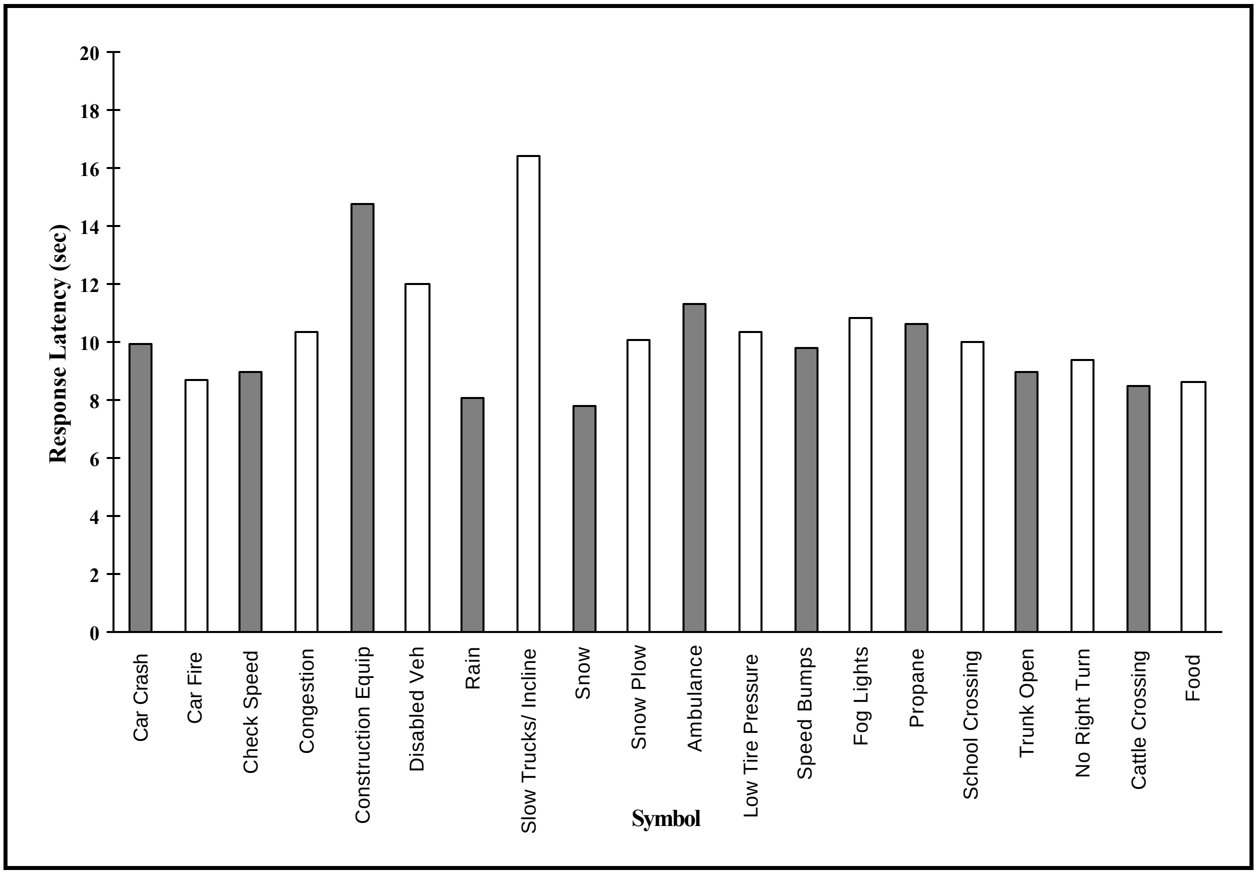


Figure 30. Response latency to symbol comprehension post-test, and symbol type.

The next analysis used a different design model in which the symbols were grouped into one of five categories: (1) very high comprehension symbol, (2) medium high comprehension symbol, (3) medium low comprehension symbol, (4) very low comprehension symbol, and (5) previously untested comprehension symbol. As such a 2 (Age) x 5 (Symbol Group) design was used. A Levene Test For Equality of Variances indicated heterogeneity of variance. As such, results presented are from the more conservative Brown-Forsythe ANOVA. The only result to reach statistical significance was age, $F(1, 16) = 5.9$, $p<0.03$. Grouping the symbols using the above mentioned categories did not yield significant results, $p>0.05$. Using response latency as the dependent measure also did not result in a significant effect for symbol group, $p>.05$. However, age was again found to be significant, $F(1, 17) = 5.03$, $p<0.04$.

The third analysis divided the previously untested comprehension symbols into two sub-groups: (1) previously presented/ previously untested comprehension symbols and (2) not previously presented/ previously untested comprehension symbols. For percent correct, age proved to be significant, $F(1, 21)$ 8.8, $p<0.008$, but symbol group did not, $p>0.05$. Using response latency, age was significant, $F(1, 19) = 6.45$, $p<0.03$, but symbol group was not significant, $p>0.05$.

CHAPTER 2. EXPERIMENT 5B

INTRODUCTION

This experiment was quite similar to the previous experiment. The main difference was that messages were presented auditorally. Auditory in-vehicle messages have been used to reduce the load on the driver's visual channel and there is evidence that, especially for alerting (Micheal and Casali, 1995) and warning functions (Graham, Hirst, and Carter, 1995), auditory messages or prompts can reduce driver response latency.

Two kinds of messages were investigated, as was the case for the previous experiment. Text messages consisted of recorded speech; this was analogous to the visual text messages previously used. Auditory icons, hereafter called earcons (Brewster, Wright, and Edwards, 1995) were analogous to the visual icons used previously. Two kinds of earcons were investigated. One set of earcons corresponded to natural sounds such as wind, tire squeal, and dripping water. Such naturalistic earcons offer greater ecological validity (Gaver, 1993) and have the potential to be processed more effectively by the driver. Complex tones, such as bells, whistles, and drum sounds, comprised the other category of earcons.

As in the previous experiment, three questions were examined:

1. How does in-vehicle message format (speech versus earcons) affect comprehension?

2. How does message format affect memory retention?

3. What impact does driver age have on memory of auditory messages?

METHOD

SUBJECTS

Eighteen subjects participated in this experiment. There were six males and six females ranging in age from 18 to 30 years old, and two males and four females over the age of 65. All subjects had a valid driver's license, drove at least two times per week, and reported not being prone to motion sickness. Younger drivers were recruited from the University of Washington. Older drivers were recruited from senior centers and recreational facilities. All subjects were paid $10.00 per hour, for approximately 3 hours of research time.

APPARATUS

The apparatus was the same as used in the previous experiment.

EXPERIMENTAL DESIGN

A 2×2×2×2×6×2×2 repeated measures design was used for the simulator portion of this experiment. Five within-subjects variables (message type, message repetition, message delay, message position, and question lure) were used with two between-subjects variables (age and gender). Table 8 below provides a complete description of all the independent variables used.

Table 8. Complete independent variable descriptions for experiment 5B.

INDEPENDENT VARIABLE	BETWEEN SUBJECTS/ WITHIN SUBJECTS	DESCRIPTION
AGE	Between Subjects	Young (18-30) Old (over 65)
GENDER	Between Subjects	Male Female
MESSAGE TYPE	Within Subjects	Earcon: - Naturalistic tone - Complex tone Speech
MESSAGE DELAY	Within Subjects	0-second delay 50-second delay
MESSAGE REPETITION	Within Subjects	Two repetitions at 0-second delay Two repetitions at 50-second delay
LURE TYPE	Within Subjects	Acoustic Cognitive-2 Cognitive-1 Acoustic and cognitive Control-2 Control-2
MESSAGE LOCATION	Within Subjects	Left side of vehicle Right side of vehicle

Independent Variables

The messages conveyed traffic, roadway, and vehicle information to drivers. Two message types were used, earcons and speech. Twenty-one earcons were developed using either naturalistic or complex tones to represent a message. Ten of the earcon messages were also presented as speech. All earcons were 300 ms in duration, approximately 65-72 dBA, and played through either the left or right channel, never in stereo. The speech messages were approximately 1s in duration, approximately 68 dBA, and played through either the left or right channel. A female voice was used to create the speech messages. Table 9 describes each message.

Table 9. Description of earcon and speech messages for experiment 5B.

MEANING	EARCON ONLY	NATURAL/ COMPLEX	DESCRIPTION
Ambulance Approaching		N	A siren that you would hear in a control room
Beach	**T**	N	Splash of water
Cattle Crossing	**T**	N	A cow “moo”
Fallen Rocks		C	Five pats on a congo drum randomly spaced
Ferry	**T**	C	A ferry horn
Gas Station		C	Two bell dings similar to the bells sounded when one drives up to the gas tanks
Impending Collision		N	A “crash” sound
Low Oil Pressure	**T**	C	Two dripping sounds
Low Tire Pressure	**T**	N	Hissing air
No Passing Zone	**T**	C	A doppler whoosh sound with a buzz sound played simultaneously
No Pedestrian Crossing		C	Two short whistles with a buzz sound played simultaneously
Passing Zone	**T**	N	A doppler whoosh sound
Pedestrian Crossing		N	One long whistle
Police Approaching		C	A British police siren
Slippery Road	**T**	N	Tire squeal
Snow Advisory		C	Sleigh bells
Speed Bumps Ahead		C	Two low congo drums pats played sequentially in an even rhythm
Telephone		C	Ringing bell
Train Station	**T**	C	Two clangs of bell with a ‘train sound’ in the background
Restaurant	**T**	N	Belch
Road Work Ahead	**T**	N	Jackhammer

Each message was repeated four times throughout the entire experiment. Each time a message was played, a question would appear on the touch-screen regarding the recent message along with two possible answers. Question delay refers to the lag time between the message offset and question onset. Questions either appeared immediately after the message (0-s delay) or 50 s after the message (50-s delay). The frequency of message repetition and message delay were completely balanced so that each message had two questions appearing with a 0-s delay and two questions appearing with a 50-s delay throughout the entire experiment. Message position refers to the side of the vehicle on which the message was played. Each message was played on the right side of the vehicle two times and the left side of the vehicle two times.

The questions were written so that specific pairings could be made between messages in order to determine whether observers were confusing similar messages (question lures). The types of questions that were investigated were as follows: (1) Acoustic lures compared messages that had similar acoustical properties (i.e., gas station consisted of two distinct bell dings and train station consisted of two similar sounding bell dings with the second bell being briefer than the first); (2) Cognitive-2 lures compared two messages that had similar goal characteristics and subjects had learned a traffic message for both possible answers (i.e., ambulance approaching and police approaching would both require the driver to locate the emergency vehicle and, if necessary, slow down and pull over to allow the emergency vehicle to pass); (3) Cognitive-1 lures compared messages that had similar goal characteristics; however, subjects learned a traffic message for one of the possible answers and the other answer was novel (i.e., the message "ambulance approaching" would be paired with "fire truck approaching." These two types of events would cause similar actions in drivers, however, the subject only learned an associated message for ambulance, not fire truck); (4) Acoustic and cognitive lures compared messages that possessed similar acoustical properties and also similar goal characteristics (i.e., "speed bumps" were two distinct congo drum beats and "fallen rocks" was a compilation of several congo drum beats. Both "speed bumps" and "fallen rocks" would cause the driver to slow down and possibly maneuver around the obstacles in the roadway); (5) Control-2 is a lure that randomly compared two sounds that did not sound similar or possess similar goal characteristics and subjects had learned an association for both possible answers (i.e., "ambulance approaching" and "speed bumps ahead" did not sound similar nor did they possess similar goals); and (6) Control-1 lures randomly compared one learned association with a novel stimulus (i.e., "ambulance approaching" and "camp ground ahead").

Given the complexity of the design and range of levels of independent variables, not all independent variables were completely balanced in respect to other variables. Each message was repeated four times across all three scenarios, twice with a 0-s delay and twice with a 50-s delay. Each message was played on the left side of the vehicle twice and on the right side of the vehicle twice. It was balanced over all messages as to how many times each message at a 0-s delay is on the left side of the vehicle versus the right, and how many times each message at a 50-s delay is played on the left side of the vehicle versus the right. Question lure was not balanced with respect to other variables since only similar acoustic and similar goal properties were investigated. The three scenarios did not possess the same number of messages, time delays, or message positions.

The entire experiment was 101 minutes in length. This was broken into scenarios one, two, and three, consisting of 32 minutes, 34 minutes, and 35 minutes, respectively. All subjects viewed the three scenarios in the same order.

Dependent Variables

The dependent variables that were used are outlined below in table 10. Data were collected in four primary phases: (1) Subject recruitment phase, (2) Pre-test message recognition phase, (3) Simulator testing phase, and (4) Post-test message recognition phase. The dependent measures that were collected during each of these four phases are detailed below.

Subject Recruitment Phase

There were two missions of the subject recruitment phase. The first was to determine subjects' suitability for this experiment. Subjects were required to have an active driver's license, drive at least twice per week, and report to rarely have difficulties with motion sickness.

After it had been determined that subjects were qualified to participate, some basic demographic information was collected (i.e., age, marital status, and number of miles driven annually).

Pre-Test Message Recall Phase

The pre-test message recognition phase consisted of teaching the subjects the meanings associated with each of the 21 earcons. The subjects were required to recall all 21 earcons on two consecutive tests. All subjects were allowed one incorrect answer on the second test.

Simulator Testing Phase

The testing phase consisted of data collection in the Battelle Human Performance Laboratory Low-Fidelity Vehicle Simulator. The following data were recorded.

Table 10. Description of dependent variables for experiment 5B.

DEPENDENT VARIABLE	DESCRIPTION
Recognition Accuracy	Measured in percent, accuracy of responses to in-vehicle system message recall questions.
Recognition Latency	Measured in milliseconds, time to respond to in-vehicle system message recall questions.
Self-Confidence Response	Measured on scale from 0 (low) to 100 (high), rated self-confidence in recognition accuracy task.
Self-Confidence Response Latency	Measured in milliseconds, time to response to self-confidence questions.
Message Position Accuracy	Measured in percent, accuracy of localizing which side of the vehicle the messages came from (right or left side).
Message Position Latency	Measured in milliseconds, time to response to localizing which side of the vehicle the message came from (left or right).

Post-Test Phase

At the conclusion of the simulator portion of the testing phase, one test of all earcon messages was given to the subjects. The purpose of the post-test was to determine if there was a change in subjects' ability to recall the earcon messages after the driving task.

PROCEDURES

The initial screening for participant suitability was done by telephone. The "Subject Selection Phone Questionnaire" and the "Driver Demographic Questionnaire" were administered at this time. The purpose of the screening procedure was to ensure a homogeneous population in terms of age groups, driving knowledge, and experience. Another goal of the screening was to rule out subjects who might be prone to motion sickness and have difficulty driving the simulator. Potential subjects who either did not have an active driver's license, drove less than twice per week, or reported experiencing motion sickness often were eliminated from the subject pool. Those who met the outlined criteria were administered a series of demographic characteristic questions and scheduled for a laboratory testing time. The demographic questionnaire was given during the telephone interview to reduce the time required of subjects for the testing session.

At the testing site, subjects were given a brief explanation of the experiment and then asked to fill out a written consent form. The experimenter then told subjects they would be listening to 21 sounds, all representing roadway or vehicle information. The experimenter told the subject the name of the sound they were about to hear (i.e., "this sound means ambulance approaching") and the sound was played. A brief explanation was given about each sound and why it represents that particular roadway or vehicle event. After all 21 sounds were played and explained to the subject, the list was played a second time. The second time through the entire list, the sound was played before the experimenter reported what the sound meant. Sounds often confused were then played for the subjects so that they could hear the discriminating properties. Finally, subjects were asked if there were any sounds that they wanted to hear again before taking the test. The subjects were given a sheet of paper with 21 blanks on it and asked to write down what each sound meant as it was played in sequence. They were allowed to have any sounds repeated or to come back to a particular sound. The experimenter then played each sound in a different sequence and waited for the subject to indicate they were ready to continue with the next sound. After all 21 earcons were played, the experimenter asked if the subject wanted to hear any repeated and, if not, the responses were evaluated to determine need for re-testing. Subjects were required to get 100 percent correct once and not get less than 95 percent correct on the following test. The subjects were required to repeat the test until they reached this criterion. If the subjects did not reach this criterion within 60 min, they were withdrawn from the study.

After reaching pre-test criterion, the experimenter read the simulator instructions, which gave a precise description of the task. The subjects were then given a 5-min-long practice scenario. The experimenter sat in the passenger seat to provide instruction about the touch-screen and vehicle operation, outline the task, and monitor and assist the subject in completing the task correctly. Subjects who correctly performed and completed the task during the practice scenario began the testing scenarios. Subjects who had difficulty completing the task (e.g., were not comfortable with the touch-screen operation) were administered a second 5-min-long practice session. If the subject continued to have problems, a final 5-min practice session was allowed. Subjects who could not perform the task after three practice sessions were withdrawn from the experiment.

After the practice session, each subject drove three simulated scenarios, lasting 32, 34, and 35 minutes, respectively. The simulator was programmed to maintain a constant speed and, therefore, in terms of operating the vehicle, subjects were only required to steer. The driver's task was as follows: (1) to safely operate the vehicle, (2) to listen for the traffic and traveler-related messages, (3) to press the right or left button on the steering wheel indicating which side of the vehicle the message was played, and (4) to respond to the questions that pertained to the auditory messages. Note that periodically, questions would appear on the touch-screen that pertained to roadway information only (i.e., “Was the traffic light you just passed green or amber?”). The purpose of these distraction questions was to help keep the driver focused on the driving events, rather than only on the messages. A total of 15 distractor questions were administered over the course of all three scenarios.

A tone was played to indicate when a question appeared on the touch-screen. After reading the question, drivers would select (touch) one of two response boxes (i.e., forced choice). For example, when the “Speed Bumps Ahead” message was played, the question would read, “What obstacle is ahead?” The response choices were “Speed Bumps Ahead” or “Fallen Rocks Ahead.” After answering the question, a follow-up question immediately appeared that queried the driver on his/her confidence of the previous response. A horizontal scale was presented on the bottom screen that ranged from 0 (very unsure) to 100 (very sure). By touching a point on the scale, drivers could indicate their confidence. Drivers were allowed 15 s to answer both questions, after which the screen blanked.

During the course of the simulated drive, the three types of messages (naturalistic sounds, complex tones, and verbal messages) would appear and, either immediately or 50 s later, drivers were asked the above noted forced-choice questions.

At the conclusion of the three scenarios, drivers would take a post-test for earcon recall. Again the subjects would be given a sheet of paper with 21 blanks and were asked to write down the meaning of the message as the experimenter played the sounds in random sequence. The subject would listen to each earcon and write down its meaning, then indicate when he/she was ready to hear the next message. After all 21 nonverbal messages were played, the subject was asked a series of questions verbally by the experimenter, and was asked to rank likes/dislikes of general properties of the verbal and nonverbal messages. Once this was completed, the subjects were paid for their time.

RESULTS

PRE-TEST MESSAGE RECALL

For this test, the 21 earcons were played and subjects had to name each sound immediately after it was presented. This test was repeated until subjects were able to recall each sound with not more than 1 error in the entire set of 21 earcons. For pre-tests one, two, and three (for those subjects who required a third test), the mean recall scores for younger drivers were 98.8 percent, 99.6 percent, and 100 percent. All younger drivers completed this phase of the experiment in less than 25 minutes.

For older drivers, pre-test scores for four tests were 72 percent, 80 percent, 87 percent, and 88 percent. After 60 minutes of training, only 3 older drivers reached the fourth pre-test. No older driver was able to meet the criterion of only one error on this test. This failure, according to self-reports from all the older drivers, was due to their inability to distinguish among the 300 ms sounds, rather than an inability to remember the names of the sounds. Since no older driver met the pre-test criterion, none proceeded to the driving phase of the experiment. The following reported results pertain to younger drivers only. A sample of the results of the "Driver Demographic Questionnaire" administered to these younger drivers is presented in table 11. The relevant ANOVA tables for experiment 5B can be found in appendix C.

Table 11. Sample of the primary results from "Driver Demographic Questionnaire" (as administered to younger drivers for experiment 5B).

DRIVER GENDER GROUP	DEMOGRAPHIC		DRIVING EXPERIENCE		
	AGE	YEARS LIVING IN SEATTLE	YEARS AS A LICENSED DRIVER	YEARS DRIVING IN SEATTLE	AVERAGE MILES DRIVEN ANNUALLY
MALES	20.3	4.6	3.5	1.4	11250
FEMALES	23.4	14	8.7	5.7	16667

IN-VEHICLE MESSAGES

Figure 31 shows that recognition accuracy was extremely high for all conditions. Effects of message delay, message type, and their interaction did not reach the 0.05 level of significance.

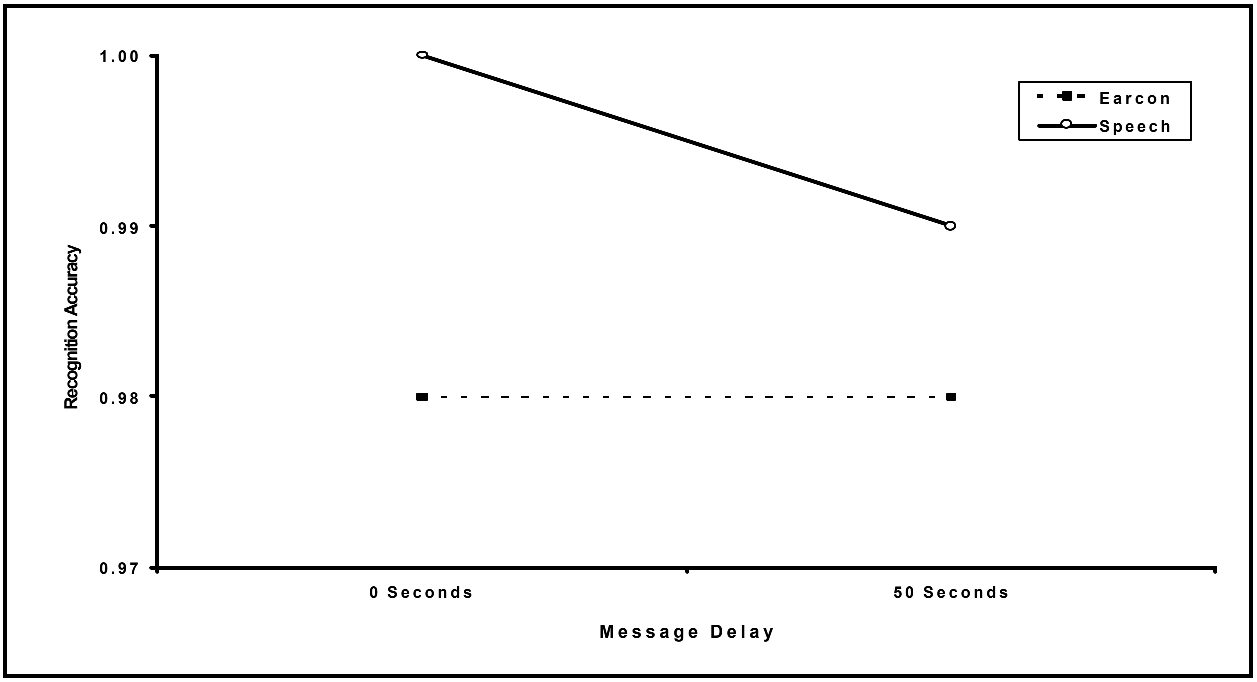


Figure 31. Recognition accuracy: Message Type x Message Delay.

Figure 32 shows the effects of message type and delay upon recognition latency. Drivers took longer to respond to earcons (M = 2.45 s) than to speech messages (M = 2.16 s), $F(1, 10) = 14.1$, $p < 0.01$. Drivers responded faster with zero message delay (M = 1.85 s) than with a 50-s delay (M = 2.77 s), $F(1,10) = 121$, $p < 0.001$.

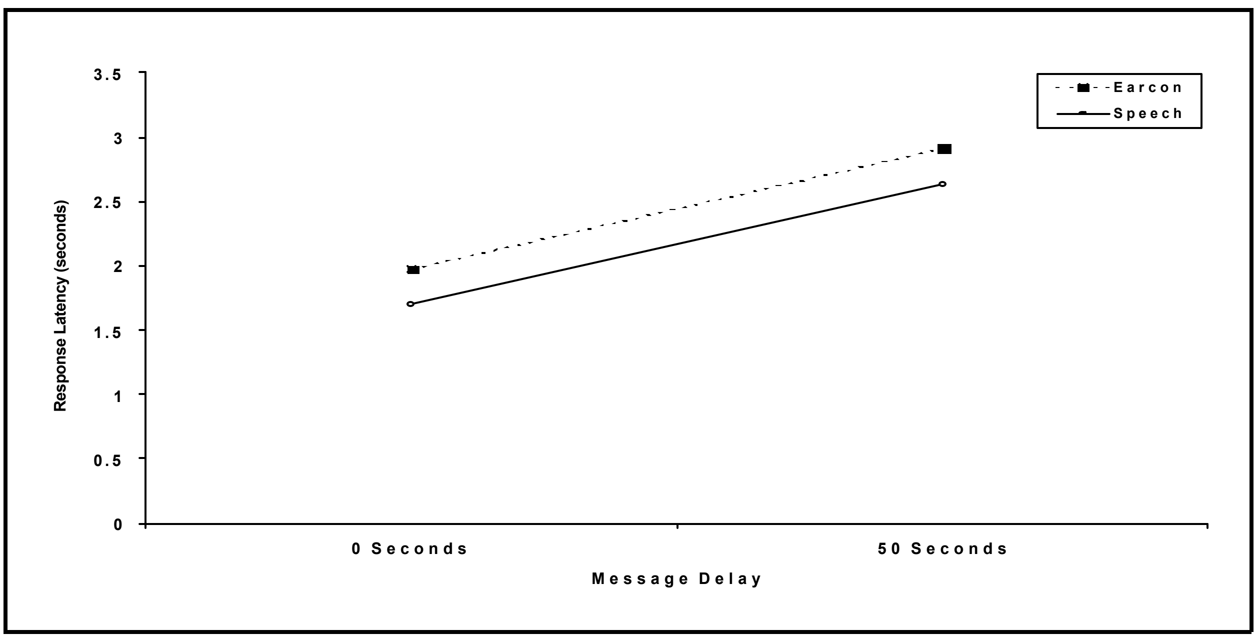


Figure 32. Recognition latency: Message Type × Message Delay.

Figure 33 shows effects of message type and delay upon subjective confidence ratings. As was the case for recognition accuracy, confidence was very high and no effects were significant at the 0.05 level.

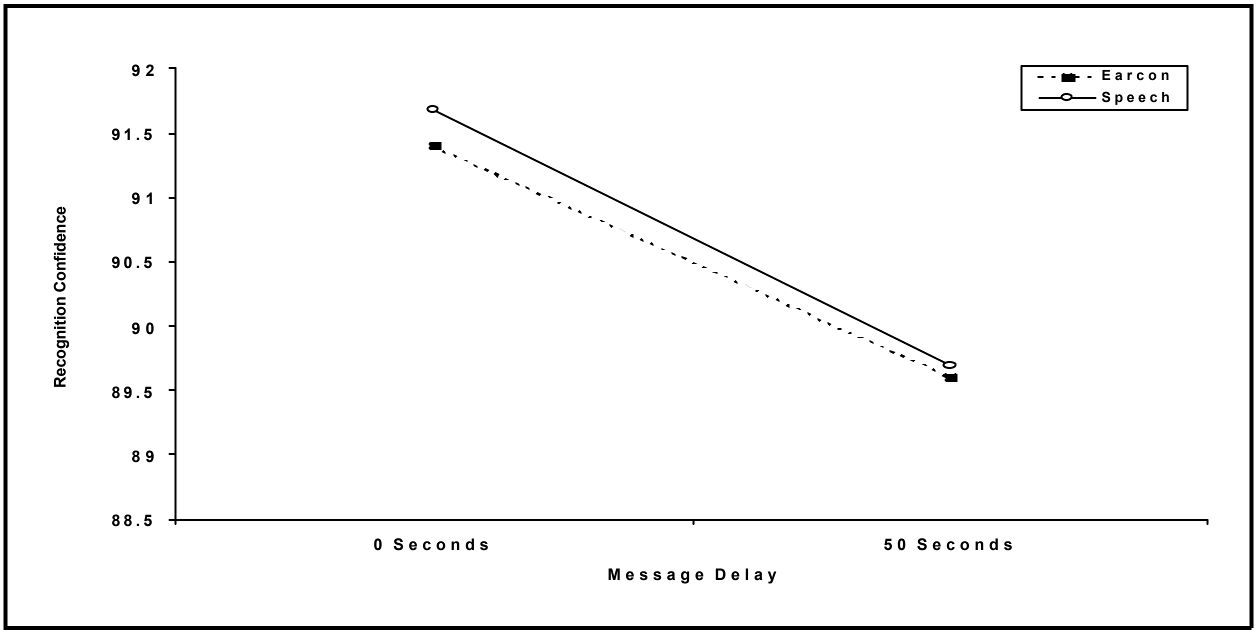


Figure 33. Recognition confidence: Message Type × Message Delay.

Figure 34 shows effects of message position and repetition upon localization accuracy, where drivers were required to press either the left or right button on the steering wheel to indicate on which side the message occurred. While messages presented on either side were localized identically for the first repetition of a message, the second repetition revealed better performance for the left-side messages, $F(1, 10) = 7.85$, $p = 0.02$. This interaction was the only statistically significant finding. The driver, of course, was closer to the left-side loudspeaker. This result may indicate that, as drivers learn about in-vehicle systems that present auditory messages, they also learn about the physical position presenting the message. However, an experiment with many more message repetitions would be required before this result could be incorporated into an ATIS guideline.

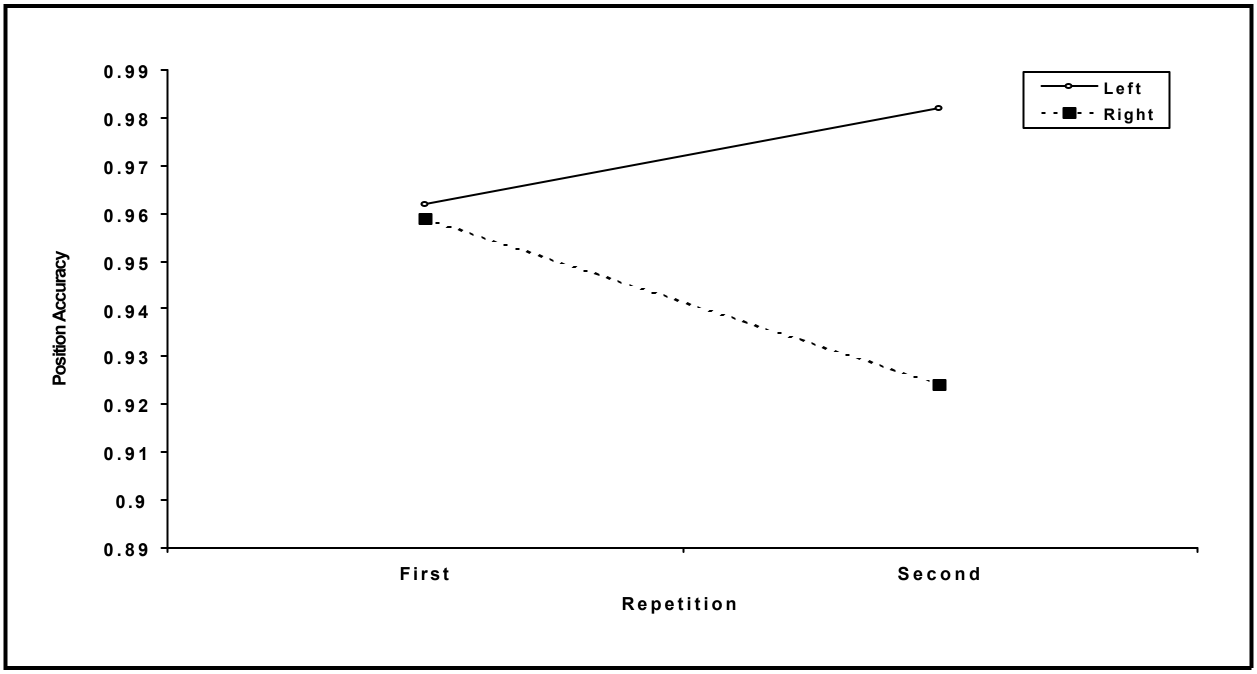


Figure 34. Position accuracy: Message Position × Repetition.

Figure 35 shows latency of the localization response, pressing the left or right button on the steering wheel. Drivers localized earcons more quickly than speech messages, $F(1, 10) = 6.63$, $p < 0.03$. There was no significant effect of message position upon latency. While an ecological psychologist might interpret this result to indicate that there is a survival benefit to the species of being better able to localize naturalistic sounds, it is not immediately apparent how this result could be incorporated into ATIS guidelines. However, this result could be useful for crash avoidance guidelines because it suggests that earcons are better for indicating the position of an external vehicle, such as an ambulance. Given that it is difficult for drivers to localize external vehicle position (Caelli and Porter, 1980), such in-vehicle information could be very helpful.

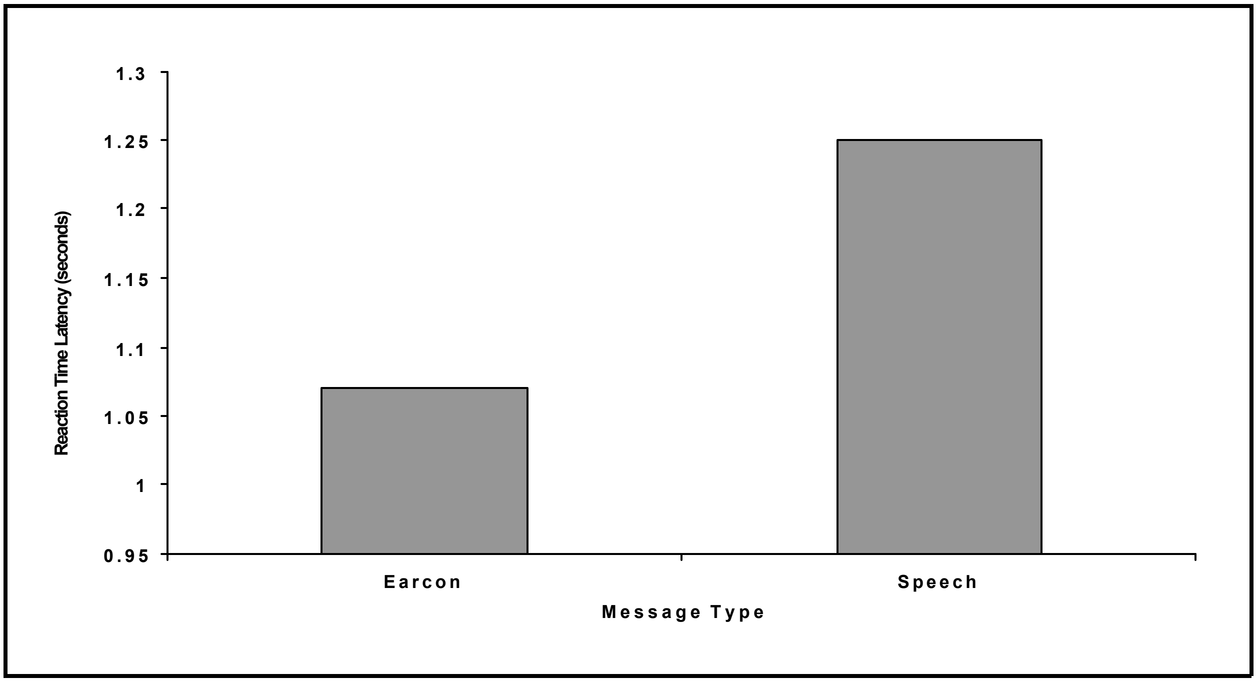


Figure 35. Position latency: Message Type.

Figure 36 shows that recognition accuracy was very high for all messages. No one message was any more accurate, and no differences reach the 0.05 level of significance.

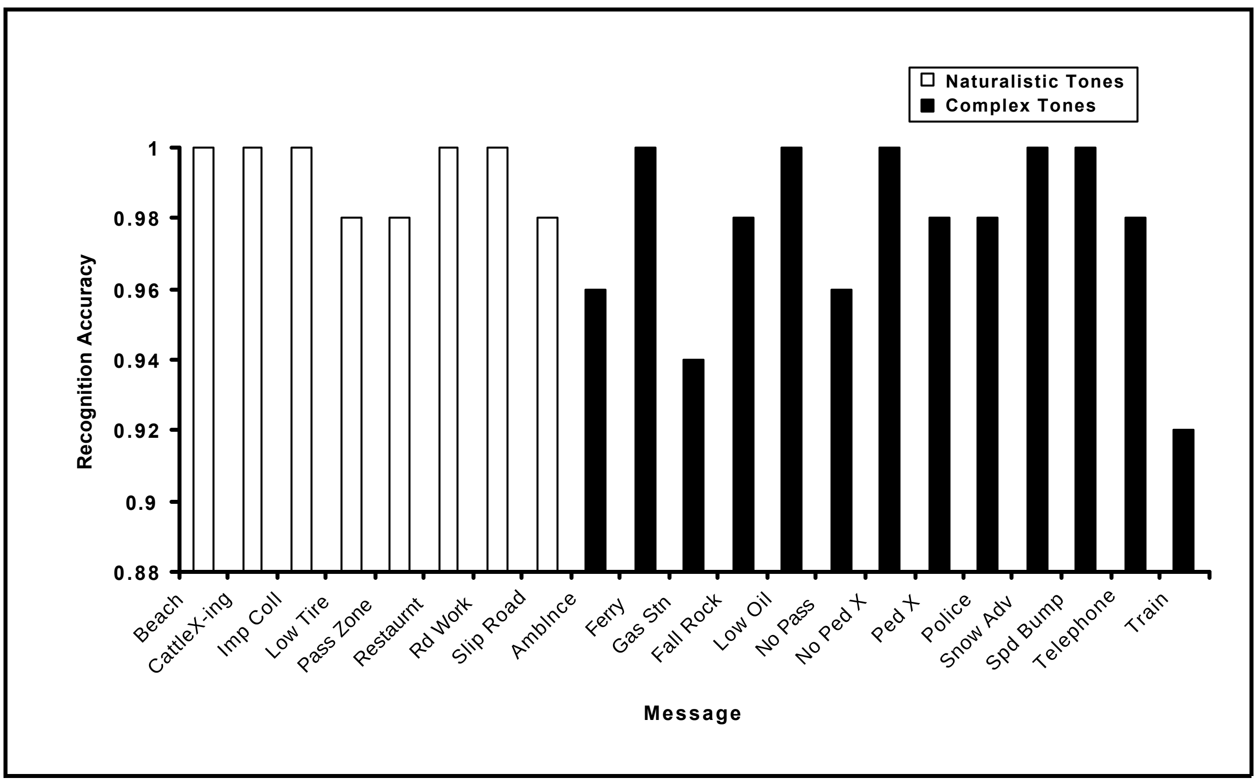


Figure 36. Recognition accuracy: Message.

Table 12 shows the results of a positive Tukey Test where each message was compared with all other messages. Only the messages where significant differences in recognition latency occurred are shown in the table.

Table 12. Recognition latency: Message.

	NO PEDESTRIAN CROSSING	PASSING ZONE	PEDESTRIAN CROSSING	RESTAURANT	SLIPPERY ROAD	SNOW ADVISORY	SPEED BUMPS AHEAD	TELEPHONE	TRAIN STATION
Beach	0.883	--	0.885	--	--	0.749	--	0.774	1.272
Cattle Crossing	--	--	--	--	--	--	--	--	0.938
Ferry Dock	--	--	--	--	--	--	--	--	1.069
Gas Station	--	--	--	--	--	--	--	--	0.98
Impending Collision	--	--	--	--	--	--	--	--	0.827
Fallen Rocks	--	--	--	--	--	--	--	--	1.093
Low Tire Pressure	0.885	--	0.887	--	--	0.751	--	0.776	1.274
Low Oil Pressure	--	--	--	--	--	--	--	--	0.974
No Passing Zone	--	--	--	--	--	--	--	--	1.068
No Pedestrian Crossing	--	--	--	0.737	--	--	0.791	--	--
Passing Zone	--	--	--	--	--	--	--	--	1.085
Pedestrian Crossing	--	--	--	0.739	--	--	0.793	--	--
Police Approaching	--	--	--	--	--	--	--	--	0.922
Restaurant	--	--	--	--	--	--	--	--	1.126
Road Work	--	--	--	--	--	--	--	--	1.008
Speed Bumps Ahead	--	--	--	--	--	--	--	--	1.18
Critical Value: 0.7328									

Figure 37 groups latencies for naturalistic versus complex earcons. This a priori comparison revealed faster latencies for naturalistic tones, t (10) = 8.83, p = 0.014.

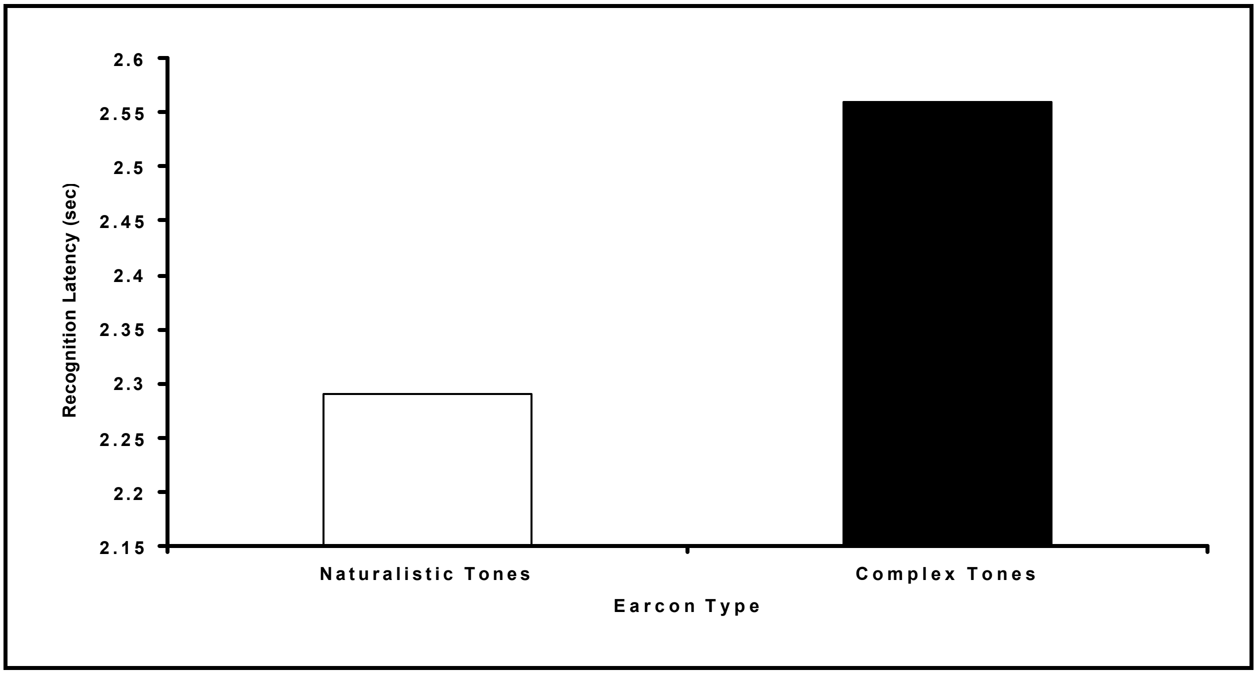


Figure 37. Recognition latency: T-test comparing naturalistic and complex tones.

POST-TEST MESSAGE RECALL

At the close of the experiment, the set of 21 messages was again administered when subjects were not driving. Mean scores were 100 percent for males and 98 percent for females. This test was included as a control condition to evaluate possible changes in message recall after driving. However, since accuracy scores were so high in all conditions, no improvement was observed.

CHAPTER 3. EXPERIMENT 5C

INTRODUCTION

In the first two experiments, memory for visual and auditory messages was evaluated while the vehicle speed in the driving simulator was controlled automatically. Cruise control was activated in these experiments because our initial goal was to study memory rather than driving, and so we tried to minimize driver workload. (Indeed, our original research plan called for these experiments to be performed without a driving simulator, but we were able to take advantage of early completion of the simulator facility to perform these experiments in the simulator instead of a test booth.) Results showed that memory was excellent when drivers were responsible only for controlling steering. In this experiment, the driver was in full control of the vehicle, being responsible for both vehicle speed and lane position. Furthermore, unlike the simpler, previous experiments where messages were presented only in a single modality, this experiment used a mixed presentation with both visual and auditory messages. Given the results of the previous experiment, where older drivers were unable to learn the auditory messages, only younger drivers were tested in this experiment. A new variable, alert, was added to warn drivers that a message would be presented. Since auditory messages have an innate alerting quality (e.g., the driver can hear the message even if looking out the windshield to the roadway), a comparison of mixed auditory and visual messages may be confounded by potential alerting qualities that differ across modalities. Thus, two kinds of auditory alerts (speech and tone) were examined; furthermore, a no-alert control condition was also included to provide baseline data.

Two questions, also asked for the previous experiments, were examined:

1. How does in-vehicle message format (modality, message format) affect comprehension?

2. How does message format affect memory retention?

METHOD

PARTICIPANTS

Thirty-six participants were recruited from the University of Washington. Participants met all selection requirements; possessed a valid driver's license, drove at least two times per week, were between 18 and 30 years old, and reported not to be prone to motion sickness. The initial screening for participant suitability was conducted via the telephone using the "Subject Selection Phone Questionnaire." If the participant passed selection criterion, the "Driver Demographic Questionnaire" was administered.

The final participant sample consisted of 18 males and 18 females, with a mean age of 23.3 years. Participants were paid $10.00 per hour for approximately 2 hours of simulated driving time.

APPARATUS

The apparatus was the same as used in the previous experiment.

EXPERIMENTAL DESIGN

Two between-subject variables (alert type and gender) and three within-subject variables (message modality, message format, and time delay) were examined in this experiment using a 3x2x3x2x2 mixed design (table 13).

Table 13. Independent variables for experiment 5C.

INDEPENDENT VARIABLE	BETWEEN-SUBJECTS/ WITHIN-SUBJECTS	DESCRIPTION
GENDER	Between Subjects	Female Male
ALERT TYPE	Between Subjects	No Alert Speech Alert Tone Alert
MESSAGE MODALITY	Within Subjects	Visual Auditory Visual + Auditory
MESSAGE FORMAT	Within Subjects	Symbolic Lexical
TIME DELAY	Within Subjects	0 Seconds 50 Seconds

Given the large number of independent variables and the complexity of the design, a completely counterbalanced design was not implemented. Message modality was ordered by creating a 3x3 Latin square containing three unique orders of message modality. This 3x3 Latin square was then repeated 4 times to account for the 12 information and traveler-related messages. Finally, these four Latin squares were repeated once more to account for the two levels of delay, for a total of eight Latin squares (table 14).

Table 14. Experimental design of ATIS: Experiment 5C

		FIRST SCENARIO			SECOND SCENARIO		
		0-SECOND DELAY			50-SECOND DELAY		
	MESSAGE	VISUAL	AUDITORY	VISUAL + AUDITORY	VISUAL	AUDITORY	VISUAL + AUDITORY
1	Fallen Rocks	1	2	3	1	2	3
2	Ambulance Approaching	2	3	1	2	3	1
3	Ferry	3	1	2	3	1	2
4	Cattle Crossing	1	2	3	1	2	3
5	Low Tire Pressure	2	3	1	2	3	1
6	Water Recreation	3	1	2	3	1	2
		50-SECOND DELAY			0-SECOND DELAY		
		VISUAL	AUDITORY	VISUAL + AUDITORY	VISUAL	AUDITORY	VISUAL + AUDITORY
7	Police Approaching	1	2	3	1	2	3
8	Speed Bumps Ahead	2	3	1	2	3	1
9	Snow Advisory	3	1	2	3	1	2
10	Pedestrian Crossing	1	2	3	1	2	3
11	Low Oil Pressure	2	3	1	2	3	1
12	Train Station	3	1	2	3	1	2

Between-Subjects Variables

The alert variable was used to determine if notification of an impending message would enhance drivers' recognition memory. One group received a 60 dbA tone alert that was 200 ms in length and consisted of two rapid pulses. The second group of participants received a 60 dbA speech alert that was 1000 ms in length and consisted of a recorded male voice stating, "incoming message." The third group of participants did not receive a message alert. Six males and six females were randomly assigned to each alert group.

Within-Subjects Variables

Three different levels of message modality were used to determine if modality (visual or auditory) would affect drivers' memory of messages and also if a bimodal (visual + auditory) message type would improve drivers' comprehension and memory over a unimodal (visual or auditory) message type. Visual messages were presented on a color monitor on the dashboard compartment of the vehicle. Auditory messages were presented through two speakers in the rear corners of the vehicle. Bimodal (visual + auditory) messages consisted of the simultaneous presentation of both a visual message and an auditory message.

Two levels of message format, symbolic and lexical, were used to determine if recognition memory is affected by the format of messages. Two types of format were investigated for both visual and auditory messages. Visual messages were presented as icons or text. Auditory messages were presented as either earcons or simple speech messages. Visual icons were rectangular in shape (7.25 cm long and 5.5 cm high) and were presented for 7 s in the center of a color monitor. Visual text messages were written in 0.5-cm capital letters using the same dimensions, duration, and location as the visual icons. The auditory earcons were presented at 57.5 dbA (+/- 2.5 dbA) for 300 ms. The auditory speech messages were presented at the same intensity for 1000 ms.

Both a 0-s time delay and a 50-s time delay condition were used in this experiment to determine how well participants would be able to remember messages over time. Time delay is defined as the length of time between the offset of the message and the onset of the question about that message.

Dependent Variables

Several measures of message comprehension, self-confidence, and driving performance were collected. These measures are summarized in table 15.

Table 15. Description of dependent variables for experiment 5C.

DEPENDENT VARIABLE	DESCRIPTION
RECOGNITION ACCURACY	Accuracy of responses to message recognition questions, measured in percent correct
RECOGNITION LATENCY	Time to respond to message recognition questions, measured in milliseconds
CONFIDENCE RATING	Rated self-confidence in recognition accuracy task, measured on a scale from 0 (low) to 100 (high)
LATENCY OF CONFIDENCE RATING	Time to respond to self-confidence questions, measured in milliseconds
MEAN LANE POSITION	The driver's lane position with respect to the centerline of the highway, measured in meters
STANDARD DEVIATION OF LANE POSITION	The driver's variable lane position with respect to the centerline of the highway, measured in meters
MEAN VELOCITY	To determine whether drivers maintained an average velocity of 35 mi/h, measured in miles per hour
STANDARD DEVIATION OF VELOCITY	The driver's variable velocity around their target velocity of 35 mi/h, measured in miles per hour
NUMBER OF TICKETS	The number of times the driver exceeds 38 mi/h
NUMBER OF ACCIDENTS	The number of times that the participants collide with an object or drive off the road

PROCEDURES

Task

At the testing site, participants were given a brief explanation of the experiment and then were asked to read and sign a consent form. Each participant was trained on 10 earcons as well as 7 icons. All participants were administered earcon recall tests and icon recall tests until they achieved 100 percent correct on one test and no more than one incorrect answer on the very next recall test. If a participant still had not achieved this criterion after 60 minutes of training and testing, they were dismissed from the experiment.

The experimenter then read the simulator instructions describing the task in detail. Participants were asked to drive a simulated vehicle on a winding, rural, two-lane highway. While they were driving, visual, auditory, or visual plus auditory messages were presented to them. Either immediately or 50 s after the message, a question was displayed querying the driver about the previous message. After reading the question, drivers selected one of two response boxes. For example, if the Speed Bumps Ahead message was presented, the question, "Caution due to?", and response choices, Speed Bumps or Road Construction, would appear on the touch-screen. After selecting an answer, a follow-up question immediately appeared on the touch-screen querying the driver to rate the level of confidence she had in her previous answer. A horizontal scale was presented on the bottom of the touch-screen ranging from 0 (Very Unsure) to 100 (Very Sure). Drivers were allowed 15 s to answer both questions. If the driver failed to answer the first question in 15 s, the confidence rating scale did not appear.

One-third of the subjects received a tone alert prior to the message presentation, one-third received a speech alert prior to message presentation, and the final one-third received no message alert at all. All participants heard a bell chime indicating when a question was present on the question display.

After the task was described by the experimenter, the participants were allowed two 5-min practice scenarios. The first practice scenario allowed the participant to experience the vehicle dynamics, where participants controlled the vehicle's speed and direction (a velocity limit was set at 40 mi/h). The second practice scenario required the participant to practice the experimental task in its entirety, which included controlling the vehicle direction and speed, attending to the traveler- and vehicle-related messages, answering questions about these messages, and rating their self-confidence.

After the 2 practice sessions, each subject drove 2 simulated scenarios, lasting 25 min each. The participant was instructed to:

- drive safely and maintain 35 mi/h,
- listen and view the traffic- and traveler-related messages,
- answer the questions about the messages, and
- indicate their confidence about the previous question.

During these two scenarios the experimenter was in an adjoining room, monitoring the participant via a small in-vehicle camera to determine if there were any difficulties during the course of the experiment.

At the conclusion of the two scenarios, drivers were administered both auditory and visual message recall tests that were identical in format to the criterion tests administered at the beginning of the testing session.

Scenarios

During both scenarios, 36 messages were presented to the participant. Of these 36 messages, 12 messages were visual only, 12 messages were auditory only, and 12 messages were visual plus auditory. Also, 18 of the 36 questions in each scenario were presented immediately after the message presentation and 18 questions were presented 50 s after the message presentation. All participants viewed the scenarios in the same order.

The content of the 12 visual, 12 auditory, and 12 visual plus auditory messages was identical and pertained to general traveler or vehicle information (see appendix B). The 12 visual messages were selected from a subset of icons and text messages that were easily recognized in a previous experiment (experiment 5A). The 12 auditory messages were selected from a subset of earcons and speech messages that were easily recognized in a previous experiment (experiment 5B). The final message selections were based on those messages that were highly recognizable in both visual and auditory modalities so that both unimodal and bimodal messages could be created.

To increase mental workload, and also to reduce the length of the experiment, messages and corresponding questions were presented interleaved. During every 50-s delay period, either a message, or corresponding question, or another message was randomly presented to the participant. For example, the visual text message Pedestrian Crossing was presented to the participant. Then the auditory speech message "Snow Advisory" was presented, followed by the question, "Caution due to?" with options of Snow Advisory and Speed Bumps Ahead. After the participant selected the correct answer and rated his/her self-confidence, the question referring to the first message appeared on the touch-screen, "What type of crossing zone did you pass?" with the options of Cattle Crossing and Pedestrian Crossing. This procedural manipulation required participants to sometimes maintain two messages in working memory.

RESULTS

DATA ANALYSES

Data were collected in four test phases: (1) driver recruitment phase; (2) pre-test message recall phase; (3) simulator testing phase; and (4) post-test message recall phase.

Driver Recruitment Phase

In the Driver Recruitment phase, drivers were asked questions (via telephone) about their age, number of years with driving experience, and number of times they drove per week. See table 16 for this demographic data.

Table 16. Sample of the primary results from "Driver Demographic Questionnaire" (as administered for experiment 5C).

DRIVER GENDER GROUP	DEMOGRAPHIC		DRIVING EXPERIENCE		
	AGE	YEARS LIVING IN SEATTLE	YEARS AS A LICENSED DRIVER	YEARS DRIVING IN SEATTLE	AVERAGE MILES DRIVEN ANNUALLY
MALES	23.8	8.7	7.4	4.7	19131
FEMALES	22.3	11.2	6.1	3.6	13421

Pre-Test Message Recall Phase

In the pre-test message recall phase, drivers were asked to recall 10 earcons and 7 icons. For the earcon pretest, only two drivers incorrectly recalled messages on the first test. The rest of the drivers received a perfect score on the first two tests. For the icon pretest, all drivers got all seven icons correct on the first two tests.

Simulator Testing Phase

The simulator testing phase consisted of collecting several objective and subjective dependent measures that included two primary categories. The first category is **message screen data,** which included the following dependent variables: (1) message recognition (percent correct), (2) recognition latency (measured in seconds), and (3) self-confidence rating of message recognition questions (on a scale from 1 to 100). The second category of dependent measures is **driving performance data,** which includes the following: (1) mean speed, (2) standard deviation of speed, (3) mean lane position, (4) standard deviation of lane position, (5) number of accidents, and (6) number of speeding tickets. These two categories are outlined below.

The message screen data were evaluated to determine how the drivers in each of the alert groups (between-subjects variable) would differ in their performance for each message modality (within-subjects variable) as well as in each delay condition (within-subjects variable). It was hypothesized that performance would be worst for the drivers not receiving an alert. We also expected that driver recognition accuracy, latency, and confidence levels would be superior with redundant or bimodal information. Finally, we hypothesized that the bimodal message modality would be more memorable than the unimodal message modality after the 50-s delay condition.

The driving performance data were broken down into three time periods surrounding and including the presentation of the message. These three time periods are as follows: 6 s prior to the presentation of a message (pre-message window); 8 s during the presentation of the message (during message window), and 6 s after the presentation of the message (post-message window). Only half of the data were eligible for the post-message window because immediately following half of the messages, drivers would answer a question about the message (0-s delay condition). We were only interested in driving performance data, not driving performance with a concurrent activity. Therefore, the post-message window was only calculated for the 50 percent of the data where there was a 50-s delay between the message presentation and the question.

The during-message window evaluated drivers' initial reactions to particular message types. We hypothesized that drivers would slow down during the visual messages, slow a moderate amount during the visual plus auditory messages, and maintain their original speed with the auditory messages. We also hypothesized that their speed deviations would be greater with the visual messages, moderate with the visual plus auditory, and smallest with the auditory messages. Lane position and standard deviation of lane position would fluctuate most with the visual messages, moderately fluctuate with the visual plus auditory, and fluctuate least with the auditory only messages. These hypotheses are based on the premise that subjects will have to avert their gaze to view the visual messages. All of the predicted effects are hypothesized to be much stronger in the first scenario than in the second.

The post-message window investigated any latent reactions to particular message types. We hypothesized that the visual message modality would cause a slower mean speed and larger speed deviations than the other two message modalities. All of the predicted effects were hypothesized to be stronger in the first scenario than in the second.

MESSAGE SCREEN DATA

Three measures of message screen data were measured: (1) recognition accuracy, (2) recognition latency, and (3) self-confidence of recognition. The relevant ANOVA tables for experiment 5C can be found in appendix C. Note that missing data comprised less than 3 percent of the entire data set. Missing data occurred because of drivers' failure to answer the recognition question within a 10-s period. If the drivers failed to answer the question, it was considered an incorrect response for the recognition accuracy data. The individual's mean latency and mean self-confidence scores for message modality and delay were used to replace missing data values. Thus, this analysis was conducted on a full data set.

Recognition Accuracy

The three types of alerts had no effect on drivers' recognition accuracy, F (2, 33) = 1.71, p = 0.196. Drivers in the tone-alert (M = 0.984), speech-alert (M = 0.969), and no-alert (M = 0.964) conditions were maintaining a very high level of recognition accuracy, which indicates a ceiling effect.

An alert by message modality interaction was found when only the visual and auditory message modality types were compared (see figure 38). Drivers maintained higher accuracy with the auditory messages than they maintained with the visual messages for all alert types, F (2, 33) = 3.85, p = 0.032. Drivers' accuracy in the visual message modality no-alert condition was significantly lower than in the visual message modality tone-alert condition [t (33) = 8.33, $p < 0.001$] and the visual speech-alert condition [t (33) = 6.67, $p < 0.01$]. There were also significant accuracy differences in the no-alert condition between the visual and the auditory message modalities, t (33) = 5.15, $p < 0.025$. These results were anticipated as it would be harder to know when a visual message was being presented without an alert. These differences between message modalities and alerts, while statistically significant, were so small that they have minimal implications for guidelines.

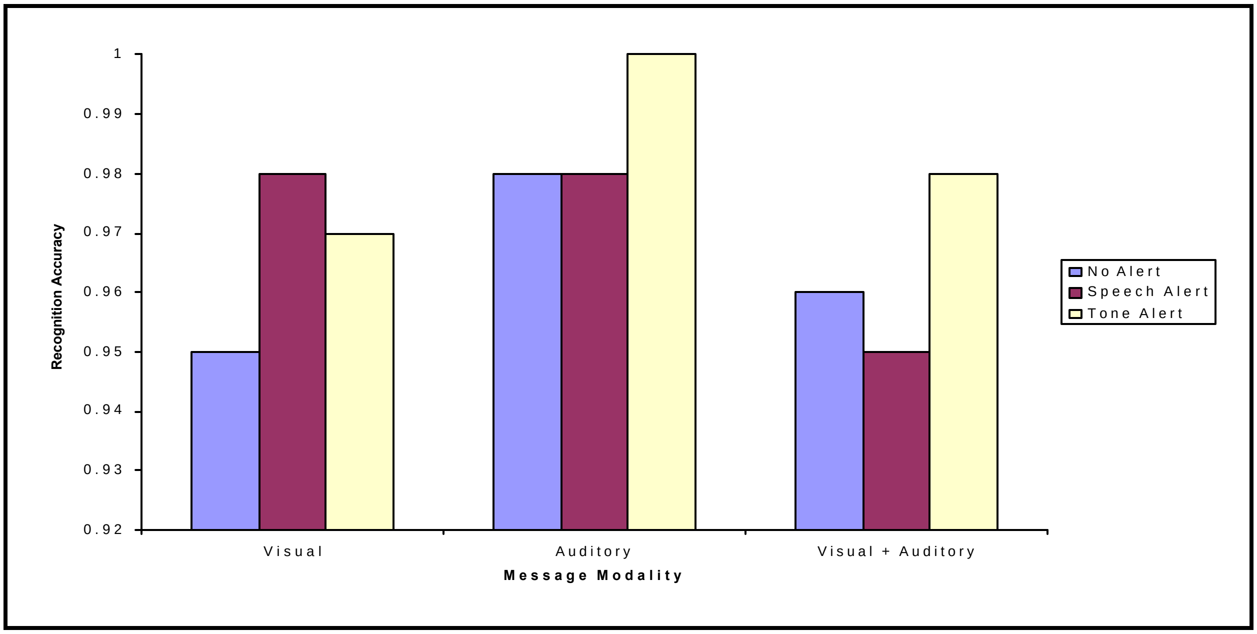


Figure 38. The effects of alert type and message modality on recognition accuracy.

There was a message modality by delay interaction where drivers' recognition accuracy was equally high with all three message types in the 0-s delay condition but significantly different in the 50-s delay condition (see figure 39), F (2, 66) = 5.95, p = 0.006. In the 50-s delay condition, drivers were more accurate with the auditory messages than they were with the visual messages [t (66) = 6.5, $p < 0.01$] or with the visual plus auditory messages [t (66) = 10.5, $p < 0.001$]. There was also a significant difference between the visual message modality and the visual plus auditory message modality with a 50-s delay, t (66) = 4.0, p = 0.05. These findings support our hypotheses that drivers seem to remember the auditory messages better than they remember the visual ones. It was surprising that drivers remembered the auditory messages better than the visual plus auditory messages. Performance differences between 0-s delay and 50-s delay intervals were significantly different for the visual, t (66) = 8.75, $p < 0.001$, auditory, t (66) = 4.5, $p < 0.025$, and visual plus auditory, t (66) = 16.75, $p <$ 0.001. This indicates that performance differences involving message modality occur when the drivers have to remember the message for 50 s.

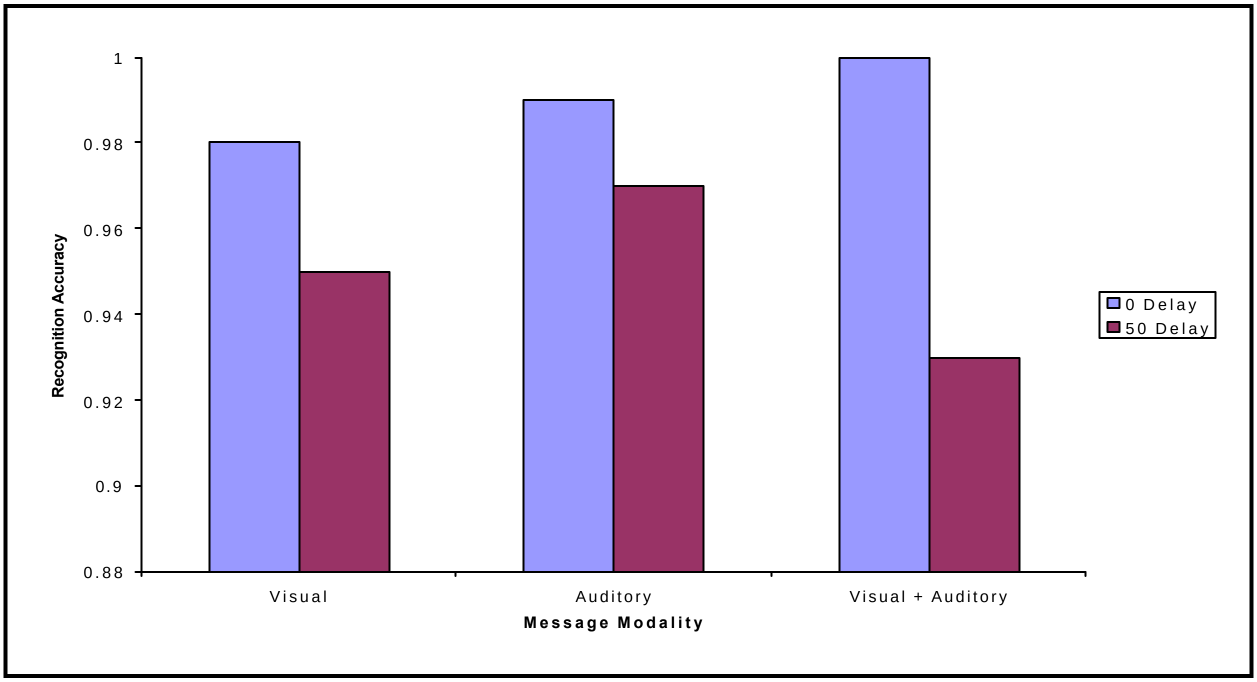


Figure 39. The effects of message modality and delay on recognition accuracy.

Figure 40 shows the interaction between message modality and message format. This interaction did not achieve significance at the 0.05 level, $F(1, 33) = 0.260$, $p = 0.612$. This result suggests that drivers perform equally well with the symbolic and lexical message formats regardless of whether they are presented in the visual or auditory modality. This finding goes against our hypotheses in that we anticipated drivers' performance to be superior with the auditory modality for both symbolic and lexical message formats, moderate in the visual modality/symbolic message format, and worst in the visual modality/lexical message format. One possible reason why our hypotheses were incorrect is because all drivers were very accurate in their responses, causing a ceiling effect.

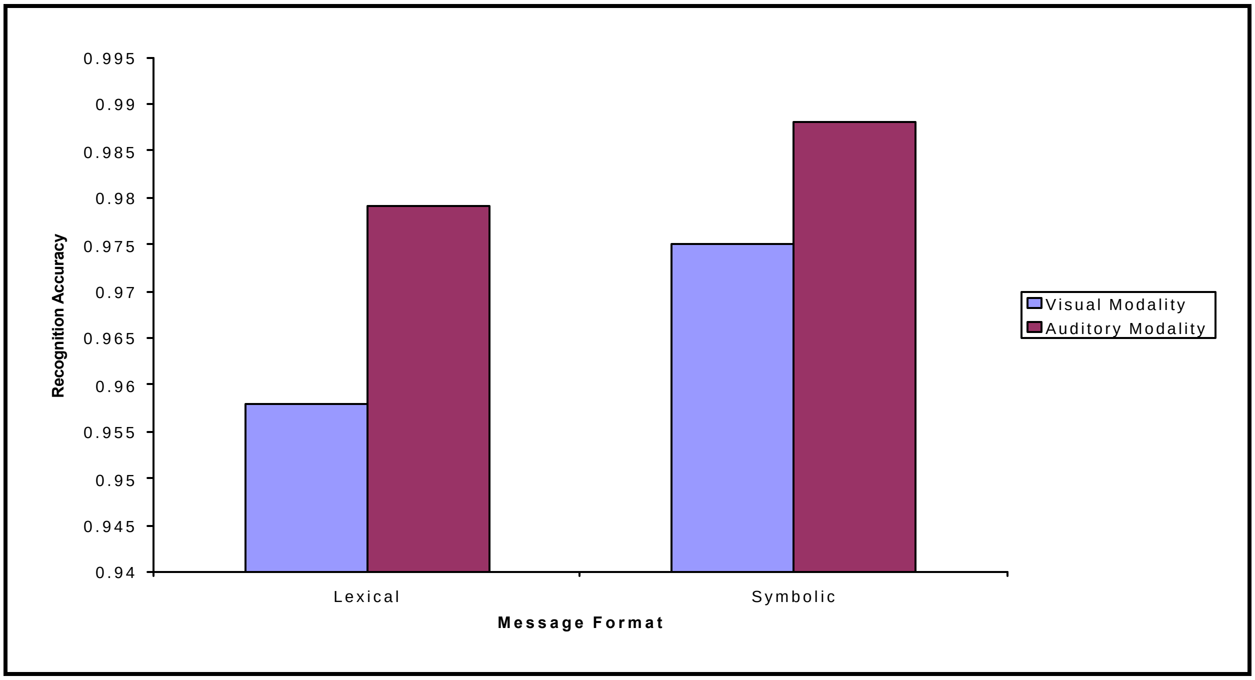


Figure 40. The effects of message modality and message format on recognition accuracy (ns).

Drivers' recognition accuracy was significantly higher with auditory messages (M = 0.984) than with visual messages (M = 0.966) and visual plus auditory messages (M = 0.966), F (2, 66) = 5.18, p = 0.01. Drivers responded more accurately to the auditory messages than to either the visual [t (66) = 6.0, $p < 0.01$] or the visual plus auditory messages [t (66) = 6.0, $p < 0.01$]. This was a surprising finding as it was expected that higher recognition accuracy rates would be associated with the visual plus auditory messages.

Drivers were generally more accurate with symbolic (M = 0.981)) message formats than they were with the lexical (M = 0.969) message formats, F (1, 33) = 4.12, p = 0.051. This finding supports our hypothesis that drivers will be able to acquire information easier when it is in a symbolic format than when it is in a lexical format. This difference was so small that it has little implications for guidelines.

There was a larger performance decrement found between scenario 1 and scenario 2 in the 50- s delay condition than was found in the 0-s delay (see figure 41), F (1, 33) = 6.61, p = 0.015. The recognition accuracy scores for scenarios 1 and 2 were significantly different for the 50-s delay condition [t (33) = 6.07, $p < 0.025$] but not with the 0-s delay condition. This finding suggests that drivers' memory for messages will improve after they become more familiar with the messages.

Drivers responded more accurately when they answered the question immediately (M = .99) than when they answered a question 50 s later (M = .95), F (1, 33) = 21.97, p = 0.001. This finding was not surprising and establishes that drivers' memory does decay slightly during the 50-s delay period.

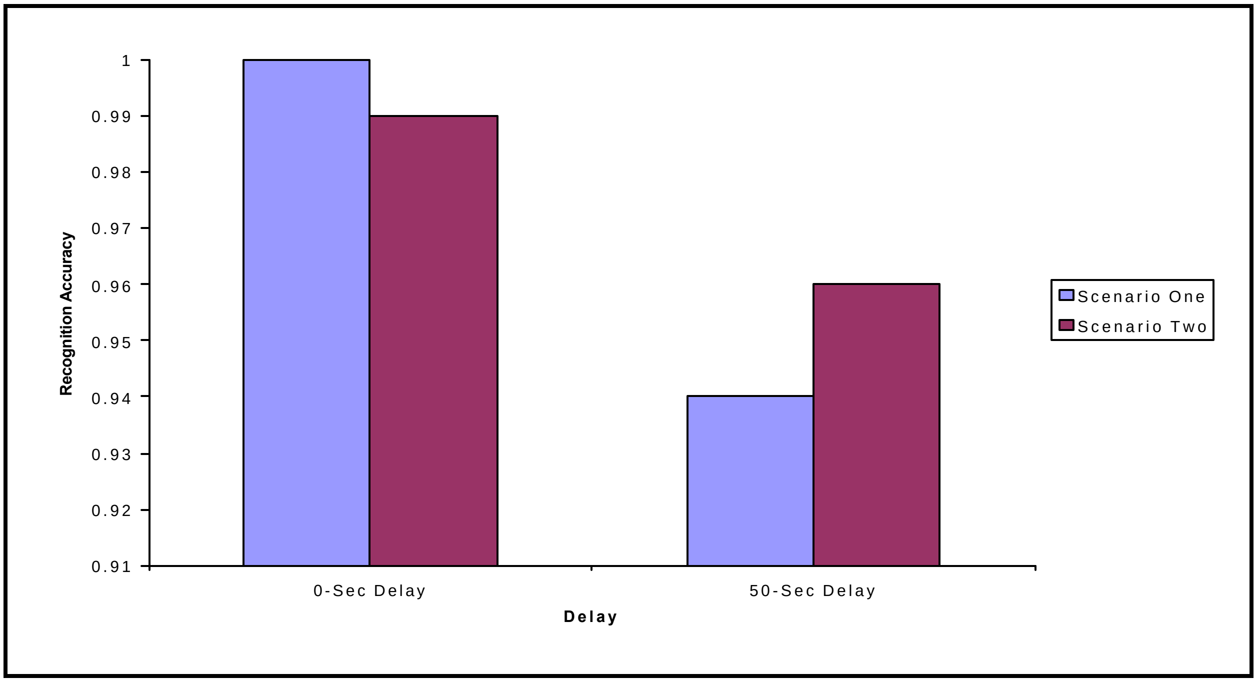


Figure 41. The effects of scenario and delay on recognition accuracy.

Recognition Latency

Figure 42 shows the alert by message modality interaction, F (2, 33) = 3.50, p = 0.042. Drivers in the no-alert condition were significantly faster with the auditory messages than with the visual messages, t (33), 6.06, p < 0.01. There were no significant differences between the visual and auditory messages for speech [t (33) = 2.61, p > 0.10] and tone alerts [t (33) = 1.12, p > 0.25]. This was an unexpected finding because it was hypothesized that drivers would be able to perform more quickly with a warning present than without a warning present. This finding may suggest that auditory messages do not require an auditory alert but visual messages do require an alert for faster response.

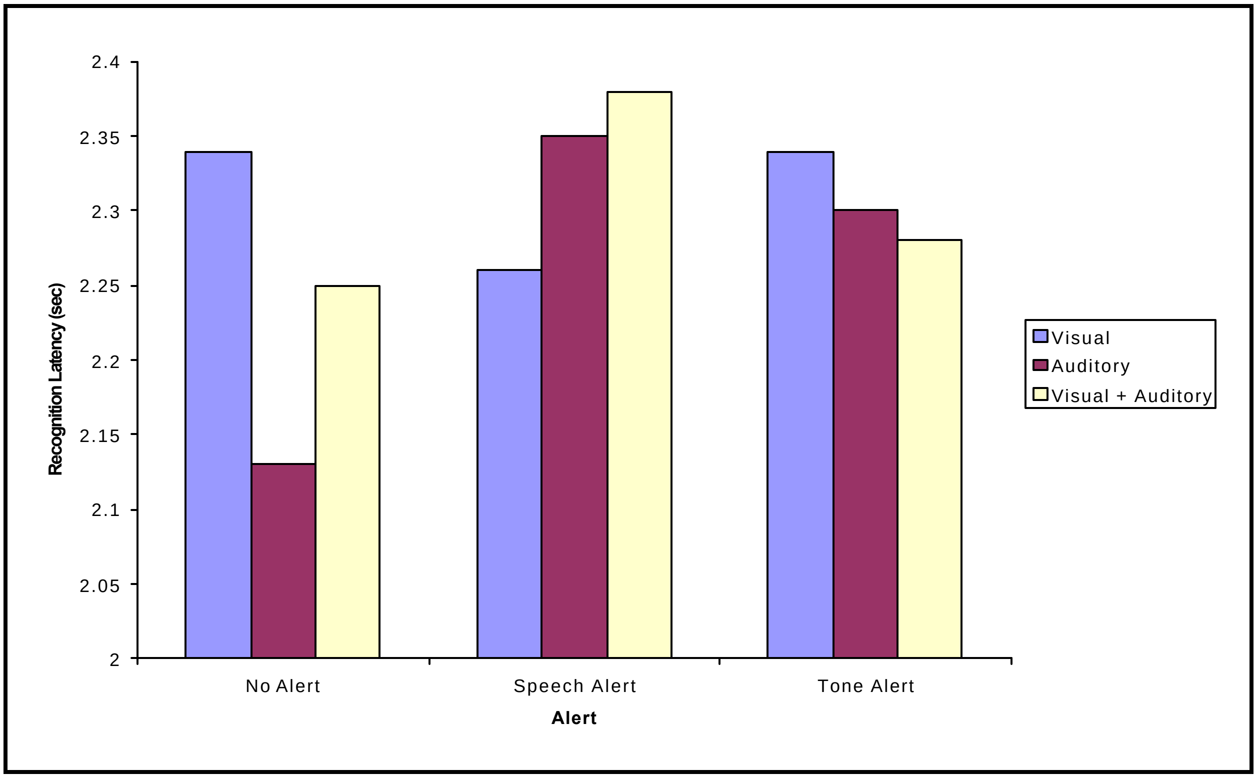


Figure 42. The effects of alert type and message modality on recognition latency.

Figure 43 shows that there was an interaction with message modality, scenario, and delay. Subjects' response times were longer in scenario 1 with a 50-s delay than in scenario 2 with a 50-s delay for all three message modalities, F (2, 66) = 3.54, p = 0.039. For scenario 2, drivers' performance with the auditory message modality was statistically equivalent to performance with the visual plus auditory message modality for the 0-s delay condition, t (66) = 0.857, p > 0.25. Also for scenario 2, performance with the auditory message modality was significantly faster with the 50-s delay than performance with the visual plus auditory, t (66) = 4.57, p < 0.025. This was the same result found with measures of recognition accuracy. It was hypothesized that drivers would perform best with the visual plus auditory message modality; however, it was not expected that performance with the auditory message modality would be equivalent or superior.

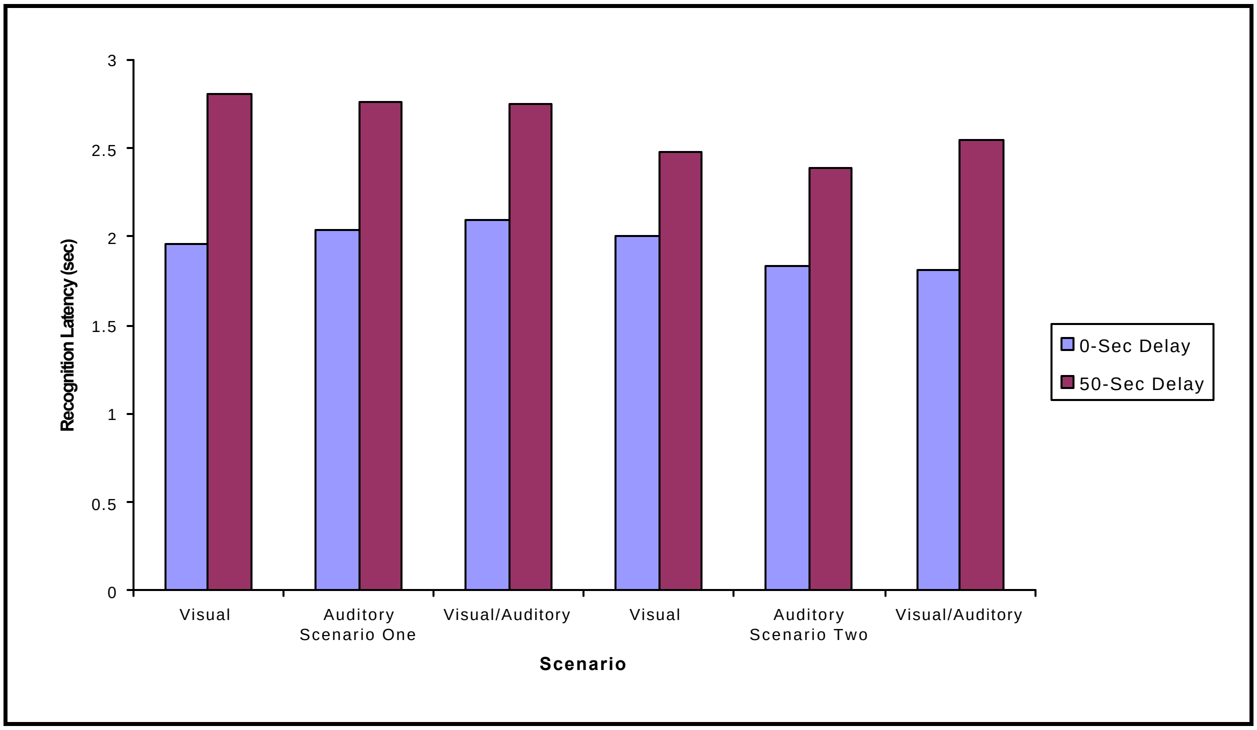


Figure 43. The effects of message modality, scenario, and delay on recognition latency.

Figure 44 depicts the scenario by message format interaction, F (1, 33) = 20.51, p = 0.001. There were significant differences between the two scenarios with the lexical message format [t (33) = 16.45, p < 0.001], but not between the two scenarios with the symbolic message format, t (33) = 2.90, p > 0.10.

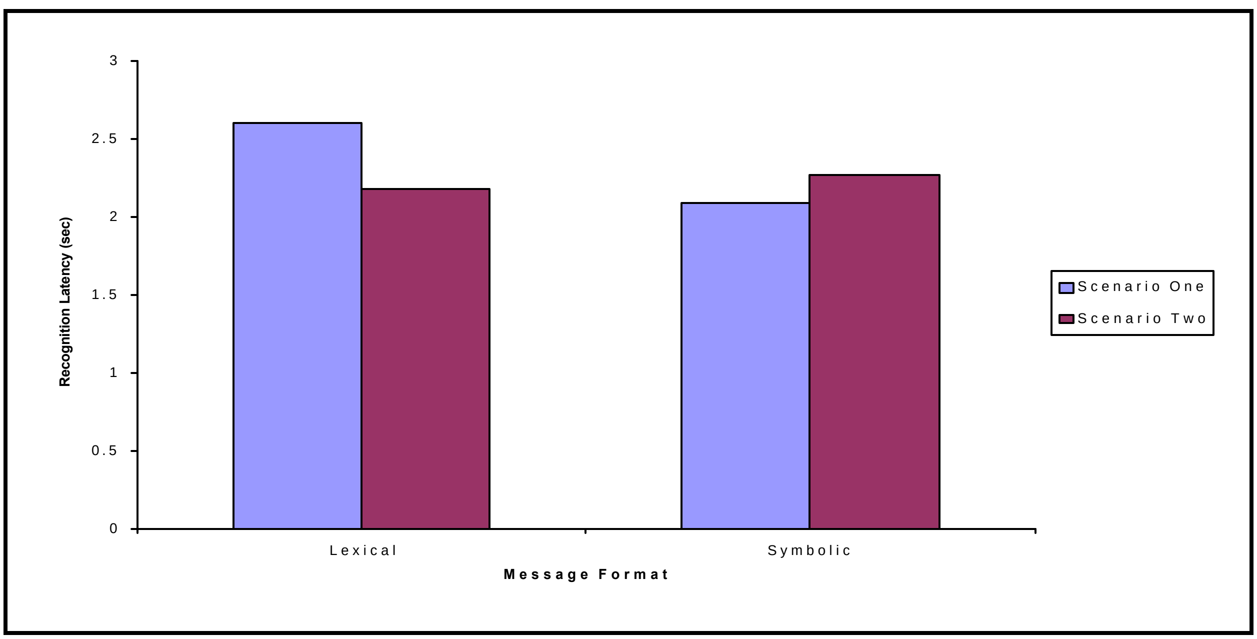


Figure 44. The effects of message format and scenario on recognition latency.

Drivers responded generally faster to the symbolic messages (M = 2.23 s) than they did to the lexical messages (M = 2.34 s), F (1, 33) = 6.56, p = 0.015. This finding supports our hypothesis that drivers' performance will be faster with the symbolic message format than with the lexical message format.

Figure 45 shows the scenario by delay interaction, F (1, 33) = 4.09, p = 0.051. Drivers' performance was significantly different between scenarios 1 and 2 for both the 0-s delay condition [t (33) = 7.14, p < 0.025] and the 50-s delay condition [t (33) = 13.81, p < 0.001]. This finding indicates that drivers' performance improved in scenario 2 for both message presentation delays. This was an expected finding as drivers' performance will usually improve over time.

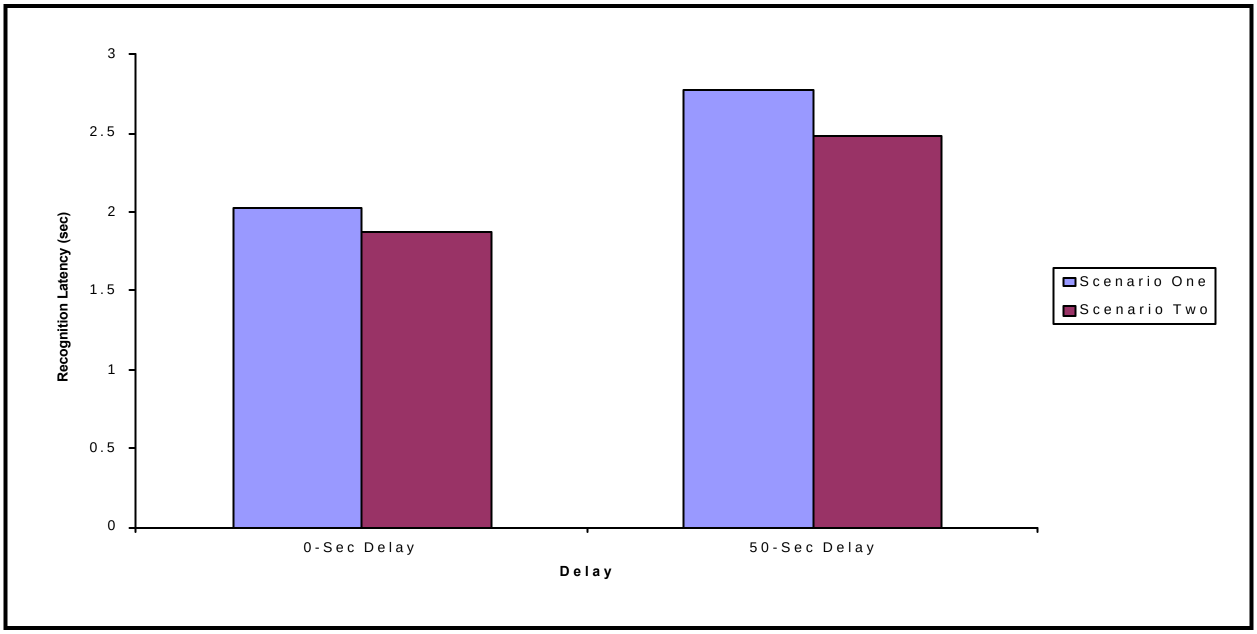


Figure 45. The effects of scenario and delay on recognition latency.

Drivers responded significantly faster in the 0-s delay condition (M = 1.959 s) than they did in the 50-s delay condition (M = 2.624 s), F (1, 33) = 45.07, p = 0.001. This finding suggests that it takes drivers longer to respond if they must remember the message for 50 s than if they answer the question immediately (0-s delay).

Drivers improved their response times over the two scenarios (first scenario M = 2.403 s, second scenario M = 2.179 s). Drivers responded as expected in that they were significantly faster in scenario 2 than they were in scenario 1, F (1, 33) = 19.38, p = 0.001.

Self-Confidence for Recognition Accuracy

Figure 46 shows that drivers were more confident overall with the tone alert for all three message modalities than with any other alert, F (4, 66) = 2.70, p = 0.038. An interesting point within this interaction was the significantly lower confidence ratings of the visual message modality and no-alert condition versus the high confidence ratings of the visual messages with the speech alert [t (66) = 12.2, p < 0.001] and tone alert [t (66) = 11.4, p < 0.001]. This supports our hypothesis that drivers were not as confident of their responses to the visual messages without an alert present. What was not hypothesized was that drivers would be more confident with the auditory message modality than with the visual plus auditory message modality. Drivers were significantly more confident of the auditory message modality than the visual plus auditory message modality in the no-alert [t (66) = 3.71, $p <$ 0.01] and in the speech-alert conditions [t (66) = 3.1, $p < 0.025$]. Drivers' confidence levels were equivalent for the tone-alert condition. This suggests that the tone alert assists drivers in their recognition confidence better than the speech alert or having no alert presented at all. Aural messages apparently do not require an alert whereas the visual messages and the visual plus auditory messages do require an alert for high self-confidence ratings.

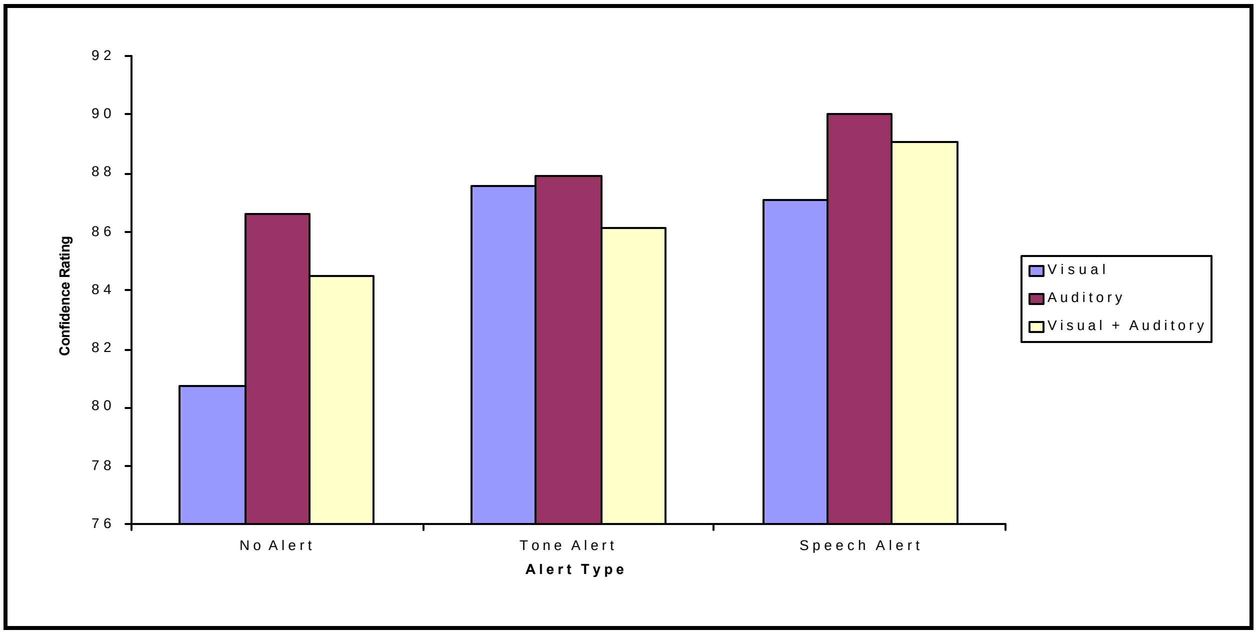


Figure 46. The effect of alert type and message modality on self-confidence.

Figure 47 shows the three-way interaction of alert by message modality by delay. The results suggest once again that the differences in confidence ratings are generally found when the drivers must remember the message format for an extended period of time, F (4, 66) = 4.29, $p = 0.004$. Confidence ratings were higher across message modalities and delay conditions with the tone- alert group than with either the speech- or no-alert groups. In the 50-s delay, confidence ratings were higher with the auditory modality versus the visual plus auditory modality for the no-alert group [t (66) = 7.20, $p < 0.001$] and the speech-alert group [t (66) = 3.41, $p < 0.025$]. Confidence ratings between the auditory and the visual plus auditory were not significantly different with the 50-s delay for the tone-alert group, t (66) = 2.17, $p < 0.100$. Once again, the auditory message modality performed well with or without an alert present, whereas the visual and visual plus auditory modalities only performed well with the tone alert present.

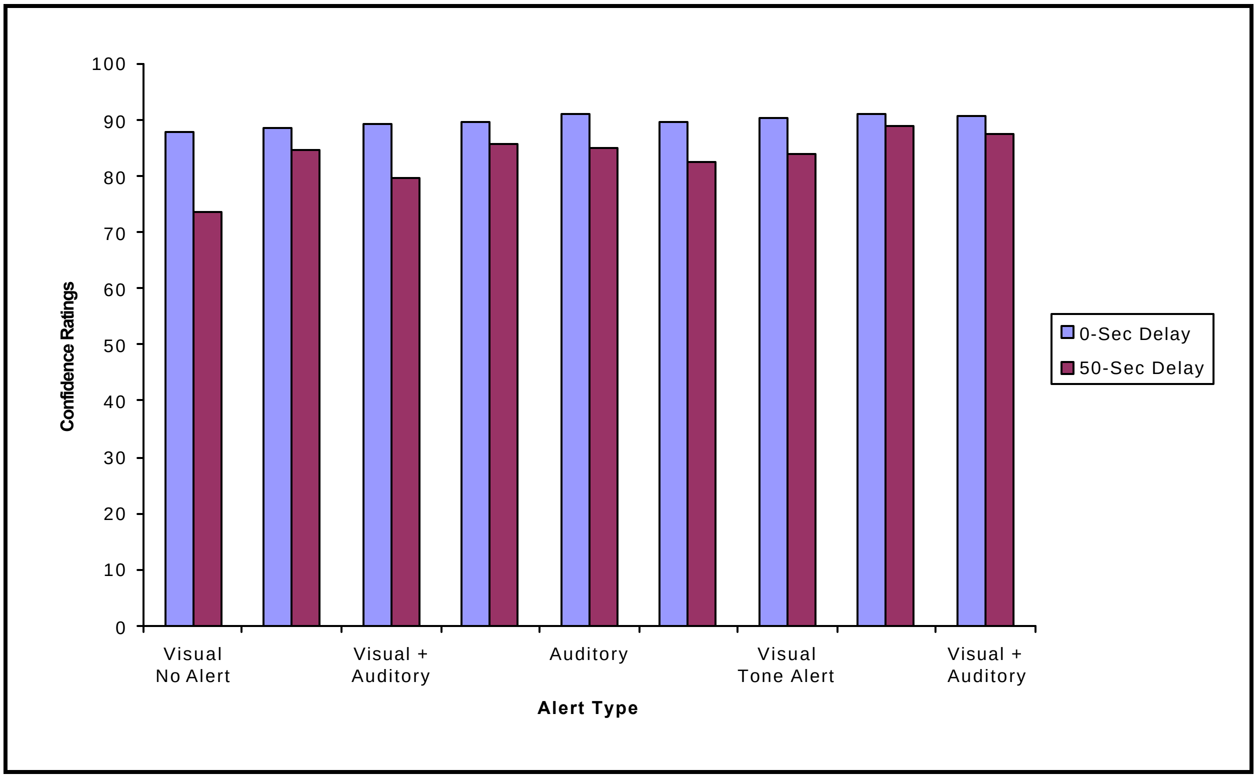


Figure 47. The effects of alert, message modality, and delay on self-confidence ratings.

Figure 48 shows the alert by message format interaction, F (2, 33) = 4.79, p = 0.015. Drivers in the no-alert group maintained higher confidence with the symbolic message format than with the lexical message format, t (33) = 9.13, p < 0.001. There were no differences between the message formats for the other two alert groups. There also was a significant difference for the lexical message formats between no alert and speech alert [t (33) = 14.18, p < 0.001] and the no alert and tone alert [t (33) = 13.70, p < 0.001]. These results suggested that presenting an alert will help increase drivers' confidence in the lexical message format but an alert is not necessary for the symbolic message format. This finding is important for design guideline development.

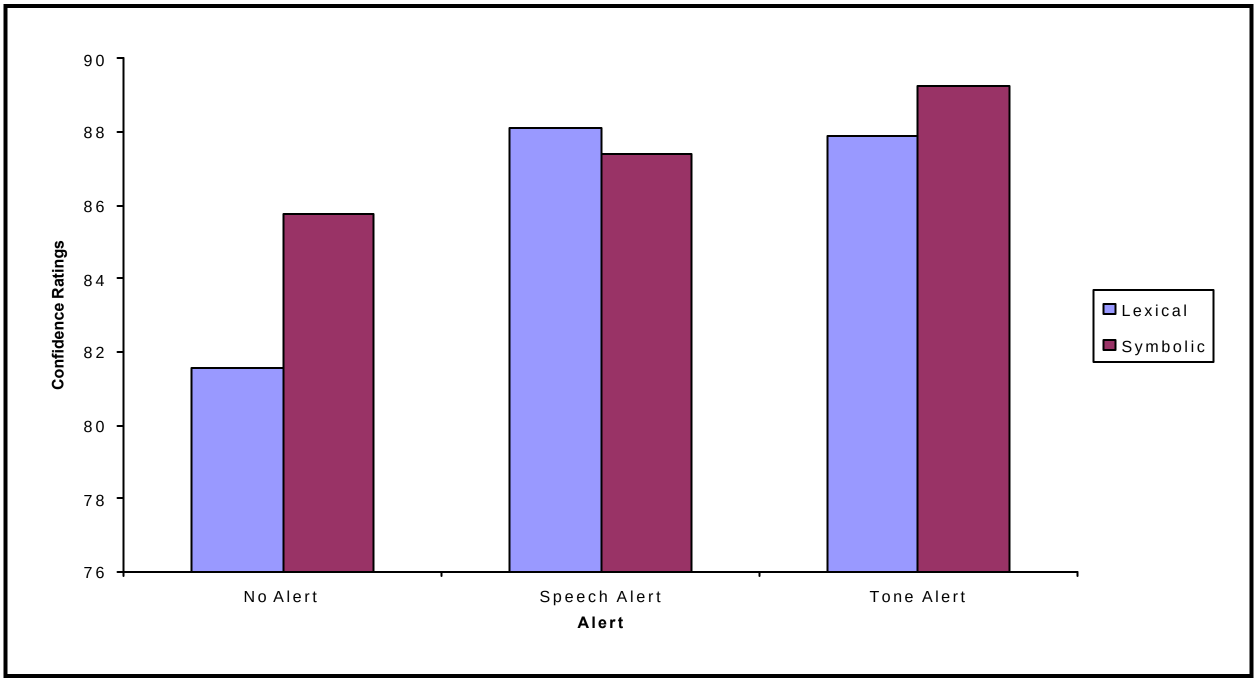


Figure 48. The effects of alert type and message format on self-confidence ratings.

Figure 49 shows the message modality by delay interaction. Drivers were more confident in the 0-s delay condition with all three message modalities; however, confidence ratings show significant differences with the 50-s delay across all three message modalities, $F\ (2, 66) = 5.16$, $p = 0.008$. There was a significant difference between the visual and the auditory message modality for the 50-s delay condition [$t\ (66) = 13.33$, $p < 0.001$], the visual and the visual plus auditory condition [$t\ (66) = 5.95$, $p < 0.01$], and the auditory and the visual plus auditory condition [$t\ (66) = 7.38$, $p < 0.1$]. These findings suggest that confidence ratings among the three message modalities only differ when the drivers were required to remember the message for an extended period of time. These findings also indicate that the auditory message modality was preferred by drivers over the visual plus auditory and the visual modalities. This result contradicts our original hypothesis that the visual plus auditory message modality would be the preferred modality.

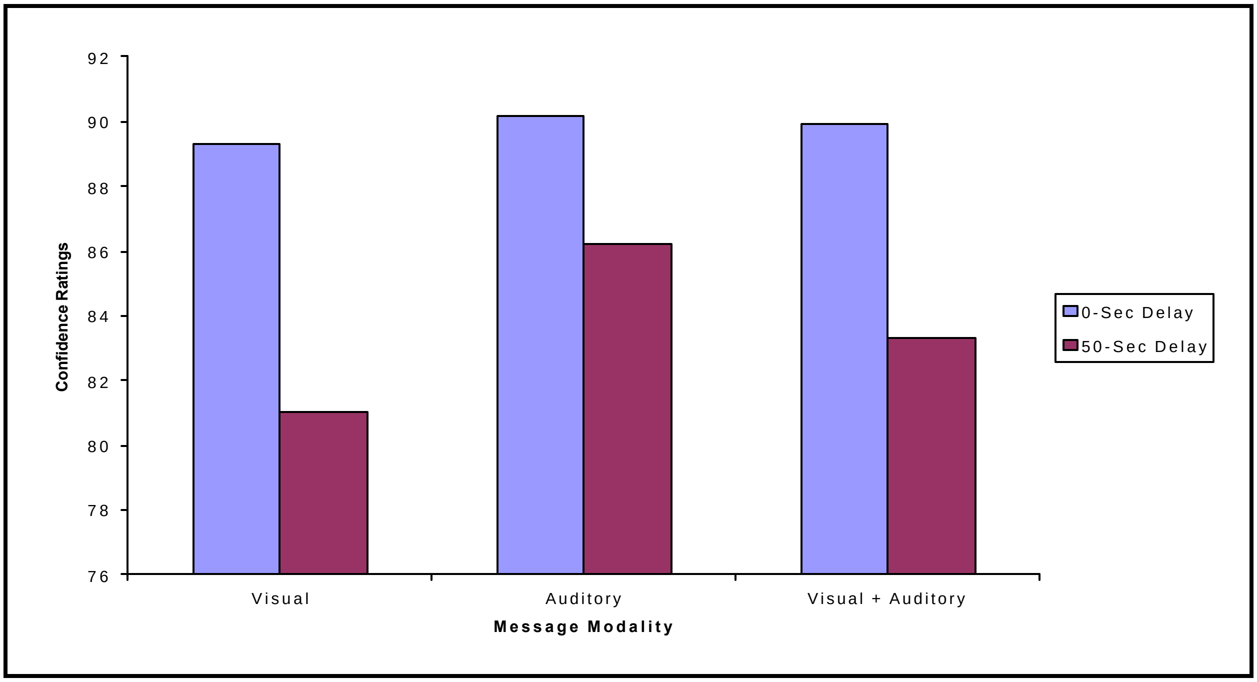


Figure 49. The effects of message modality and delay on self-confidence ratings.

Figure 50 shows the message modality by message format interaction. Confidence levels differed between the visual/lexical and the visual/symbolic messages, while no differences existed between the auditory/lexical and auditory/symbolic messages, F (1, 33) = 6.02, $p = 0.02$. There was a significant difference between the confidence ratings of the visual/lexical messages and the visual/symbolic messages [t (33) = 8.60, $p = 0.01$]. There were no differences between the visual/symbolic messages and the auditory/symbolic messages [t (33) = 3.75, $p > 0.05$] or the auditory/symbolic and auditory/lexical [t (33) = 0.099, $p > 0.25$], which suggests that drivers were least confident of the visual lexical message. This finding supports our original hypothesis and has important implications for the design guidelines.

Drivers' confidence ratings indicated that the auditory message modality was preferred (M = 88.2), visual plus auditory messages were moderately high (M = 86.6), and the visual messages were lowest (M = 85.2), F (2, 66) = 7.12, $p = 0.002$. Confidence ratings for auditory messages were significantly higher than the visual plus auditory messages [t (66) = 4.86, $p < 0.025$] and the visual messages [t (66) = 9.23, $p < 0.001$]. The visual messages were significantly lower than the visual plus auditory messages [t (66) = 4.37, $p < 0.025$]. This finding was surprising in that it was expected that drivers would be more confident of the visual plus auditory message modality than either the visual or the auditory message modality.

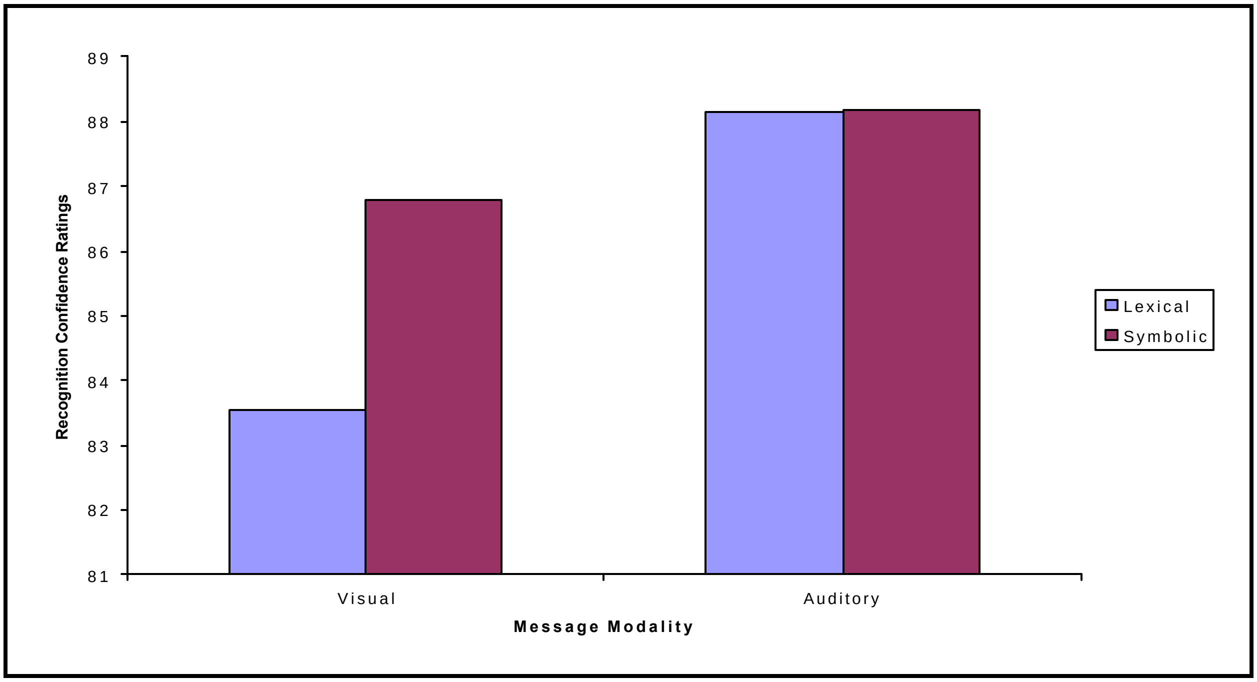


Figure 50. The effects of message modality and message format on recognition confidence.

The scenario by delay interaction indicated that confidence ratings for scenarios 1 and 2 were much higher in the 0-s delay condition than they were in the 50-s delay condition (see figure 51), F (1, 33) = 8.66, $p = 0.006$. Confidence rating differences for the 50-s delay condition were significantly lower in scenario 1 than in scenario 2 [t (33) = 8.89, $p < 0.01$]. This finding suggests that, over time, drivers gained more confidence in their answers in the 50-s delay condition.

Drivers were more confident of their answers when there was no delay between the message presentation time (M = 89.9) and their answers versus when there was a 50-s delay (M = 84.5), F (1, 33) = 24.93, $p = 0.001$.

Drivers' confidence ratings were significantly higher for the second scenario (M = 87.5) than they were for the first (M = 85.9), F (1, 33) = 6.06, $p = 0.019$.

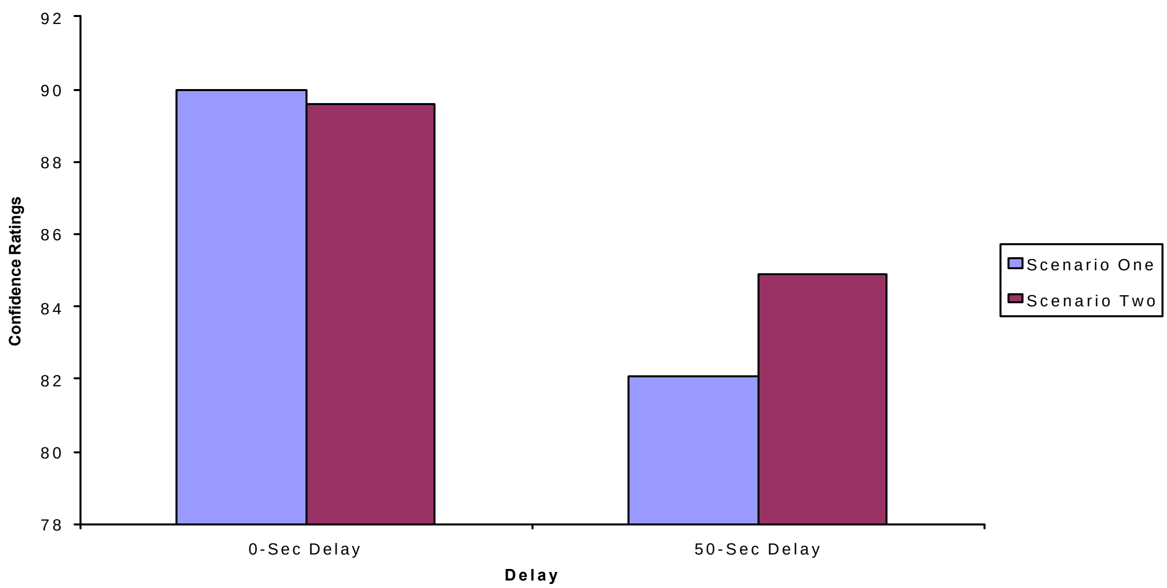


Figure 51. The effects of scenario and delay on self-confidence ratings.

DRIVING PERFORMANCE DATA

Figure 52 shows mean speed as a function of Window and Message Modality. Mean speed was slightly lower for the Pre-Window (50.4 ft/s) than for the During (50.5 ft/s) or Post-Window (50.6 ft/s), $F(2,66) = 3.60$, $p < 0.001$. While there were other statistically significant effects, their magnitude was too small to be important for guideline development.

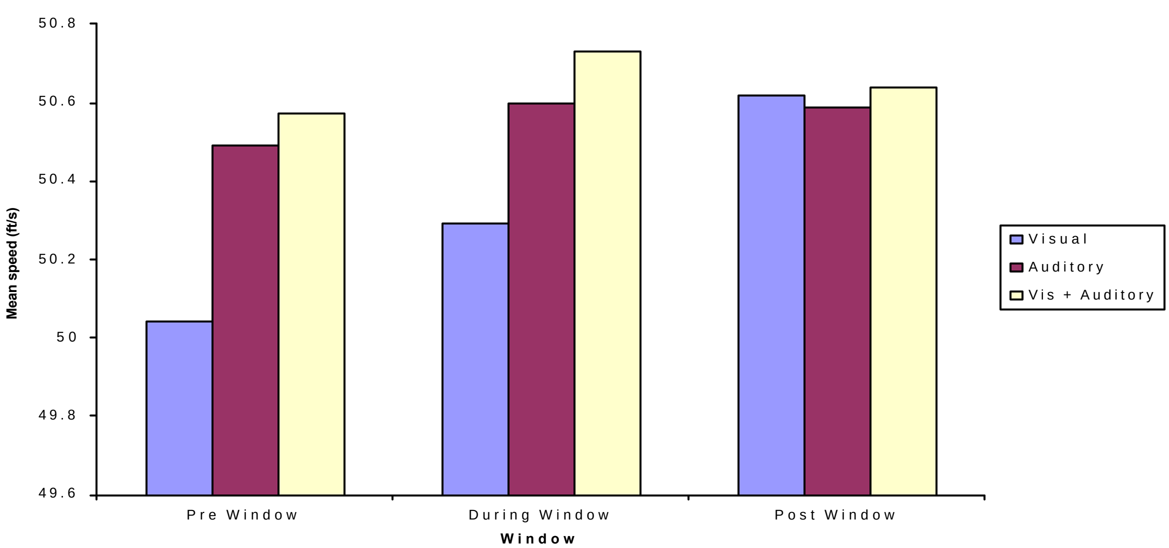


Figure 52. The effect of window and message modality on mean speed.

Figure 53 shows that the standard deviation of speed was slightly higher for the Pre-Window (0.328 ft) than for the During (0.211) for Post-Window (0.203 ft), $F(2, 66) = 38.6$, $p < 0.001$. The significant interaction between Window and Modality found in figure 53, $F(4, 132) = 23.9$, $p < 0.001$, shows that the decrease from Pre- to During-Window is greater for auditory or visual messages relative to combined auditory plus visual messages. By the time the Post Window is entered, the initial effects of message modality have disappeared. These findings are consistent with the hypothesis that processing in-vehicle messages temporarily suppresses accelerator control actions for uni-modal messages. However, it is not at all clear why control actions are not suppressed by messages that combine visual and auditory stimuli.

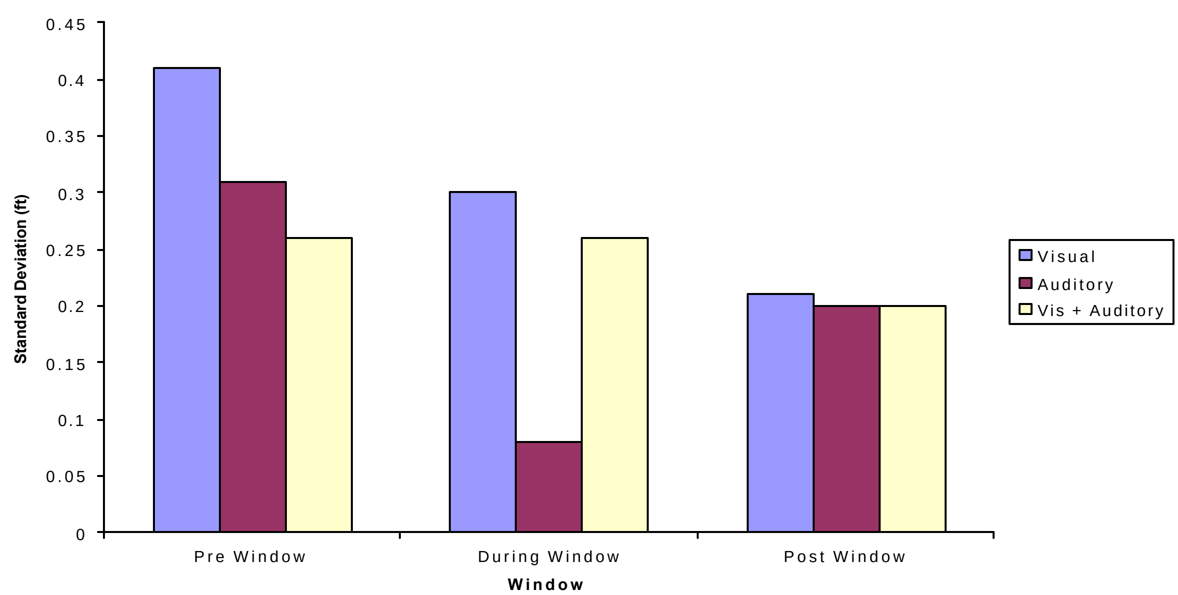


Figure 53. The effects of message modality and window on standard deviation of speed.

Figure 54 shows standard deviation of lane position is greater for the Pre-Window (0.700 ft) than for the During (0.519 ft) or Post-Window (0.579 ft), $F(2, 66) = 82.8$, $p < 0.001$. The significant interaction displayed in this figure, $F(4, 132) = 107$, $p < 0.001$, is largely due to a decrease in auditory message standard deviation for the During Window. This finding is consistent with the hypothesis that auditory messages command attention (Lee et al., 1998) and so suppress steering-wheel control actions.

These results taken together show that in-vehicle message presentation does not impair driving performance. Indeed, there is some evidence that processing in-vehicle messages tends to suppress lateral and longitudinal control actions by the driver, resulting in less variable positioning of the vehicle on the simulated roadway.

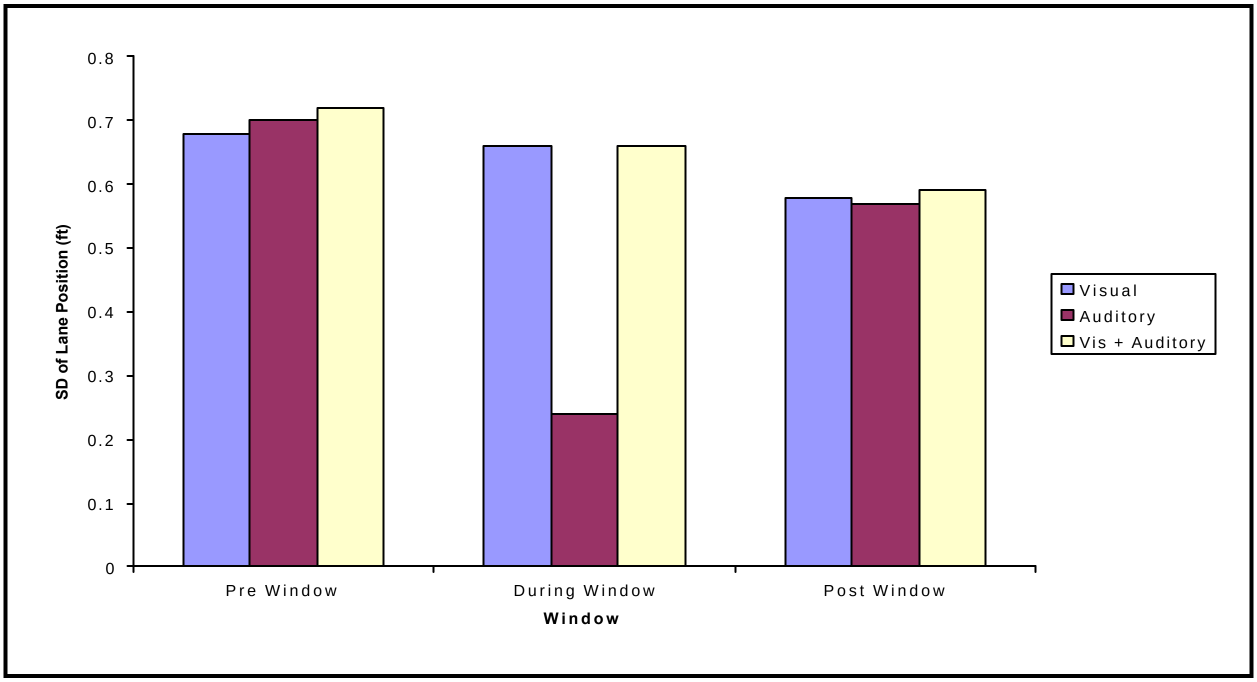


Figure 54. Effects of window and message modality on standard deviation of lane position.

CHAPTER 4. DISCUSSION

DRIVER AGE

Older drivers have considerable difficulty in comprehending the meaning of in-vehicle icons and earcons. The most important finding in these experiments was the complete inability of older drivers to learn the set of auditory earcons (experiment 5B). While younger drivers had almost perfect recall scores, not a single older driver was able to continue with the simulator portion of the experiment because they could not learn the earcons even with extended practice. This has very strong implications for design guidelines. Earcons, especially if of brief duration as in experiment 5B, may not be at all suitable for older drivers.

In sharp contrast, younger drivers fared quite well with earcons. All younger drivers learned the set of earcons in less than 25 minutes and had recall scores above 98 percent. Furthermore, recognition accuracy during simulated driving was also about 98 percent. Naturalistic earcons were responded to more quickly than complex earcons (figure 37).

While older drivers were able to learn the set of visual icons, their comprehension scores, although high (90 percent), were less than scores of younger drivers (99 percent). This implies that icons are highly appropriate as in-vehicle messages for both older and younger drivers. This is especially true when familiar symbols are used (experiment 5A).

MESSAGE FORMAT AND COMPREHENSION

For visual messages, recognition scores were similar for text and icons, except for low-comprehension icons that were not familiar to many drivers (experiment 5A). For auditory messages, recognition scores for earcons and speech messages were also similar (experiment 5B). Contrary to our expectations, recognition accuracy was not improved for redundant visual plus auditory messages (experiment 5C). While scores were generally quite high, this result cannot be attributed to a ceiling effect because visual plus auditory scores were slightly but significantly lower (97 percent) than for auditory messages (98 percent). These results allow designers to use whatever message format and modality is most convenient.

MESSAGE FORMAT AND MEMORY

For visual messages, a 50-s delay imposed a slight loss in recognition, decreasing from 95 percent at zero delay to 90 percent. However, this loss was similar regardless of message format. Furthermore, this memory effect was largely confined to older drivers (experiment 5A). For auditory messages, similar results were obtained with the younger drivers showing no significant memory loss; recall that older drivers were not tested because they could not learn the earcons. However, in experiment 5C when drivers were in full control of the vehicle, the delay produced a small but significant memory decrease on the order of 4 percent that was especially pronounced for visual plus auditory messages (figure 39). Thus, while all message formats were comprehended about equally, they were not remembered equally well. Auditory messages seem to be most resistant to delay and this is an important point for designers to consider. Furthermore, symbolic and lexical message formats were equally resistant to delay so that designers may use the most convenient format.

VEHICLE CONTROL

In general, no adverse effects of in-vehicle message presentation were found when examining lateral and longitudinal vehicle control. In experiments 5A and 5B where drivers only controlled steering and the vehicle cruise-control maintained a constant speed, driving lane position and standard deviation of lane position were not altered when driving performance in a pre-message window was compared with a during-message window. In experiment 5C when drivers had full vehicle control, neither steering nor speed control was impaired by message presentation. While these results are quite encouraging for ATIS designers, since they imply that in-vehicle messages will not decrease road safety, nevertheless they should be interpreted with some caution because they were obtained in a simulator. No driver has ever died in a simulator accident. On-road replication of these findings will be useful.

APPENDIX A: QUESTIONNAIRES

SUBJECT SELECTION PHONE QUESTIONNAIRE

Subject Name ___________________________

Sub ID _______ Age _______ Gender ____ (1=M, 2=F)

Subject Selection Phone Questionnaire
Spring 1995

Note To Experimenter: DO NOT read the following "Purpose" to subjects.

Purpose: Before a subject can be selected to participate in Task K/Experiment 5A, he or she must have an active driver's license, drive at least twice per week, & not be prone to motion sickness.

Questions:

1) Do you have an active Driver's License? Yes (1) No (2)

2) How many times per week do you drive in Seattle or the surrounding areas?
< 1X (1) 1X (2) 2-3X (3) 4 + (4)

3) How often do you experience motion sickness when driving?
Never (1) Sometimes (2)* Often (3)**

* Experimenter: if subject answers "sometimes" to experiencing motion sickness, ask them further questions to try and assess if this is likely to be a problem in the simulator. If so, go to **!

** Experimenter: if the subject answers "often" to experiencing motion sickness, inform them of the following:

One potential risk with any simulator study is the possibility of "simulator sickness." Simulator sickness is similar to the motion sickness that some experience when traveling in a vehicle. Because you often experience motion sickness, there might be a chance of you experiencing motion sickness from our simulator. We don't want this to happen, so unfortunately you won't be able to participate in this study. We do however, greatly appreciate your time and interest, and if you like, we can put you on our list for other experiments. That way, if we have a need for subjects at any time in the future, we will contact you.

Scoring:

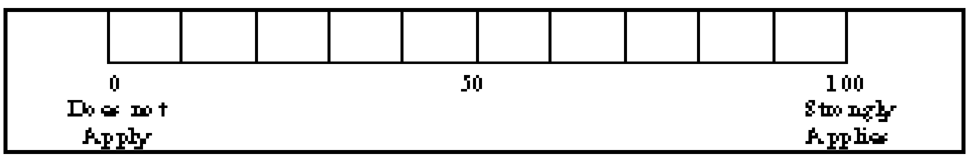
1) All subjects MUST have an active driver's license.
2) Subjects must drive at least 2 times/week.
3) Subjects must not experience motion sickness "often."

EXPERIMENT 5A PRE-STUDY QUESTIONNAIRE:
HOW COMFORTABLE ARE YOU WITH COMPUTERS?
(QUESTIONNAIRE REPEATED IN POST-STUDY.)

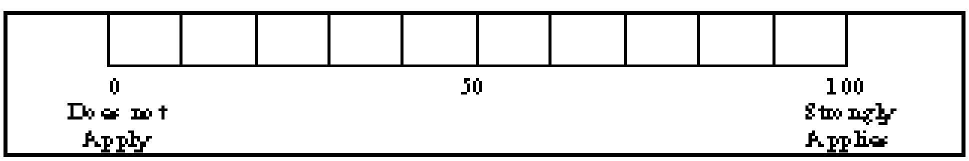
Experiment 5A
Post-Study Questionnaire
How Comfortable Are You With Computers?

It is important to us to understand how comfortable you feel with computers. Please mark with an "X" to indicate how much each statement below applies to you. Marking toward the 100 would indicate that a statement strongly applies. Marking toward the 0 would indicate that it does not apply.

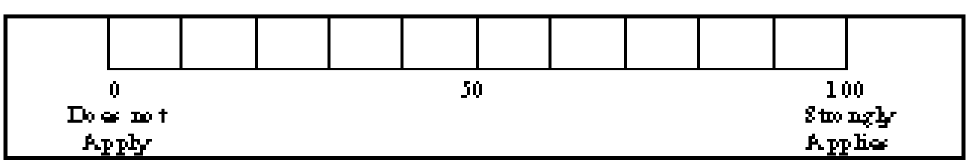
1. I am sure I could do work with computers.



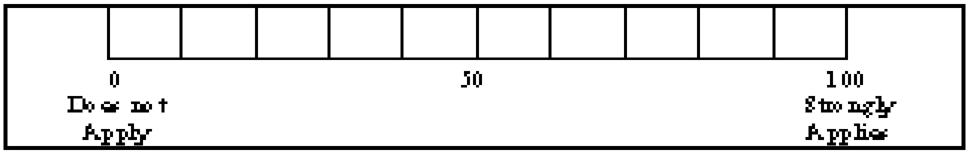
2. I would like working with computers.



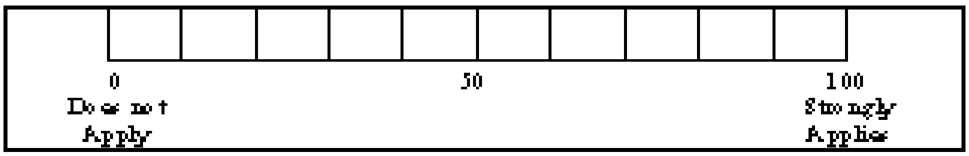
3. I would feel comfortable working with computers.



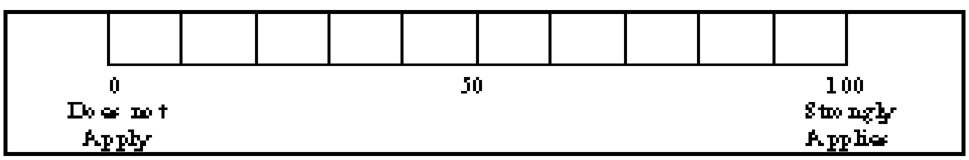
4. Working with a computer would make me very nervous.



5. I do as little work with computers as possible.

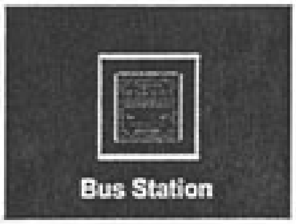


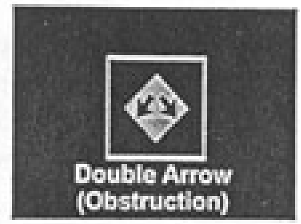
6. I think using a computer would be very hard for me.

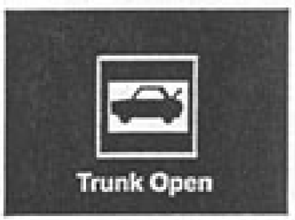


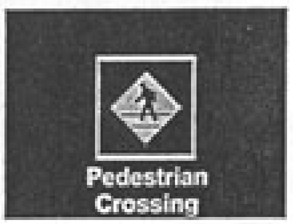
APPENDIX B: SYMBOL MESSAGES

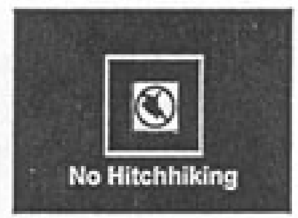
EXPERIMENT 5A DATA COLLECTION

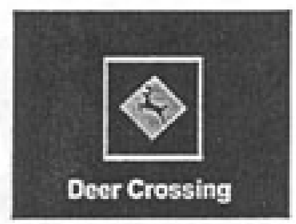


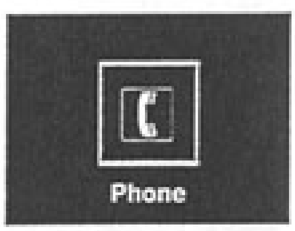


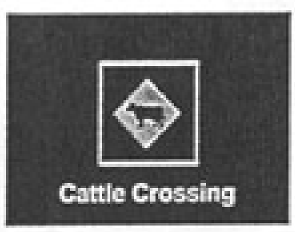


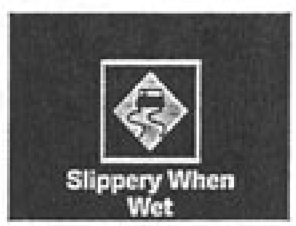


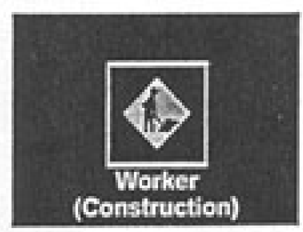


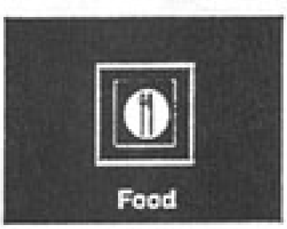


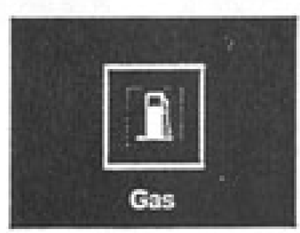




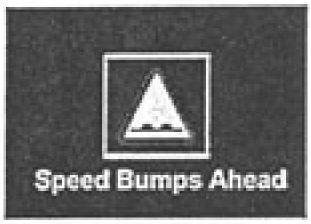


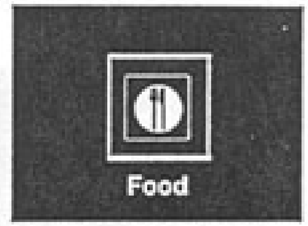


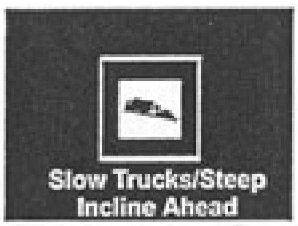


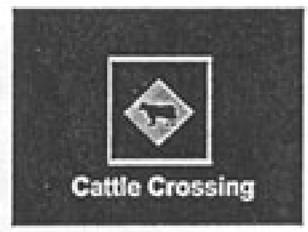


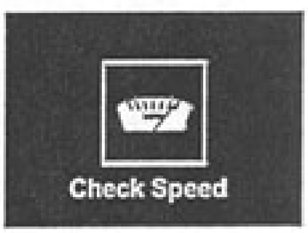
EXPERIMENT 5A POST TEST.

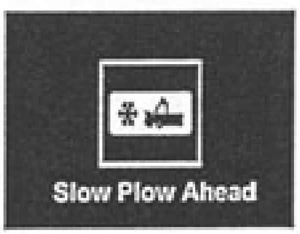


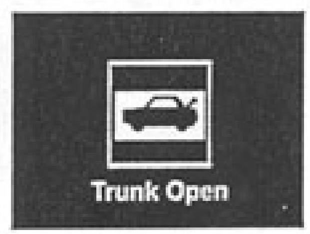


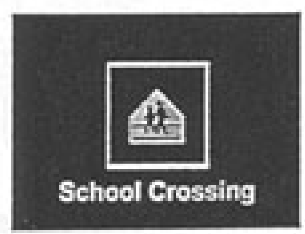




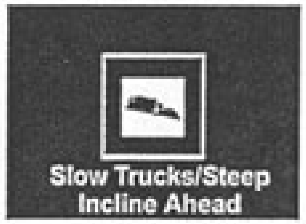


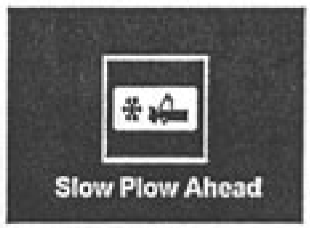


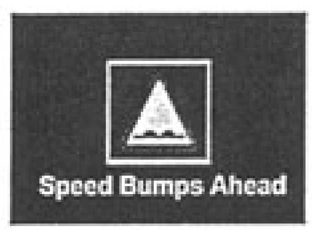




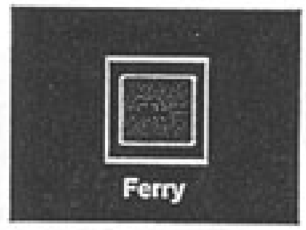
EXPERIMENT 5A UNFAMILIAR POST TEST COMPREHENSION.

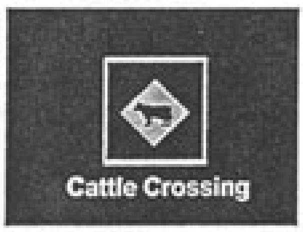






EXPERIMENT 5C DATA COLLECTION.





APPENDIX C. RESULTS ANOVA TABLES

EXPERIMENT 5A - RECOGNITION ACCURACY

Table 17. Analysis of variance for gender, age, repetition, delay, and symbol type.

SOURCE OF VARIATION	df	MS	F	p	GREENHOUSE-GEISSER
Gender	1	0.072	1.44	0.25	
Age	**1**	**0.27**	**5.39**	**0.036**	
Gender x Age	1	0.027	0.54	0.475	
Error	14	0.05			
Repetition	1	0.024	0.95	0.348	
Repetition x Gender	1	0.005	0.2	0.665	
Repetition x Age	1	0.005	0.21	0.654	
Repetition x Gender x Age	1	0.002	0.09	0.767	
Error	14	0.025			
Delay	**1**	**0.199**	**30.72**	**0.0001**	
Delay x Gender	1	0.017	2.6	0.129	
Delay x Age	**1**	**0.071**	**10.97**	**0.005**	
Delay x Gender x Age	1	0.027	4.1	0.062	
Error	14				
Repetition x Delay	1	0.004	0.29	0.599	
Repetition x Delay x Gender	1	0.017	1.33	0.268	
Repetition x Delay x Age		0.006	0.49	0.495	
Repetition x Delay x Gender x Age	1	0.018	1.38	0.26	
Error	14	0.013			
Symbol Type	**5**	**0.099**	**3.24**	**0.011**	**0.05**
Symbol Type x Gender	5	0.012	0.39	0.853	0.693
Symbol Type x Age	5	0.035	1.14	0.346	0.335
Symbol Type x Gender x Age	5	0.027	0.9	0.488	0.425
Error	70	0.031			

Table 17. **Analysis of variance for gender, age, repetition, delay, and symbol type (continued).**

SOURCE OF VARIATION	df	MS	F	p	GREENHOUSE-GEISSER
Repetition x Symbol Type	5	0.035	2.47	0.04	0.084
Repetition x Symbol Type x Gender	5	0.022	1.53	0.192	0.227
Repetition x Symbol Type x Age	5	0.005	0.33	0.892	0.775
Repetition x Symbol Type x Gender x Age	5	0.007	0.52	0.764	0.65
Error	70	0.014			
Delay x Symbol Type	5	0.012	0.97	0.443	0.405
Delay x Symbol Type x Gender	5	0.011	0.91	0.477	0.428
Delay x Symbol Type x Age	5	0.011	0.92	0.471	0.424
Delay x Symbol Type x Gender x Age	5	0.001	0.1	0.992	0.939
Error	70	0.012			
Repetition x Delay x Symbol Type	**5**	**0.137**	**5.92**	**0.0001**	**0.003**
Repetition x Delay x Symbol Type x Gender	5	0.065	2.82	0.022	0.057
Repetition x Delay x Symbol Type x Age	5	0.011	0.46	0.804	0.69
Repetition x Delay x Symbol Type x Gender x Age	5	0.001	0.06	0.998	0.974
Error	70	0.023			

EXPERIMENT 5A - RECOGNITION LATENCY

Table 18. Analysis of variance for gender, age, repetition, delay, and symbol type.

SOURCE OF VARIATION	df	MS	F	p	GREENHOUSE-GEISSER
Gender	1	29.89	1.82	0.198	
Age	1	0.38	0.02	0.882	
Gender x Age	1	12.61	0.77	0.395	
Error	14	16.38			
Repetition	**1**	**11.55**	**5.19**	**0.039**	
Repetition x Gender	1	6.37	2.86	0.112	
Repetition x Age	1	1.34	0.60	0.451	
Repetition x Gender x Age	1	1.44	0.65	0.435	
Error	14	2.23			
Delay	**1**	**113.87**	**77.28**	**0.001**	
Delay x Gender	1	0.43	0.29	0.599	
Delay x Age	1	0.01	0.01	0.943	
Delay x Gender x Age	1	0.02	0.01	0.915	
Error	14	1.47			
Repetition x Delay	1	1.33	1.16	0.300	
Repetition x Delay x Gender	1	0.14	0.01	0.912	
Repetition x Delay x Age	1	0.14	0.12	0.736	
Repetition x Delay x Gender x Age	1	0.47	0.41	0.532	
Error	14	1.15			
Symbol Type	**5**	**8.97**	**10.67**	**0.001**	**0.001**
Symbol Type x Gender	5	0.72	0.86	0.513	0.454
Symbol Type x Age	**5**	**3.05**	**3.63**	**0.006**	**0.029**
Symbol Type x Gender x Age	5	0.74	0.88	0.500	0.444
Error	70	0.84			

Table 18. Analysis of variance for gender, age, repetition, delay, and symbol type (continued).

SOURCE OF VARIATION	df	MS	F	p	GREENHOUSE-GEISSER
Repetition x Symbol Type	**5**	**3.30**	**4.38**	**0.002**	**0.016**
Repetition x Symbol Type x Gender	5	0.158	0.21	0.958	0.847
Repetition x Symbol Type x Age	5	0.823	1.09	0.374	0.357
Repetition x Symbol Type x Gender x Age	5	0.714	0.94	0.457	0.412
Error	70	0.755			
Delay x Symbol Type	5	0.10	0.13	0.984	0.933
Delay x Symbol Type x Gender	5	0.36	0.47	0.795	0.694
Delay x Symbol Type x Age	5	0.77	1.01	0.417	0.395
Delay x Symbol Type x Gender x Age	5	0.33	0.43	0.824	0.722
Error	70	0.76			
Repetition x Delay x Symbol Type	**5**	**6.61**	**8.25**	**0.001**	**0.001**
Repetition x Delay x Symbol Type x Gender	5	0.91	1.13	0.351	0.346
Repetition x Delay x Symbol Type x Age	5	1.06	1.33	0.263	0.378
Repetition x Delay x Symbol Type x Gender x Age	5	1.78	2.33	0.061	0.101
Error	70	0.80			

EXPERIMENT 5A - RECOGNITION CONFIDENCE

Table 19. Analysis of variance for gender, age, repetition, delay, and symbol type.

SOURCE OF VARIATION	df	MS	F	p	GREENHOUSE-GEISSER
Gender	1	158.8	0.08	0.777	
Age	1	3498.6	1.84	0.196	
Gender x Age	1	559.2	0.29	0.596	
Error	14	1897.9			
Repetition	1	314.1	1.18	0.296	
Repetition x Gender	1	14.29	0.05	0.820	
Repetition x Age	1	0.60	0.00	0.963	
Repetition x Gender x Age	1	239.3	0.90	0.360	
Error	14	266.6			
Delay	**1**	**2350.5**	**8.26**	**0.012**	
Delay x Gender	1	3.91	0.01	0.901	
Delay x Age	1	6.00	0.02	0.887	
Delay x Gender x Age	1	370.8	1.30	0.273	
Error	14	284.6			
Repetition x Delay	**1**	**746.7**	**6.40**	**0.024**	
Repetition x Delay x Gender	1	85.3	0.73	0.407	
Repetition x Delay x Age	1	123.4	1.05	0.323	
Repetition x Delay x Gender x Age	1	37.4	0.32	0.581	
Error	14	116.7			
Symbol Type	**5**	**1686.1**	**6.11**	**0.001**	**0.002**
Symbol Type x Gender	5	312.0	1.13	0.352	0.345
Symbol Type x Age	5	161.7	0.58	0.710	0.609
Symbol Type x Gender x Age	5	224.8	0.82	0.543	0.481
Error	70	275.7			

Table 19. Analysis of variance for gender, age, repetition, delay, and symbol type (continued).

SOURCE OF VARIATION	df	MS	F	p	GREENHOUSE-GEISSER
Repetition x Symbol Type	5	460.7	2.78	0.024	0.062
Repetition x Symbol Type x Gender	5	256.7	1.55	0.185	0.222
Repetition x Symbol Type x Age	5	205.6	1.24	0.299	0.306
Repetition x Symbol Type x Gender x Age	5	326.4	1.97	0.094	0.143
Error	70	165.5			
Delay x Symbol Type	5	646.6	2.96	0.018	0.078
Delay x Symbol Type x Gender	5	227.3	1.04	0.400	0.358
Delay x Symbol Type x Age	5	99.1	0.45	0.809	0.610
Delay x Symbol Type x Gender x Age	5	75.3	0.35	0.884	0.678
Error	70	218.2			
Repetition x Delay x Symbol Type	5	213.0	1.89	0.107	0.149
Repetition x Delay x Symbol Type x Gender	5	10.72	0.18	0.968	0.899
Repetition x Delay x Symbol Type x Age	5	70.0	0.62	0.684	0.598
Repetition x Delay x Symbol Type x Gender x Age	5	28.55	0.25	0.937	0.849
Error	70	112.6			

EXPERIMENT 5A - RECOGNITION CONFIDENCE LATENCY

Table 20. Analysis of variance for gender, age, repetition, delay, and symbol type.

SOURCE OF VARIATION	df	MS	F	p	GREENHOUSE-GEISSER
Gender	1	16.7	2.59	0.13	
Age	**1**	**70.68**	**10.97**	**0.005**	
Gender x Age	1	8.42	1.31	0.272	
Error	14	6.44			
Repetition	**1**	**40.93**	**11.41**	**0.005**	
Repetition x Gender	1	11.59	3.23	0.094	
Repetition x Age	1	6.77	1.89	0.191	
Repetition x Gender x Age	1	2.23	0.62	0.444	
Error	14	3.59			
Delay	1	2.16	1.72	0.210	
Delay x Gender	1	2.63	2.10	0.170	
Delay x Age	1	1.90	1.51	0.239	
Delay x Gender x Age	1	2.56	2.05	0.174	
Error	14	1.25			
Repetition x Delay	1	1.70	1.06	0.321	
Repetition x Delay x Gender	1	1.94	1.21	0.290	
Repetition x Delay x Age	1	1.91	1.19	0.294	
Repetition x Delay x Gender x Age	1	2.12	1.32	0.269	
Error	14	1.60			
Symbol Type	5	2.95	1.88	0.108	0.177
Symbol Type x Gender	5	0.85	0.54	0.747	0.568
Symbol Type x Age	5	2.40	1.53	0.193	0.237
Symbol Type x Gender x Age	5	0.79	0.51	0.771	0.586
Error	70	1.57			

Table 20. Analysis of variance for gender, age, repetition, delay, and symbol type (continued).

SOURCE OF VARIATION	df	MS	F	p	GREENHOUSE-GEISSER
Repetition x Symbol Type	5	1.88	1.06	0.389	0.341
Repetition x Symbol Type x Gender	5	2.03	1.14	0.346	0.319
Repetition x Symbol Type x Age	5	2.46	1.38	0.241	0.266
Repetition x Symbol Type x Gender x Age	5	0.81	0.46	0.807	0.571
Error	70	1.78			
Delay x Symbol Type	5	2.26	1.74	0.137	0.203
Delay x Symbol Type x Gender	5	1.03	0.79	0.560	0.432
Delay x Symbol Type x Age	5	1.22	0.94	0.460	0.381
Delay x Symbol Type x Gender x Age	5	1.56	1.20	0.316	0.307
Error	70	1.30			
Repetition x Delay x Symbol Type	5	1.46	1.23	0.305	0.303
Repetition x Delay x Symbol Type x Gender	5	1.18	0.99	0.430	0.370
Repetition x Delay x Symbol Type x Age	5	1.24	1.04	0.399	0.353
Repetition x Delay x Symbol Type x Gender x Age	5	1.17	0.99	0.433	0.371
Error	70	1.19			

EXPERIMENT 5A - MEAN LANE POSITION

Table 21. Analysis of variance for gender, age, repetition, message type and PDA.

SOURCE OF VARIATION	df	MS	F	p	GREENHOUSE-GEISSER
Gender	1	58.4	2.21	0.159	
Age	1	12.2	0.46	0.507	
Gender x Age	1	22.3	0.85	0.373	
Error	14	26.4			
Repetition	**1**	**85.8**	**27.0**	**0.001**	
Repetition x Gender	1	0.28	0.09	0.771	
Repetition x Age	1	6.60	2.07	0.172	
Repetition x Gender x Age	1	0.16	0.05	0.828	
Error	14	3.19			
Message Type	5	2.45	2.25	0.059	0.101
Message Type x Gender	5	0.44	0.39	0.855	0.749
Message Type x Age	5	0.78	0.69	0.930	0.553
Message Type x Gender x Age	5	0.56	0.50	0.778	0.675
Error	70	1.13			
Repetition x Message Type	5	2.08	1.62	0.167	0.191
Repetition x Message Type x Gender	5	0.98	0.76	0.581	0.540
Repetition x Message Type x Age	5	0.72	0.56	0.731	0.670
Repetition x Message Type x Gender x Age	**5**	**3.50**	**2.72**	**0.027**	**0.047**
Error	70	1.29			
PDA	2	1.72	3.37	0.049	0.058
PDA x Gender	2	1.38	2.70	0.085	0.095
PDA x Age	2	0.22	0.43	0.652	0.622
PDA x Gender x Age	2	0.64	1.25	0.301	0.299
Error	28	0.51			

Table 21. Analysis of variance for gender, age, repetition, message type, and PDA (continued).

SOURCE OF VARIATION	df	MS	F	p	GREENHOUSE-GEISSER
Repetition x PDA	2	1.25	1.25	0.302	0.300
Repetition x PDA x Gender	2	0.05	0.04	0.956	0.942
Repetition x PDA x Age	2	0.36	0.36	0.702	0.679
Repetition x PDA x Gender x Age	2	0.07	0.07	0.934	0.917
Error	28	1.00			
Message Type x PDA	**10**	**1.50**	**3.27**	**0.001**	**0.001**
Message Type x PDA x Gender	10	0.37	0.80	0.627	0.555
Message Type x PDA x Age	10	0.36	0.79	0.641	0.566
Message Type x PDA x Gender x Age	10	0.69	1.51	0.143	0.197
Error	140	0.46			
Repetition x Message Type x PDA	10	0.72	1.30	0.236	0.276
Repetition x Message Type x PDA x Gender	10	0.13	0.23	0.993	0.943
Repetition x Message Type x PDA x Age	10	0.53	0.96	0.479	0.444
Repetition x Message Type x PDA x Gender x Age	10	0.41	0.75	0.676	0.582
Error	140	0.55			

EXPERIMENT 5A - STANDARD DEVIATION LANE POSITION

Table 22. Analysis of variance for gender, age, repetition, message type and PDA.

SOURCE OF VARIATION	df	MS	F	p	GREENHOUSE-GEISSER
Gender	1	0.53	0.14	0.71	
Age	1	9.30	2.46	0.14	
Gender x Age	1	0.78	0.21	0.66	
Error	14	3.79			
Repetition	1	1.45	3.34	0.089	
Repetition x Gender	1	0.05	0.11	0.750	
Repetition x Age	1	0.08	0.19	0.671	
Repetition x Gender x Age	1	0.36	0.83	0.379	
Error	14	0.44			
Message Type	5	0.51	2.34	0.050	0.089
Message Type x Gender	5	0.37	1.70	0.147	0.184
Message Type x Age	5	0.18	0.84	0.528	0.479
Message Type x Gender x Age	5	0.25	1.12	0.358	0.352
Error	70	0.22			
Repetition x Message Type	5	0.37	2.38	0.047	0.071
Repetition x Message Type x Gender	5	0.05	0.30	0.908	0.853
Repetition x Message Type x Age	5	0.04	0.25	0.938	0.890
Repetition x Message Type x Gender x Age	5	0.11	0.70	0.629	0.582
Error	70	0.16			
PDA	2	0.29	1.44	0.255	0.256
PDA x Gender	2	0.21	1.06	0.359	0.349
PDA x Age	2	0.27	1.36	0.272	0.272
PDA x Gender x Age	2	0.01	0.04	0.958	0.933
Error	28	0.20			

Table 22. Analysis of variance for gender, age, repetition, message type and PDA (continued).

SOURCE OF VARIATION	df	MS	F	p	GREENHOUSE-GEISSER
Repetition x PDA	2	0.19	1.85	0.175	0.179
Repetition x PDA x Gender	2	0.03	0.24	0.788	0.769
Repetition x PDA x Age	2	0.04	0.36	0.699	0.681
Repetition x PDA x Gender x Age	2	0.04	0.39	0.679	0.662
Error	28	0.10			
Message Type x PDA	10	0.21	1.52	0.139	0.204
Message Type x PDA x Gender	10	0.08	0.56	0.842	0.705
Message Type x PDA x Age	10	0.24	1.76	0.074	0.144
Message Type x PDA x Gender x Age	10	0.04	0.31	0.976	0.881
Error	140	0.14			
Repetition x Message Type x PDA	10	0.13	1.57	0.120	0.163
Repetition x Message Type x PDA x Gender	10	0.06	0.73	0.692	0.627
Repetition x Message Type x PDA x Age	10	0.07	0.89	0.541	0.525
Repetition x Message Type x PDA x Gender x Age	10	0.06	0.69	0.737	0.666
Error	140	0.08			

EXPERIMENT 5A - CRASHES

Table 23. Analysis of variance for gender, age, repetition, message type and PDA.

SOURCE OF VARIATION	df	MS	F	p	GREENHOUSE-GEISSER
Gender	1	0.09	0.01	0.964	
Age	**1**	**0.27**	**6.66**	**0.022**	
Gender x Age	1	0.01	0.03	0.860	
Error	14	0.04			
Repetition	1	0.11	2.71	0.122	
Repetition x Gender	1	0.01	0.02	0.882	
Repetition x Age	1	0.06	1.62	0.224	
Repetition x Gender x Age	1	0.09	0.05	0.833	
Error	14	0.04			
Message Type	5	0.03	2.51	0.038	0.102
Message Type x Gender	5	0.01	1.00	0.424	0.377
Message Type x Age	5	0.02	1.56	0.184	0.230
Message Type x Gender x Age	5	0.01	0.78	0.570	0.464
Error	70	0.01			
Repetition x Message Type	5	0.01	0.34	0.890	0.692
Repetition x Message Type x Gender	5	0.01	0.71	0.617	0.484
Repetition x Message Type x Age	5	0.01	0.48	0.788	0.600
Repetition x Message Type x Gender x Age	5	0.01	0.37	0.869	0.671
Error	70	0.01			
PDA	2	0.01	1.16	0.328	0.324
PDA x Gender	2	0.01	1.06	0.361	0.355
PDA x Age	2	0.02	2.22	0.128	0.135
PDA x Gender x Age	2	0.01	1.20	0.317	0.314
Error	28	0.01			

Table 23. Analysis of variance for gender, age, repetition, message type and PDA (continued).

SOURCE OF VARIATION	df	MS	F	p	GREENHOUSE-GEISSER
Repetition x PDA	2	0.02	1.01	0.378	0.367
Repetition x PDA x Gender	2	0.01	0.65	0.530	0.505
Repetition x PDA x Age	2	0.01	0.69	0.509	0.486
Repetition x PDA x Gender x Age	2	0.01	0.37	0.693	0.657
Error	28	0.02			
Message Type x PDA	10	0.02	1.49	0.151	0.239
Message Type x PDA x Gender	10	0.01	0.42	0.934	0.701
Message Type x PDA x Age	10	0.02	1.51	0.141	0.233
Message Type x PDA x Gender x Age	10	0.01	0.84	0.589	0.461
Error	140	0.01			
Repetition x Message Type x PDA	**10**	**0.03**	**3049**	**0.001**	**0.021**
Repetition x Message Type x PDA x Gender	10	0.01	0.46	0.912	0.721
Repetition x Message Type x PDA x Age	**10**	**0.03**	**3.20**	**0.001**	**0.030**
Repetition x Message Type x PDA x Gender x Age	10	0.01	0.73	0.699	0.459
Error	140	0.01			

EXPERIMENT 5A: SYMBOL COMPREHENSION POST-TEST: ACCURACY

Table 24. Analysis of variance for age (2) and symbols (20).*

SOURCE	df	F	p
Age	**1, 84**	**14.84**	**0.0002**
Symbol	19, 84	1.71	0.051
Age x Symbol	19, 84	1.42	0.142

Table 25. Analysis of variance for age (2) and symbol group (5).*

SOURCE	df	F	p
Age	**1, 16**	**5.9**	**0.027**
Symbol	4, 15	1.36	0.296
Age x Symbol	4, 15	1.23	0.339

Table 26. Analysis of variance for age (2) and symbol group (6).*

SOURCE	df	F	p
Age	**1, 21**	**8.8**	**0.007**
Symbol	5, 20	1.21	0.342
Age x Symbol	5, 20	1.11	0.385

*Variances not assumed to be equal; Brown-Forsythe used.

EXPERIMENT 5A: SYMBOL COMPREHENSION POST-TEST: LATENCY

Table 27. Analysis of variance for age (2) and symbols (20).*

SOURCE	df	F	p
Age	**1, 95**	**10.91**	**0.0013**
Symbol	**19, 95**	**5.07**	**0.00001**
Age x Symbol	19, 95	0.94	0.5357

Table 28. Analysis of variance for age (2) and symbol group (5).*

SOURCE	df	F	p
Age	**1, 17**	**5.03**	**0.039**
Symbol	4, 17	0.80	0.539
Age x Symbol	4, 17	2.26	0.105

Table 29. Analysis of variance for age (2) and symbol group (6).*

SOURCE	df	F	p
Age	**1, 19**	**6.45**	**0.02**
Symbol	5, 18	0.79	0.57
Age x Symbol	5, 18	2.09	0.113

*Variances not assumed to be equal; Brown-Forsythe used.

EXPERIMENT 5B: RECOGNITION ACCURACY

Table 30. Analysis of variance for gender, message type, delay, and repetition.

SOURCE OF VARIATION	df	MS	F	p
Gender	1	0.00036	0.26	0.624
Error	10	0.0014		
Message	1	9.00	2.30	0.160
Message x Gender	1	0.00155	1.10	0.318
Error	10	0.0014		
Delay	1	0.0001	0.14	0.716
Delay x Gender	1	0.00009	0.11	0.742
Error	10	0.00074		
Message x Delay	1	0.0001	0.09	0.769
Message x Delay x Gender	1	0.00088	0.77	0.400
Error	10	0.00114		
Repetition	1	0.00082	2.60	0.138
Repetition x Gender	1	0.0001	0.33	0.579
Error	10	0.00032		
Message x Repetition	1	0.00241	2.17	0.171
Message x Repetition x Gender	1	0.0001	0.09	0.766
Error	10	0.00111		
Delay x Repetition	1	0.00036	0.49	0.499
Delay x Repetition x Gender	1	0.00357	4.90	0.051
Error	10	0.00073		
Message x Delay x Repetition	1	0.00155	3.15	0.106
Message x Delay x Repetition x Gender	1	0	0	0.949
Error	10	0.00049		

EXPERIMENT 5B - RECOGNITION ACCURACY

Table 31. Analysis of variance for gender, earcon, delay, and repetition.

SOURCE OF VARIATION	df	MS	F	p	GREENHOUSE-GEISSER
Gender	1	0.04	0.75	0.407	
Error	10	0.05			
Earcon	20	0.03	1.52	0.076	0.220
Earcon x gender	20	0.01	0.89	0.605	0.472
Error	200	0.17			
Delay	1	0	0	1.000	
Delay x gender	1	0.02	0.65	0.441	
Error	10	0.02			
Earcon x delay	20	0.02	1.02	0.435	0.407
Earcon x delay x gender	20	0.01	0.46	0.977	0.765
Error	200	0.02			
Repetition	1	0.06	4.21	0.067	
Repetition x gender	1	0	0	1.000	
Error	10	0.02			
Earcon x repetition	20	0.03	1.29	0.186	0.294
Earcon x repetition x gender	20	0.01	0.62	0.893	0.609
Error	200	0.02			
Delay x repetition	1	0.04	2.50	0.145	
Delay x repetition x gender	1	0.04	2.50	0.145	
Error	10	0.01			
Earcon x delay x repetition	20	0.03	1.81	0.022	0.138
Earcon x delay x repetition x gender	20	0.01	0.71	0.815	0.603
Error	200	0.02			

EXPERIMENT 5B: RECOGNITION LATENCY

Table 32. Analysis of variance for gender, message type, delay, and repetition.

SOURCE OF VARIATION	df	MS	F	p
Gender	1	2.073	1.55	0.242
Error	10	1.34		
Message	**1**	**2.03**	**14.13**	**0.004**
Message x Gender	1	0.17	1.21	0.297
Error	10	0.14		
Delay	**1**	**20.33**	**121.27**	**0.001**
Delay x Gender	1	0.39	2.34	0.157
Error	10	0.17		
Message x Delay	1	0.00025	0	0.967
Message x Delay x Gender	1	0.00007	0	0.983
Error	10	0.14		
Repetition	**1**	**0.79**	**18.34**	**0.002**
Repetition x Gender	1	0.00002	0	0.985
Error	10	0.04		
Message x Repetition	1	0.07	1.70	0.222
Message x Repetition x Gender	**1**	**0.21**	**5.19**	**0.046**
Error	10	0.04		
Delay x Repetition	1	0.01	0.11	0.747
Delay x Repetition x Gender	1	0.03	0.56	0.473
Error	10	0.06		
Message x Delay x Repetition	1	0.14	4.50	0.060
Message x Delay x Repetition x Gender	1	0.04	1.15	0.309
Error	10	0.03		

EXPERIMENT 5B: RECOGNITION LATENCY

Table 33. Analysis of variance for gender, earcon, delay, and repetition.

SOURCE OF VARIATION	df	MS	F	p	GREENHOUSE-GEISSER
Gender	1	10.98	0.78	0.399	
Error	10	14.14			
Earcon	**20**	**5.54**	**6.17**	**0.000**	**0.001**
Earcon x gender	20	1.08	1.21	0.252	0.317
Error	200	0.90			
Delay	**1**	**212.01**	**69.97**	**0.000**	
Delay x gender	1	4.23	1.4	0.265	
Error	10	3.03			
Earcon x delay	20	1.16	1.19	0.268	0.3262
Earcon x delay x gender	20	0.91	0.93	0.544	0.474
Error	200	0.97			
Repetition	**1**	**13.89**	**21.36**	**0.001**	
Repetition x gender	1	2.12	3.26	0.101	
Error	10	0.65			
Earcon x repetition	20	1.20	1.32	0.172	0.270
Earcon x repetition x gender	20	0.70	0.77	0.748	0.581
Error	200	0.91			
Delay x repetition	1	0.90	0.88	0.369	
Delay x repetition x gender	1	1.49	1.46	0.254	
Error	10	1.02			
Earcon x delay x repetition	20	1.75	2.23	0.003	0.060
Earcon x delay x repetition x gender	20	0.72	0.92	0.568	0.484
Error	200	0.79			

EXPERIMENT 5B: RECOGNITION CONFIDENCE

Table 34. Analysis of variance for gender, message type, delay, and repetition.

SOURCE OF VARIATION	df	MS	F	p
Gender	1	159.54	2.01	0.186
Error	10	79.20		
Message	1	0.90	0.14	0.711
Message x Gender	1	0.38	0.06	0.809
Error	10	6.19		
Delay	1	84.54	3.52	0.090
Delay x Gender	1	0.22	0.01	0.925
Error	10	24.01		
Message x Delay	1	0.34	0.02	0.900
Message x Delay x Gender	1	0.28	0.01	0.909
Error	10	20.24		
Repetition	1	0.16	0.02	0.882
Repetition x Gender	1	19.62	2.87	0.121
Error	10	6.84		
Message x Repetition	1	1.68	0.10	0.761
Message x Repetition x Gender	1	5.61	0.33	0.581
Error	10	17.19		
Delay x Repetition	1	7.77428	1.35	0.272
Delay x Repetition x Gender	1	0.93634	0.16	0.695
Error	10	5.74497		
Message x Delay x Repetition	1	10.61277	2.19	0.170
Message x Delay x Repetition x Gender	1	0.11213	0.02	0.882
Error	10	4.8464		

EXPERIMENT 5B: RECOGNITION CONFIDENCE

Table 35. Analysis of variance for gender, earcon, delay, and repetition.

SOURCE OF VARIATION	df	MS	F	p	GREENHOUSE-GEISSER
Gender	1	2578.56	3.08	0.118	
Error	8	837.87			
Earcon	20	158.82	1.11	0.344	0.369
Earcon x gender	20	130.97	0.91	0.569	0.466
Error	160	143.16			
Delay	1	108.36	1.99	0.196	
Delay x gender	1	206.83	3.79	0.087	
Error	8	54.51			
Earcon x delay	20	151.62	1.25	0.220	0.309
Earcon x delay x gender	20	111.38	0.92	0.563	0.465
Error	160	121.15			
Repetition	1	3.62	0.02	0.887	
Repetition x gender	1	2.19	0.01	0.912	
Error	8	168.74			
Earcon x repetition	20	171.34	1.43	0.115	0.249
Earcon x repetition x gender	20	121.71	1.02	0.446	0.412
Error	160	119.76			
Delay x repetition	1	127.30	2.09	0.187	
Delay x repetition x gender	1	100.30	1.64	0.236	
Error	8	61.02			
Earcon x delay x repetition	20	127.63	1.28	0.203	0.304
Earcon x delay x repetition x gender	20	74.30	0.74	0.777	0.548
Error	160	100.07			

EXPERIMENT 5C: RECOGNITION ACCURACY

Table 36. Analysis of variance for alert, scenario, message modality, and message format.

SOURCE OF VARIATION	df	MS	F	p
Alert	2	0.01	1.25	0.299
Error	33	0.01		
Scenario	1	0.0001	0.03	0.862
Scenario x Alert	2	0.01	2.79	0.076
Error	33	0.00315		
Message Modality	**1**	**0.02**	**15.18**	**0.0005**
Message Modality x Alert	**2**	**0.01**	**3.85**	**0.032**
Error	33	0.00143		
Scenario x Message Modality	1	0.00087	0.38	0.541
Scenario x Message Modality x Alert	2	0.00203	0.09	0.420
Error	33	0.00227		
Message Format	**1**	**0.01**	**4.12**	**0.050**
Message Format x Alert	2	0.00125	0.44	0.646
Error	33	0.00283		
Scenario x Message Format	1	0.00241	0.40	0.530
Scenario x Message Format x Alert	2	0.00936	1.56	0.225
Error	33	0.00599		
Message Modality x Message Format	1	0.00087	0.26	0.611
Message Modality x Message Format x Alert	2	0.00608	1.85	0.174
Error	33	0.00329		
Scenario X Message Modality x Message Format	1	0.00241	0.63	0.434
Scenario X Message Modality x Message Format x Alert	2	0.00993	2.58	0.091
Error	33	0.00385		

EXPERIMENT 5C: RECOGNITION ACCURACY

Table 37. Analysis of variance for alert, scenario, message modality, and delay.

SOURCE OF VARIATION	df	MS	F	p	GREENHOUSE-GEISSER
Alert	2	0.02	1.71	0.196	
Error	33	0.01			
Scenario	1	0.0041	1.20	0.281	
Scenario x Alert	2	0.01	1.82	0.178	
Error	33	0.0034			
Message Modality	**2**	**0.01**	**5.18**	**0.008**	**0.014**
Message Modality x Alert	4	0.0045	1.62	0.179	0.192
Error	66	0.0028			
Scenario x Message Modality	2	0.0032	1.11	0.335	0.333
Scenario x Message Modality x Alert	4	0.0027	0.94	0.445	0.442
Error	66	0.0028			
Delay	**1**	**0.17**	**21.97**	**0.001**	
Delay x Alert	2	0.02	2.35	0.111	
Error	33	0.01			
Scenario x Delay	**1**	**0.03**	**6.61**	**0.015**	
Scenario x Delay x Alert	2	0.01	1.60	0.217	
Error	33	0.0039			
Message Modality x Delay	**2**	**0.02**	**5.95**	**0.004**	**0.006**
Message Modality x Delay x Alert	4	0.0042	1.14	0.347	0.346
Error	66	0.0037			
Scenario x Message Modality x Delay	2	0.0012	0.40	0.675	0.673
Scenario x Message Modality x Delay x Alert	4	0.0040	1.30	0.278	0.279
Error	66	0.0031			

EXPERIMENT 5C: RECOGNITION ACCURACY

Table 38. Analysis of variance for alert, scenario, message modality, format, and delay.

SOURCE OF VARIATION	df	MS	F	p
Alert	2	0.02	0.83	0.443
Error	33	0.03		
Scenario	1	0.00043	0.02	0.876
Scenario x Alert	2	0.02	1.39	0.264
Error	33	0.02		
Message Modality	**1**	**0.08**	**9.67**	**0.004**
Message Modality x Alert	**2**	**0.03**	**4.08**	**0.026**
Error	33	0.01		
Scenario x Message Modality	1	0.00391	0.35	0.557
Scenario x Message Modality x Alert	2	0.01	0.96	0.394
Error	33	0.01		
Format	1	0.00444	0.29	0.592
Format x Alert	2	0.01	0.67	0.517
Error	33	0.02		
Scenario x Format	1	0.03	1.14	0.294
Scenario x Format x Alert	2	0.01	0.33	0.721
Error	33	0.02		
Message Modality x Format	1	0.00028	0.02	0.888
Message Modality x Format x Alert	2	0.03	2.23	0.124
Error	33	0.01		
Scenario X Message Modality x Format	1	0.01	0.5	0.485
Scenario X Message Modality x Format x Alert	2	0.02	1.67	0.204
Error	33	0.00139		
Delay	**1**	**0.17**	**6.87**	**0.013**
Delay x Alert	2	0.01	0.34	0.714
Error	33	0.03		

Table 38. Analysis of variance for alert, scenario, modality, format, and delay (continued).

SOURCE OF VARIATION	df	MS	F	p
Scenario x Delay	1	0.02	1.30	0.262
Scenario x Delay x Alert	2	0.04	2.39	0.108
Error	33	0.02		
Message Modality x Delay	1	0.03	2.44	0.128
Message Modality x Delay x Alert	2	0.02	1.44	0.252
Error	33	0.01		
Scenario x Message Modality x Delay	1	0.00444	0.35	0.557
Scenario x Message Modality x Delay x Alert	2	0.02	1.72	0.195
Error	33	0.01		
Format x Delay	1	0.00391	0.23	0.633
Format x Delay x Alert	2	0.01	0.63	0.537
Error	33	0.02		
Scenario x Format x Delay	1	0.01	0.41	0.525
Scenario x Format x Delay x Alert	2	0.00048	0.03	0.974
Error	33	0.02		
Message Modality x Format x Delay	1	0.00043	0.03	0.868
Message Modality x Format x Delay x Alert	2	0.02	1.04	0.365
Error	33	0.02		
Scenario x Message Modality x Format x Delay	1	0.02	1.45	0.237
Scenario x Message Modality x Format x Delay x Alert	2	0.03	2.78	0.077
Error	33	0.01		

EXPERIMENT 5C: RECOGNITION LATENCY

Table 39. Analysis of variance for alert, scenario, message modality, and delay.

SOURCE OF VARIATION	df	MS	F	p	GREENHOUSE-GEISSER
Alert	2	0.34	0.09	0.918	
Error	33	4.02			
Scenario	**1**	**5.42**	**19.38**	**0.001**	
Scenario x Alert	2	0.15	0.54	0.590	
Error	33	0.28			
Message Modality	2	0.12	0.94	0.397	0.390
Message Modality x Alert	4	0.33	2.51	0.050	0.056
Error	66	0.13			
Scenario x Message Modality	2	0.18	1.35	0.268	0.266
Scenario x Message Modality x Alert	4	0.06	0.45	0.775	0.729
Error	66	0.13			
Delay	**1**	**47.74**	**45.07**	**0.001**	
Delay x Alert	2	1.54	1.45	0.249	
Error	33	1.06			
Scenario x Delay	**1**	**0.56**	**4.09**	**0.051**	
Scenario x Delay x Alert	2	0.07	0.52	0.598	
Error	33	0.00389			
Message Modality x Delay	2	0.03	0.17	0.847	0.842
Message Modality x Delay x Alert	4	0.18	1.14	0.347	0.347
Error	66	0.16			
Scenario x Message Modality x Delay	**2**	**0.48**	**3.54**	**0.035**	**0.039**
Scenario x Message Modality x Delay x Alert	4	0.04	0.27	0.896	0.883
Error	66	0.14			

EXPERIMENT 5C: RECOGNITION LATENCY

Table 40. Analysis of variance for alert, scenario, message modality, message format, and delay.

SOURCE OF VARIATION	df	MS	F	p
Alert	2	1.03	0.19	0.832
Error	33	5.55		
Scenario	**1**	**8.56**	**14.31**	**0.0006**
Scenario x Alert	2	0.09	0.15	0.862
Error	33	0.60		
Message Modality	1	0.80	1.06	0.311
Message Modality x Alert	2	1.16	1.55	0.227
Error	33	0.75		
Scenario x Message Modality	1	0.12	0.23	0.632
Scenario x Message Modality x Alert	2	0.59	1.18	0.319
Error	33	0.50		
Message Format	1	0.71	1.54	0.224
Message Format x Alert	2	0.28	0.61	0.551
Error	33	0.46		
Scenario x Message Format	1	0.45	1.81	0.188
Scenario x Message Format x Alert	2	0.52	2.06	0.144
Error	33	0.25		
Message Modality x Message Format	1	0.41	1.13	0.296
Message Modality x Message Format x Alert	2	0.50	1.39	0.264
Error	33	0.36		
Scenario X Message Modality x Message Format	1	1.05	3.79	0.060
Scenario X Message Modality x Message Format x Alert	2	0.14	0.51	0.607
Error	33	0.28		
Delay	**1**	**69.17**	**44.75**	**0.000**
Delay x Alert	2	1.23	0.80	0.459
Error	33	1.55		

Table 40. Analysis of variance for alert, scenario, message modality, message format, and delay (continued).

SOURCE OF VARIATION	df	MS	F	p
Scenario x Delay	1	1.13	2.56	0.119
Scenario x Delay x Alert	2	0.13	0.30	0.741
Error	33	0.44		
Message Modality x Delay	1	0.60	1.26	0.270
Message Modality x Delay x Alert	2	0.94	1.97	0.155
Error	33	0.47		
Scenario x Message Modality x Delay	1	0.01	0.04	0.853
Scenario x Message Modality x Delay x Alert	2	0.19	0.49	0.617
Error	33	0.38		
Message Format x Delay	1	0.20	0.42	0.519
Message Format x Delay x Alert	2	0.26	0.57	0.573
Error	33	0.47		
Scenario x Message Format x Delay	1	0.59	1.04	0.316
Scenario x Message Format x Delay x Alert	2	0.48	0.84	0.439
Error	33	0.57		
Message Modality x Message Format x Delay	1	0.65	1.35	0.253
Message Modality x Message Format x Delay x Alert	2	0.47	0.98	0.387
Error	33	0.48		
Scenario x Message Modality x Message Format x Delay	1	0.53	1.28	0.267
Scenario x Message Modality x Message Format x Delay x Alert	2	0.31	0.74	0.486
Error	33	0.42		

EXPERIMENT 5C: RECOGNITION LATENCY

Table 41. Analysis of variance for alert, scenario, message modality, and message format.

SOURCE OF VARIATION	df	MS	F	p
Alert	2	0.21	0.08	0.926
Error	33	2.65		
Scenario	**1**	**3.30**	**15.3**	**0.0004**
Scenario x Alert	2	0.19	0.87	0.427
Error	33	0.22		
Message Modality	1	0.21	1.37	0.250
Message Modality x Alert	**2**	**0.55**	**3.5**	**0.042**
Error	33	0.16		
Scenario x Message Modality	1	0.34	1.65	0.208
Scenario x Message Modality x Alert	2	0.06	0.29	0.749
Error	33	0.20		
Message Format	**1**	**0.97**	**6.56**	**0.015**
Message Format x Alert	2	0.08	0.52	0.602
Error	33	0.15		
Scenario x Message Format	**1**	**6.53**	**20.51**	**0.0001**
Scenario x Message Format x Alert	2	0.84	2.63	0.09
Error	33	0.32		
Message Modality x Message Format	1	0.38	2.38	0.13
Message Modality x Message Format x Alert	2	0.01	0.61	0.55
Error	33	0.16		
Scenario X Message Modality x Message Format	1	0.28	2.28	0.14
Scenario X Message Modality x Message Format x Alert	2	0.06	0.50	0.61
Error	33	0.12		

EXPERIMENT 5C: RECOGNITION CONFIDENCE

Table 42. Analysis of variance for alert, scenario, message modality, and delay.

SOURCE OF VARIATION	df	MS	F	p	GREENHOUSE-GEISSER
Alert	2	871.90	2.89		0.697
Error	33	301.60			
Scenario	1	154.30	5.32	0.028	
Scenario x Alert	2	13.83	0.48	0.625	
Error	33	28.99			
Message Modality	**2**	**324.70**	**7.12**	**0.002**	**0.004**
Message Modality x Alert	**4**	**123.13**	**2.70**	**0.038**	**0.052**
Error	66	31.08			
Scenario x Message Modality	2	26.64	0.86	0.429	0.406
Scenario x Message Modality x Alert	4	83.86	2.70	0.038	0.053
Error	66	31.08			
Delay	**1**	**4267.14**	**24.93**	**0.001**	
Delay x Alert	2	268.34	1.57	0.224	
Error	33	171.19			
Scenario x Delay	**1**	**278.64**	**8.66**	**0.006**	
Scenario x Delay x Alert	2	42.83	1.33	0.278	
Error	33	32.16			
Message Modality x Delay	**2**	**165.32**	**5.16**	**0.008**	**0.008**
Message Modality x Delay x Alert	**4**	**137.38**	**4.29**	**0.004**	**0.004**
Error	66	32.02			
Scenario x Message Modality x Delay	2	2.95	0.06	0.941	0.935
Scenario x Message Modality x Delay x Alert	4	35.93	0.74	0.569	0.564
Error	66	48.65			

EXPERIMENT 5C: RECOGNITION CONFIDENCE

Table 43. Analysis of variance for alert, scenario, message modality, and message format.

SOURCE OF VARIATION	df	MS	F	p
Alert	2	666.98	3.02	0.062
Error	33	220.86		
Scenario	**1**	**182.43**	**6.06**	**0.019**
Scenario x Alert	2	15.34	0.51	0.605
Error	33	30.10		
Message Modality	**1**	**648.96**	**10.11**	**0.003**
Message Modality x Alert	2	188.63	2.94	0.067
Error	33	64.18		
Scenario x Message Modality	1	19.35	0.60	0.44
Scenario x Message Modality x Alert	2	76.91	2.38	0.11
Error	33	32.35		
Message Format	**1**	**191.62**	**6.21**	**0.018**
Message Format x Alert	**2**	**147.91**	**4.79**	**0.015**
Error	33	30.87		
Scenario x Message Format	**1**	**932.93**	**14.68**	**0.001**
Scenario x Message Format x Alert	2	171.87	2.70	0.082
Error	33	63.57		
Message Modality x Message Format	**1**	**183.02**	**6.02**	**0.020**
Message Modality x Message Format x Alert	2	41.15	1.35	0.272
Error	33	30.39		
Scenario X Message Modality x Message Format	1	91.37	2.31	0.138
Scenario X Message Modality x Message Format x Alert	2	68.68	1.73	0.192
Error	33	39.62		

EXPERIMENT 5C: RECOGNITION CONFIDENCE

Table 44. Analysis of variance for alert, scenario, message modality, message format, and delay.

SOURCE OF VARIATION	df	MS	F	p
Alert	2	742.36	1.59	0.220
Error	33	467.78		
Scenario	1	519.75	3.70	0.063
Scenario x Alert	2	56.51	0.40	0.672
Error	33	140.64		
Message Modality	**1**	**2026.46**	**10.18**	**0.003**
Message Modality x Alert	2	365.26	1.83	0.176
Error	33	199.06		
Scenario x Message Modality	1	0.216	0	0.971
Scenario x Message Modality x Alert	2	199.75	1.29	0.290
Error	33	155.17		
Message Format	1	355.46	2.85	0.101
Message Format x Alert	2	358.55	2.87	0.071
Error	33	124.89		
Scenario x Message Format	1	0.95	0.01	0.918
Scenario x Message Format x Alert	2	127.61	1.45	0.250
Error	33	88.09		
Message Modality x Message Format	**1**	**584.01**	**6.81**	**0.014**
Message Modality x Message Format x Alert	2	96.23	1.12	0.338
Error	33	85.72		
Scenario X Message Modality x Message Format	1	264.27	1.5	0.229
Scenario X Message Modality x Message Format x Alert	2	1.42	0.01	0.992
Error	33	176.05		
Delay	**1**	**5937.92**	**21.51**	**0.001**
Delay x Alert	2	302.01	1.09	0.347
Error	33	276.02		

Table 44. Analysis of variance for alert, scenario, message modality, message format, and delay (continued).

SOURCE OF VARIATION	df	MS	F	p
Scenario x Delay	1	29.18	0.2	0.659
Scenario x Delay x Alert	2	100.97	0.68	0.511
Error	33	147.53		
Message Modality x Delay	**1**	**1146.64**	**8.00**	**0.008**
Message Modality x Delay x Alert	2	448.84	3.13	0.057
Error	33	143.40		
Scenario x Message Modality x Delay	1	140.87	0.91	0.347
Scenario x Message Modality x Delay x Alert	2	3.65	0.02	0.978
Error	33	154.46		
Message Format x Delay	1	16.00	0.09	0.765
Message Format x Delay x Alert	2	361.57	2.05	0.145
Error	33	176.47		
Scenario x Message Format x Delay	1	375.63	2.52	0.122
Scenario x Message Format x Delay x Alert	2	166.41	1.12	0.340
Error	33	148.89		
Message Modality x Message Format x Delay	1	103.62	0.47	0.499
Message Modality x Message Format x Delay x Alert	2	7.21	0.03	0.968
Error	33	222.11		
Scenario x Message Modality x Message Format x Delay	1	132.68	0.75	0.394
Scenario x Message Modality x Message Format x Delay x Alert	2	507.14	2.86	0.072
Error	33	177.58		

EXPERIMENT 5C: MEAN SPEED

Table 45. Analysis of variance for alert, message modality, window, and delay.

SOURCE OF VARIATION	df	MS	F	p	GREENHOUSE-GEISSER
Alert	2	105.42	2.36	0.111	
Error	33	44.75			
Message Modality	**2**	**6.37**	**5.48**	**0.006**	**0.006**
Message Modality x Alert	4	0.73	0.63	0.644	0.643
Error	66	1.16			
Window	**2**	**3.60**	**12.44**	**0.000**	**0.0003**
Window x Alert	4	0.47	1.64	0.175	0.196
Error	66	0.29			
Message Modality x Window	**4**	**1.59**	**5.77**	**0.000**	**0.001**
Message Modality x Window x Alert	8	0.16	0.58	0.797	0.745
Error	132	0.28			
Delay	1	3.01	3.96	0.055	
Delay x Alert	2	0.65	0.86	0.432	
Error	33	0.76			
Message Modality x Delay	**2**	**6.67**	**10.30**	**0.001**	**0.001**
Message Modality x Delay x Alert	4	0.69	1.07	0.381	0.380
Error	66	0.65			
Window x Delay	**2**	**1.19**	**3.47**	**0.037**	**0.041**
Window x Delay x Alert	4	0.38	1.09	0.368	0.367
Error	66	0.34			
Message Modality x Window x Delay	**4**	**1.76**	**5.62**	**0.001**	**0.003**
Message Modality x Window x Delay x Alert	8	0.18	0.59	0.785	0.711
Error	132	0.31			

EXPERIMENT 5C: STANDARD DEVIATION FOR SPEED

Table 46. Analysis of variance for alert, message modality, window, and delay.

SOURCE OF VARIATION	df	MS	F	p	GREENHOUSE-GEISSER
Alert	2	0.37	0.73	0.490	
Error	33	0.51			
Message Modality	**2**	**0.66**	**28.90**	**0.001**	**0.001**
Message Modality x Alert	4	0.03	1.35	0.260	0.260
Error	66	0.02			
Window	**2**	**1.05**	**38.61**	**0.001**	**0.001**
Window x Alert	4	0.03	1.12	0.355	0.351
Error	66	0.03			
Message Modality x Window	**4**	**0.40**	**23.87**	**0.001**	**0.001**
Message Modality x Window x Alert	8	0.02	0.90	0.515	0.498
Error	132	0.02			
Delay	**1**	**0.11**	**6.14**	**0.019**	
Delay x Alert	2	0.02	0.90	0.415	
Error	33	0.02			
Message Modality x Delay	2	0.04	1.27	0.289	0.285
Message Modality x Delay x Alert	4	0.05	1.85	0.129	0.145
Error	66	0.03			
Window x Delay	**2**	**0.05**	**3.72**	**0.030**	**0.045**
Window x Delay x Alert	4	0.01	0.55	0.699	0.645
Error	66	0.01			
Message Modality x Window x Delay	4	0.02	0.74	0.568	0.498
Message Modality x Window x Delay x Alert	8	0.03	1.56	0.142	0.186
Error	132	0.02			

EXPERIMENT 5C: MEAN LANE POSITION

Table 47. Analysis of variance for alert, message modality, window, and delay.

SOURCE OF VARIATION	df	MS	F	p	GREENHOUSE-GEISSER
Alert	2	1.72	0.33	0.720	
Error	33	5.20			
Message Modality	**2**	**1.44**	**11.53**	**0.001**	**0.001**
Message Modality x Alert	4	0.08	0.68	0.610	0.603
Error	66	0.12			
Window	2	0.03	0.24	0.790	0.722
Window x Alert	4	0.10	0.73	0.574	0.537
Error	66	0.14			
Message Modality x Window	**4**	**0.26**	**4.02**	**0.004**	**0.005**
Message Modality x Window x Alert	8	0.07	1.00	0.439	0.4364
Error	132	0.07			
Delay	1	0.22	2.01	0.166	
Delay x Alert	2	0.09	0.81	0.454	
Error	33	0.11			
Message Modality x Delay	**2**	**0.88**	**4.15**	**0.020**	**0.025**
Message Modality x Delay x Alert	4	0.15	0.73	0.573	0.558
Error	66	0.21			
Window x Delay	2	0.04	0.75	0.475	0.473
Window x Delay x Alert	4	0.04	0.63	0.641	0.638
Error	66	0.06			
Message Modality x Window x Delay	**4**	**0.20**	**3.38**	**0.011**	**0.017**
Message Modality x Window x Delay x Alert	8	0.03	0.55	0.820	0.792
Error	132	0.06			

EXPERIMENT 5C: STANDARD DEVIATION FOR LANE POSITION

Table 48. Analysis of variance for alert, message modality, window, and delay.

SOURCE OF VARIATION	df	MS	F	p	GREENHOUSE-GEISSER
Alert	2	0.03	0.07	0.937	
Error	33	0.50			
Message Modality	**2**	**1.61**	**79.90**	**0.001**	**0.001**
Message Modality x Alert	4	0.03	1.34	0.266	0.273
Error	66	0.02			
Window	**2**	**1.82**	**82.76**	**0.001**	**0.001**
Window x Alert	4	0.04	1.67	0.168	0.175
Error	66	0.02			
Message Modality x Window	**4**	**1.35**	**106.59**	**0.001**	**0.001**
Message Modality x Window x Alert	8	0.004	0.29	0.968	0.949
Error	132	0.01			
Delay	1	0.005	0.38	0.539	
Delay x Alert	2	0.02	2.01	0.150	
Error	33	0.01			
Message Modality x Delay	2	0.003	0.22	0.802	0.798
Message Modality x Delay x Alert	4	0.01	0.56	0.696	0.693
Error	66	0.02			
Window x Delay	2	0.00	0.09	0.910	0.879
Window x Delay x Alert	4	0.01	1.10	0.366	0.362
Error	66	0.01			
Message Modality x Window x Delay	**4**	**0.03**	**2.75**	**0.031**	**0.042**
Message Modality x Window x Delay x Alert	8	0.004	0.42	0.906	0.877
Error	132	0.01			

REFERENCES

Brewster, S., Wright, P. C., & Edwards, A. D. N. (1995). Parallel earcons: Reducing the length of audio messages. *International Journal of Human-Computer Studies, 43*, 153-175.

Broadbent, D. E. (1971). *Decision and stress*. New York: Academic Press.

Brown, J. (1959). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10,* 12-21.

Caelli, T., & Porter, D. (1980). On difficulties in localizing ambulance sirens. *Human Factors, 22*, 719-724.

Dewar, R. E., Kline, D. W., & Swanson, H. A. (January, 1994). Age differences in the comprehension of traffic sign symbols. Paper presented at the 73rd Meeting of the Transportation Research Board. Washington, DC: Transportation Research Board.

Dingus, T. A., & Hulse, M. C. (1993). Human factors research recommendations for the development of design guidelines for Advanced Traveler Information Systems. *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting*, 1067-1071.

Federal Highway Administration (1988). *Manual on uniform traffic control devices for streets and highways*. Washington, DC: U.S. Department of Transportation.

Gaver, W. W. (1993). How do we hear in the world? Explorations in ecological acoustics. *Ecological Psychology, 5*(4), 285-313.

Graham, R., Hirst, S. J., & Carter, C. (1995). Auditory icons for collision avoidance warnings. *Intelligent Transportation: Serving the User Through Deployment. Proceedings of the 1995 ITS AMERICA Annual Conference*, 1057-1063.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika.*

Hawkins, H. E., Womack, K. N., & Mounce, J. M. (1993). Driver comprehension of warning signs. Paper presented at the 72nd Annual Meeting of the Transportation Research Board. Washington, DC: Transportation Research Board.

Huynh, H., & Feldt, L. S. (1970). Conditions under which mean square ratios in repeated measurement designs have fixed F-distributions. *Journal of the American Statistical Association*, *65*, 1582-1589.

ITS America. (1995). *Safety and human factors research needs*. Presented by the Safety and Human Factors Committee. Washington, DC: ITS America.

Lee, J. D., Stone, S. R., Gore, B. F., Colton, C., Macauley, J., Kinghorn, R. A., Campbell, J. L., Finch, M., & Jamieson, G. (1998). *Development of human factors guidelines for advanced traveler information systems and commercial vehicle operations. Design alternatives for in-vehicle information displays* (FHWA-RD-96-147). McLean, VA: Office of Safety and Traffic Operations R&D, Federal Highway Administration.

Micheal, S. G., & Casali, J. G. (1995). Auditory prompts: Effects on visual acquisition time and accuracy in a dashboard-mounted navigational display task. *Proceedings of the 3rd Annual Mid-Atlantic Human Factors Conference*, 15-21.

Norman, D. A. (1988). *The design of everyday things*. New York: Doubleday.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58,* 193-198.

Saunby, C. S., Farber, E. I., & DeMello, J. (1988). Driver understanding and recognition of automotive ISO symbols. Paper presented at the International Congress and Exposition (SAE Technical Paper Series No. 880056). Warrendale, PA: Society of Automotive Engineers.