DIRECT
Operational Field Test Evaluation
Human Factors

Submitted to:
Michigan Department of Transportation
Traffic & Safety Division
Transportation Systems Section
425 West Ottawa Street
P.O. Box 30050
Lansing, MI 48909

August 1998
Contract No. 94-1519 DAB

Prepared by:
University of Michigan
EECS Department
ITS Research Laboratory

Jill Fleming
Paul Green
Stewart Katz

200 Engineering Programs Building
2609 Draper Drive
Ann Arbor, MI 48109-2 101
Phone: (734) 764-4333
FAX: (734) 763-1674

EECS-ITS LAB-DT98-003

Technical Report UMTRI-98-22 June, 1998
(also listed as EECS-ITS LAB-DT98-003)

Driver Performance and Memory for Traffic Messages: Effects of the Number of Messages, Audio Quality, and Relevance

Jill Fleming
Paul Green
Stewart Katz

Technical Report Documentation Page

1 Report No UMTRI-98-22	2 Government Accession No	3 Recipient s Catalog No
4 Title and Subtltle **Driver Performance and Memory for Traffic Messanes: Effects of the Number of Messages, Audio Quality, and Relevance**		5 Repor Date June, 1998 6 Performlng Organization Code none
7 Auth or(s) Jill Fleming Paul Green, Stewart Katz		8 Performing Organlzation Report No UMTRI-98-22
9 Performlng Organization Name and Address The University of Michigan Transportation Research Institute (UMTRI) 2901 Baxter Rd, Ann Arbor, Michigan 48109-2150		10 Work Unit no (TRAIS) 11 Contract of Granl No
12. Sponsonng Agency Name and Address Michigan Department of Transportation Traffic and Safety Division Transportation Systems Section 425 W. Ottawa (P.O. Box 30050), Lansing, MI 48909 Attention: Dr. Kunwar Rajendra		13 Type of Report and Period Covered final 14 Sponsoring Agency Code
15 Supplementary Notes also identified as report # EECS-ITS LAB-DT98-003		

Abstract

In this experiment, 32 licensed drivers (16 young, 16 old) drove on an expressway. On each trial (96 per subject), 1 to 3 traffic messages containing 6 to 14 items were presented. ("l-94 eastbound at Southfield freeway, continuing construction, right lane blocked, 3 mile backup."). Imagining they were driving from Ann Arbor to Detroit on l-94, subjects identified messages relevant to that route and recalled them. Messages were either of good or poor audio quality (to simulate poor reception).

Drivers familiar with the route correctly recognized about 85 % of the relevant messages. Typically 4 items were recalled regardless of message length, with the road and crossroad being most common. Drivers recalled the direction to which the message pertained (e.g., l-94 east) only 39 percent of the time. Drivers believed the traffic information system was safe and useful to listen to while driving (approximately 9 on a lo-point scale) and would pay $177 on average for one, though most subjects were not willing to pay anything.

Of the message characteristics, message audio quality had the largest adverse impact on performance (increasing speed variance). Differences were also found between (1) driving, (2) driving while listening to messages (slight increases in speed variance), and (3) driving while speaking with the experimenter (further increases in speed variance). This is in contrast to claims that dealing with auditory information while driving has no impact on driving workload and is not a safety concern.

17 Key Words ITS, 'human factors, ergonomics, driving, usability, safety, traffic information, auditory interfaces	18. Dlstnbutlon Statement No restrictions, This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classlf. (of thts report) none	20. Secunty Classlf. (of this page) none	21. No of pages 87	22. Price

Form DOT F 1700 7 (8-72) **Reproduction of completed page authorized**

Driver Performance and Memory of Traffic Messages: Effects of the Number of Messages, Audio Quality, and Relevance

UMTRI Technical Report 98-22
Jill Fleming, Paul Green, and Stew Katz

University of Michigan,
Ann Arbor, Michigan, USA

1 ISSUES

1. How does recall vary with message content (number of messages, terms, and relevant messages), message quality, and driver differences (age and sex)?

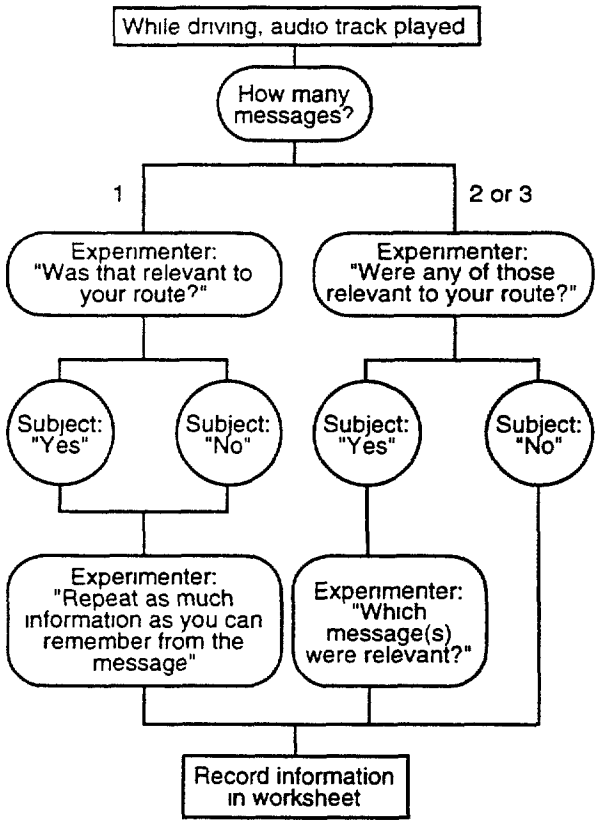
2. How do driver performance (speed, headway, lateral postition) and control inputs (throttle, steering) vary with the message and driver characteristics?

3. How easy and safe to use do drivers rate auditory traffic information systems relative to other in-vehicle tasks?

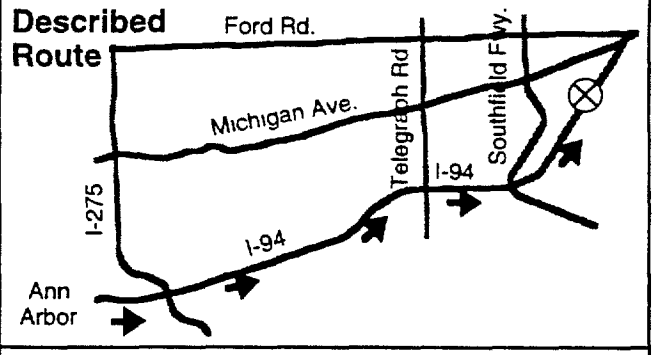
4. What is the rated usefulness of traffic information systems and of each information element? Would drivers use such systems? How much would they pay for them?

2 METHOD

Sequence



Described Route



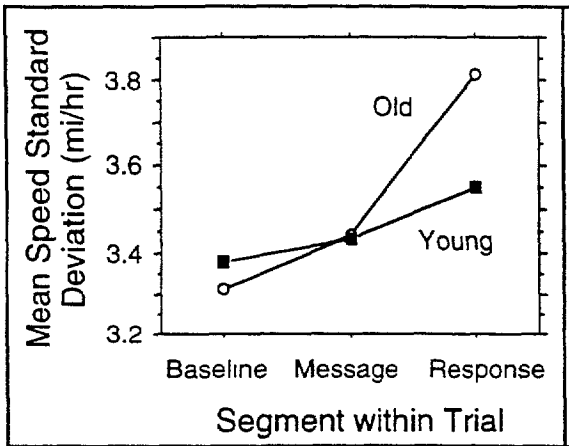
Subjects

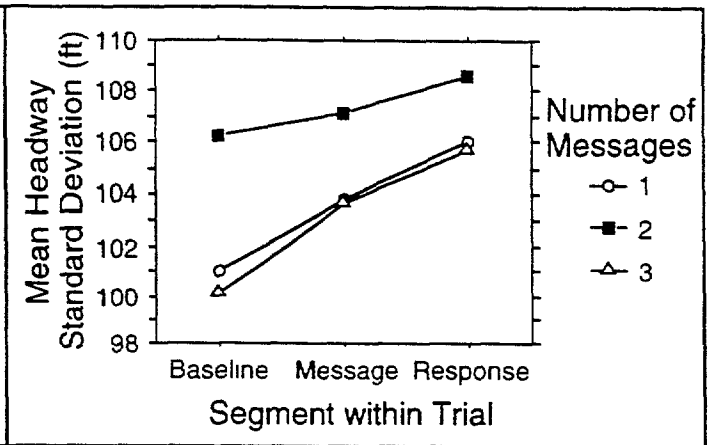
	Young		Old	
Time	Female	Male	Female	Male
Morning	4	4	4	4
Afternoon	4	4	4	4

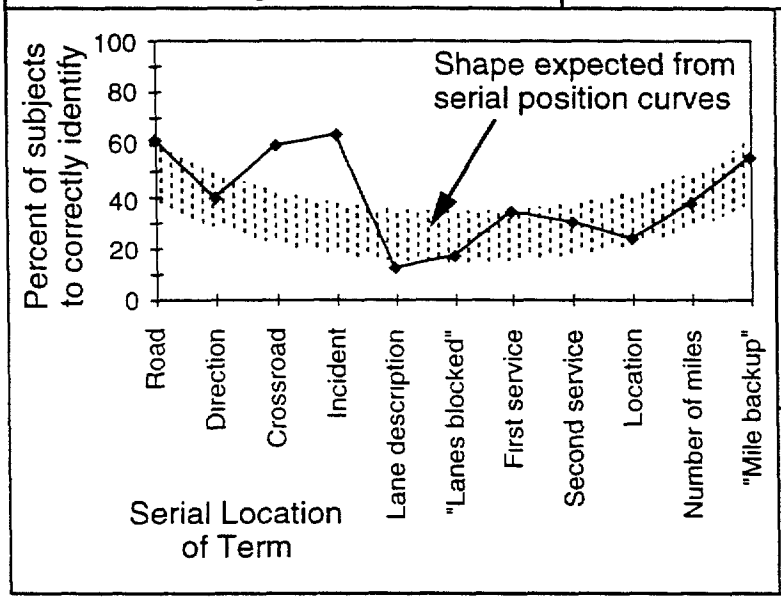
Iterations

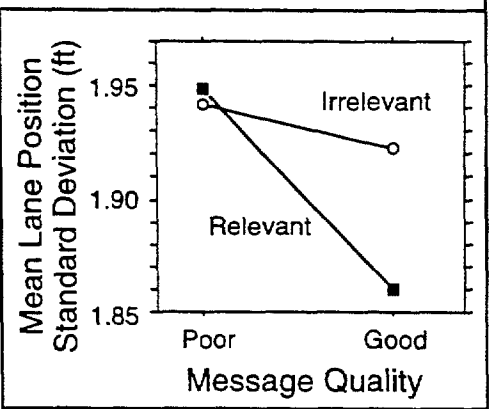
	Message Relevance			
	Relevant		Irrelevant	
	Quality			
Num. of messages	Good	Poor	Good	Poor
1	8	8	8	8
2	8	8	8	8
3	8	8	8	8

3 RESULTS









Amount Willing to Pay (U.S. $)	Number of Subjects
0 - 50	12
70 - 100	6
200 - 500	6

4 CONCLUSIONS

1. Poor audio quality combined with lack of relevance, more than one message, and age led to reduced recognition of the messages. Drivers recalled less than half of the information presented. Approximately 4 terms from each traffic message were recalled, regardless of the amount presented.

2. Poor audio quality, lack of relevance, more than one message, and response tasks led to poorer driving performance.

3. Drivers generally felt that the system was safe for them to use, but did not believe that it was safe for inexperienced drivers.

4. Subjects felt that the information would be useful when driving in a familiar area, but not in an unfamiliar area. On average, drivers were willing to pay $117 (U.S.) for this type of system. However, the most common response was $0.

PREFACE

This report describes an on-the-road experiment conducted as part of the DIRECT (Driver Information Radio using Experimental Communication Technologies) Project, an operational field test of various low-cost traffic information systems sponsored by the Michigan Department of Transportation (MDOT) using funds provided by the U.S. Department of Transportation, Federal Highway Administration (FHWA). In this project 4 mechanisms for presenting traffic information to drivers were evaluated: (1) Low Power Highway Advisory Radio (LPHAR), (2) Automatic Highway Advisory Radio (AHAR), (3) cellular call in, and (4) Radio Broadcast Data System (RBDS).

LPHAR is a descendant of Highway Advisory Radio (HAR), a system that utilizes roadside signs with lights. A flashing light indicates when drivers should tune to a local station for traffic information. Unlike variable message signs, the message is not limited by the size of the sign or the vehicle speed which determines the time available to read the message. The visual distraction of reading the sign is also eliminated. LPHAR, a radio signal, has a range of 1.0 to 1.8 miles, localizing the message and, therefore, allowing for multiple messages in a region.

AHAR is similar to HAR (from the driver's perspective) except that the information is presented to the driver automatically, interrupting (if the driver so chooses) any broadcast the driver is listening to at the moment. However, the equipment used can be quite different from that of HAR.

In cellular call in, the driver calls a particular phone number for traffic information. Options considered at various times included individual phone numbers for each road and a single phone number after which the driver entered the route number. Eventually, to provide more locally specific information, other data could also be needed (e.g., nearest exit).

In the RBDS system, a system that originated in Europe, traffic information is presented on special channels that can be received by modified automotive radios. Also known as RDS-TMC (Radio Data System - Traffic Management Channel), a display with text messages may also be provided. The RDS-TMC system provides for interrupting ongoing broadcasts. To avoid driver overload and the presentation of all possible area traffic messages, a filter indicating the current route must be programmed by the driver. During the planning of this program, the authors were unaware of the driver task of entering the filter, a task that could prove challenging to many drivers and deserves examination.

To examine the merits of these systems, 5 aspects pertaining to the implementation were investigated. In the natural use study (1), 150 drivers were loaned vehicles fitted with these systems for 2 weeks or 2 months for their own use (Reed, Hanafik, Richeson, and Underwood, 1998). Survey data concerning their usage and opinions were obtained. The simulation and modeling effort (2) examined the improvements in traffic flow in the Detroit area as a function of various levels of market penetration of these systems (Underwood, Juna, Gurusamy, Hadj-Alouane, and Hadj-Alouane, 1998). Technical performance and costs were examined (3) to determine how well these systems functioned (signal quality, message accuracy, etc.) and to estimate

production costs (Ristenbatt and Shahine, 1998). Pan of this effort included the collection of intelligibility data. Institutional and organization issues (4) were considered as part of the broader project view (Richeson, Underwood, and Waldman, 1998).

This particular report describes the final area (5), the human factors research conducted as part of the DIRECT project. Of particular interest was the safety and performance impacts of using the 4 systems of interest. The initial discussions of this project centered on providing a broad human factors evaluation of all systems. However, given the limited funds available, such an approach would have been superficial, adding little to the scientific literature.

The approach taken was therefore to identify the safety and human factors issues, and target those that seemed most important and common to all systems. Also considered was the extent to which the research would provide new information, not information that would duplicate the literature. Issues of concern were reading the RDS-TMC display, pressing buttons to retrieve messages, listening and responding to messages, and dialing the cellular phone. Given the limited number of characters on the display when the system was initially being discussed and uncertainty about how it would be implemented, RDS-TMC display issues were set aside for future efforts. Further, most systems only required a single button press to retrieve information, a task that was not thought to pose much risk to drivers. Set aside for future investigation was retrieval of information (the keying task) for a cell phone. The initial design only required calling a particular number, a task examined in the literature (Goodman, Bents, Tijerina, Wierwille, Lerner, and Benel, 1997). However, later implementation may require navigation through menus, a task deserving examination.

The focus, therefore, was on tasks common to all interfaces, such as listening to traffic messages, determining if traffic messages were relevant, and attempting to recall those messages as a function of the amount of information presented. Eventually results from such efforts will include enhanced design guidelines for the presentation of auditory traffic information.

The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the Michigan State Transportation Commission, the Michigan Department of Transportation, or the Federal Highway Administration.

INTRODUCTION **1**
Previous Research 1
Research Issues Investigated 6

TEST PLAN **9**
Test Participants 9
Instrumented Car 10
Traffic Messages 13
Audio Compact Disc 17
Test Route 17
Test Activities and Their Sequence 18

RESULTS **21**
Recognition of Relevant Messages 21
Message Content Recall 23
Ratings of Safety, Ease of Use, Usefulness, and Willingness to Pay 28
Driving Performance and Message Characteristics 31

CONCLUSIONS **39**
How Well Did Drivers Recognize Relevant Messages? 39
How Well Did Drivers Recall Information from Traffic Messages? 39
Did Drivers Believe It Was Safe for People to Listen to Traffic Messages While Driving? 40
How Useful Was the Traffic Information System? 41
How Useful Were the Traffic Information Elements? 41
How Much Were Drivers Willing to Pay for a Traffic System? 41
Did Use of the Traffic Information System Affect Driving Performance and What Were the Effects of Various Message Characteristics? 42
What Should Be Done in Future Studies? 44

REFERENCES **47**

APPENDIX A - MATCHING TASK **49**

APPENDIX B - TEST VEHICLE LAYOUT **51**

APPENDIX C – INSTRUCTIONS **53**

APPENDIX D - SAMPLE OF RECALL AND RECOGNITION WORKSHEET **57**

APPENDIX E - SUBJECT SURVEY **59**

APPENDIX F - RECALL AS A FUNCTION OF MESSAGE LENGTH **61**

APPENDIX G - ANOVA OF DRIVING DATA **67**

APPENDIX H - DETAILED DISCUSSION OF THE DRIVING DATA.....69
Mean Speed 69
Standard Deviation of Speed 70
Standard Deviation of Throttle 72
Mean Headway.....73
Standard Deviation of Headway.....74
Standard Deviation of Lateral Position.....76
Standard Deviation of Steering Wheel Angle 77

INTRODUCTION

This project was conducted as part of the U.S. Intelligent Transportation Systems (ITS) program. The goal of this program is to improve the efficiency and speed with which goods and people are moved, to make transportation safer, and to make travel more enjoyable. This particular project concerns the movement of motor vehicles over public roads. To a significant degree, the movement of goods and people is hampered by congestion. Although congestion has been a problem both for expressways and city streets, the emphasis here is on expressways.

One way to improve system efficiency is to provide drivers with better information about congestion, so that they might drive around congested areas or alter departure times. A variety of methods have been developed for that purpose. In the U.S., the most common method for people to obtain traffic information while driving is from traffic reports broadcasted by AM/FM stations, Although such information is generally complete, it may be dated (due to delays in updates of the radio stations by the police or traffic control centers, or because messages are presented periodically, e.g., every 20 minutes). Further, broadcasts generally cover an entire metropolitan area, even though drivers are only interested in a small portion that applies to their route.

In some cities variable message signs are popular. However, these signs can be distracting to read, and for long messages, a source of congestion rather than congestion relief.

Previous Research

Consequently, there has been considerable interest in audio-based systems that provide localized traffic information, especially systems that provide information about an entire trip so that alternatives can be considered at the beginning of a trip, not after one is caught in congestion.

One of the alternatives considered was Highway Advisory Radio (HAR), a system developed by the FHWA in the early 1980's (Turnage, 1980). When problems occur, drivers are advised by a flashing light on a sign to tune their radios to a particular frequency for further information. In fact, it was in conjunction with the development of HAR that virtually all of the research on understanding of auditory information while driving was conducted in the late 1970's and early 1980's. Other related work on auditory route guidance (e.g., the Back Seat Driver research at MIT, Davis, 1989; Davis and Schmandt, 1989) will not be covered here.

The initial work on the presentation and retention of auditory messages while driving was conducted by Gatling of FHWA (Gatling, 1975, 1976, 1977). All of his studies followed the same basic format. (See Green, 1992 for a summary.) Subjects drove a car on a limited access highway while either tape-recorded messages were presented or slides were shown on a screen in the vehicle (simulating a head-up display (HUD)). Gatling's performance measure was the percentage of subjects making a "route error," that is, not recalling the entire message correctly. Variables manipulated included the modality of the information (auditory versus visual), the repetition of auditory

messages, driver age (young versus old), and the number of items in the message. ("Next right exit; for Boston; via route 213; 3 miles" would be a four-item message.)

In the first of 4 experiments described by Gatling (1975), subjects heard messages containing 1 to 6 chunks of information, presented either once or twice. During a 5 to 15 second delay drivers read aloud unrelated messages (e.g., "slow - automobile accident in right lane") that interfered with rehearsal of the to-be-remembered message.

As expected, there was no effect of the delay on recall since the duration of the interfering task in the delay period was fixed. Card, Moran, and Newell (1983) state that the half-life for working memory (middleman estimate) is 73 seconds for one chunk and 7 seconds for three chunks. Using a 4-second interference period (to read the message), the predicted values are 96 % and 67 % correct, reasonably close to the 97 % and 46 % measured. Error rates were linearly related to the number of units in the message, being 100 percent (no one recalled the entire message) for older drivers at 5 units/message and 100 percent for younger drivers at 6 units. Presenting test messages a second time improved recall by about 15 %. Figure 1 shows some example results.

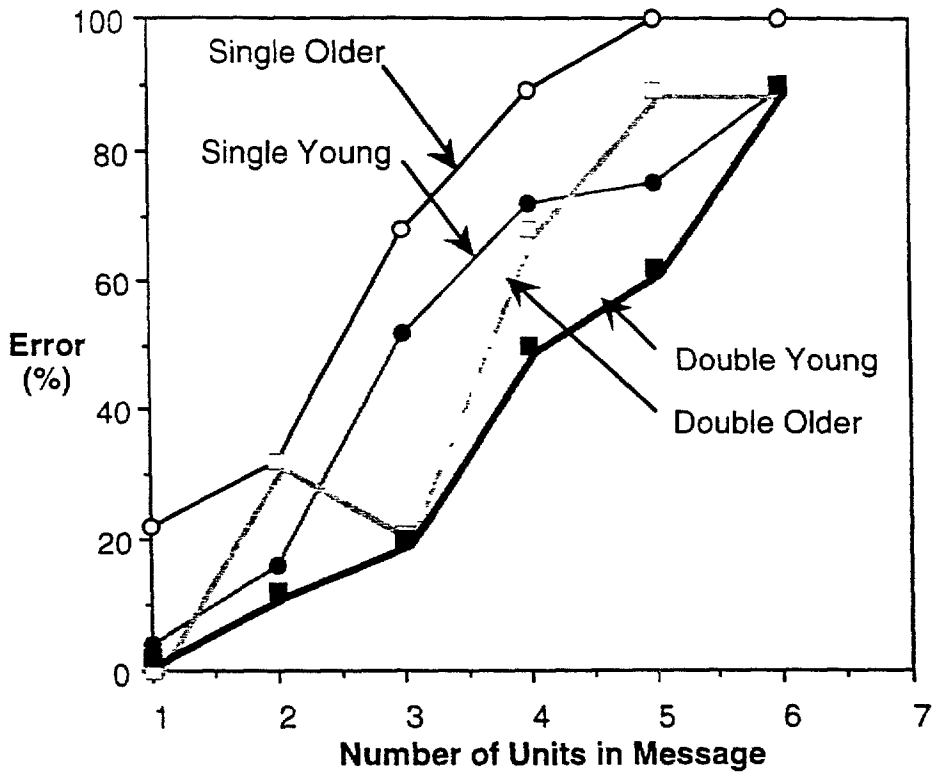


Figure 1. Percent error versus message size, experiment 1 of Gatling (1975). Note: Single or double refers to presentation of the message once or twice.

Probing the drivers for just one piece of information (experiment 2) raised the level at which 100 % errors occurred to 8 units/message. Repeating a message had the same effect as in the first experiment. Recognition was not identical for all information

elements, being about 80-90 % for exits, streets, towns, distances, and turns, but only 50 % for route numbers. The third experiment also showed problems with recall of numeric data while the fourth concerned the time to tune the radio to obtain a message.

Gatling (1977) found that providing additional prose ("This is a warning message that...") led to slightly better retention (by about 5 %) than more succinct messages.

In the last experiment in this series, Gatling found that for messages with 2 route numbers or less, there were fewer errors with visual messages. For 3 or more routes, auditory messages led to fewer errors.

The Gatling research comprises an interesting set of experiments concerning memory of route information. Notable is the use of on-the-road context to assess recall. Unfortunately, the nature of the interfering task was not precisely defined and the duration of the interfering task varied, so the results could be linked more closely to other research in the psychological literature.

Subsequent human factors research was conducted by the Texas Transportation Institute (TTI) to develop design guidelines for HAR (Dudek, Huchingson, and Stockton, 1981; Dudek, Huchingson, and Brackett, 1983). In those experiments, subjects driving on an interstate highway were presented with a tape-recorded HAR message advising of an accident and a diversion route. Drivers then attempted to recall the message and drive the route. Four experiments were conducted. The first examined the number of information elements (exit street, direction of turn, etc.) and the language style (complete sentences/conversation; short form; staccato, e.g., "overturned truck ahead"). The longer messages (10 versus 8 or 6 information elements) and less conversation formats led to more errors. In the second experiment, internal repetition (where the key elements were repeated as part of the message, "turn right on Kingman and take Kingman to Anderson") were found to lead to fewer errors than for external repetition (where the message was given followed by "I repeat …' In study 3, adding landmarks and information on the number of traffic lights helped people negotiate routes even though it lengthened messages. In study 4, unfamiliar drivers were found to make about the same number of driving errors when given turn information as did familiar drivers without turn information. This research led to the following guidelines:

1. Although language style was not found to be critical, a terse message style was preferred by drivers. Unnecessary wordiness is inefficient in communicating messages in a HAR system.

2. If unfamiliar drivers are diverted, the routes should not exceed 4 turns and 4 names, including the Interstate (8-unit problems).

3. The description of the diversion route should be repeated at least once, either with internal or external redundancy or with both.

4. Prominent landmarks may be mentioned in a HAR message whenever there is a risk the driver may not see the place to turn. The number of

traffic lights is useful but should be avoided whenever any of the lights are flashing.

5. When the driving population is known to be largely commuters or highly familiar with the area, the route description may be shortened by omitting turn directions. (Dudek, Huchingson, and Brackett, 1983, p. 9).

Additional recommendations appear in Turnage (1981), p. 25-26.

Audible messages:

1.... The length of a HAR message should be such that the motorist will hear it at least twice while within the HAR zone of coverage...

4, Motorists remember names better than numbers. The greater the frequency of route numbers in a message, the greater the number of route errors made.

5. Motorists retain cautionary messages better than informational messages.

Visual signaling:

1. It takes a motorist about 60 seconds after seeing the first advance visual HAR sign to turn on his radio and tune to the appropriate frequency...

3. Motorists have been found to interpret the degree of urgency to sign messages as follows:

TRAFFIC ALERT	greatest degree of urgency
TRAFFIC ADVISORY	moderately urgent
TRAFFIC INFORMATION	least urgent

4. It is very important that the motorist not be led to expect a message when the HAR station is not operating.

As was suggested earlier, it is peculiar that all of these studies concern human short-term memory, but there has been little effort to connect these applied studies with the results from the psychological literature. For those unfamiliar with the literature, there are 3 forms of human memory linked in a serial fashion. The first form, perceptual store, is involved with the immediate readout of information. Information is stored physically and is generally lost if not attended to within just over a second or less. Depending on the sensory modality for which information is being stored, perceptual memory has a capacity of 5-17 items. This form of memory is operating when people glance at something, look elsewhere, and attempt to recall what was just seen. At the other end of the sequence is long-term memory. Information in long-term memory is usually stored semantically and its capacity is unlimited. This is usually the type of memory being invoked when people say they have memorized something such as

their own name, the presidents of the United States, multiplication tables, the scent of a rose, the sound of a robin, or the taste of chocolate.

Connecting those two forms of memory is short-term or working memory. Short-term memory has a capacity of about 7 chunks of information. Information is maintained in short-term memory by rehearsal and decays exponentially with time if not rehearsed. For example, after looking up a telephone number, one repeats the number to oneself "7 6 3, 3 7 9 5; 7 6 3, 3 7 9 5" to avoid,forgetting it. If the repetition process is interrupted ("What time is it? It is 3:15 p.m."), the to-be-remembered information is often forgotten. Short-term memory along with some aspects of long-term memory are operating when drivers are asked to remember traffic information.

Short-term memory recall and recognition probabilities are determined by a limited set of rules. These rules have implications for auditory traffic messages.

1. Information is stored in chunks, units over which people group information. For example, the string "J M X" would be 3 chunks of information to most people (unless it was there initials or some other memorable item), while NBC would be 1 chunk.

2. Generally, people can keep 7 chunks of information in memory, at least in laboratory situations. For highly reliable recall in real world situations, confine the to-be-remembered information to 4 chunks. This is consistent with the 4 turn-4 road results in the literature.

3. Since information decays exponentially with time, the decay rate is often specified as a half-life, the time period over which half of the information initially available is lost. That time is just over 70 seconds for 1 chunk of information, 7 seconds for 3 chunks.

4. The ability to rehearse information depends on what occurs in the period between presentation and recall. For example, being asked to count backwards by 7s from 119 is likely to block all rehearsal of a previously presented phone number. Rehearsal is an active process, so not rehearsing can also cause information loss. There is no statistical evidence on the extent to which real driving interferes with rehearsal of auditory information and how that should be accounted for in calculations, though such interference is believed to be minimal.

5. When the capacity of the short-term store is exceeded, the information that is least important and/or oldest is deleted, depending upon user needs. For traffic information, the main road and the nearest exit or crossroad are priority items. In contrast, the number of cars in a crash is lower priority and more likely to be deleted when overload occurs.

As a consequence of these rules, when people are asked to memorize a list of directions, the curve shown in Figure 2 often results. In brief, the first few items in the list are remembered because there are fewer items in the list when they are first encountered , so they can be rehearsed a greater number of times. For example, for the first item, that item is the only one to repeat. For the last few items, the time

between presentation and recall is less than for items earlier in the list, so less decay occurs. One would expect to see such recall functions for auditory traffic messages.

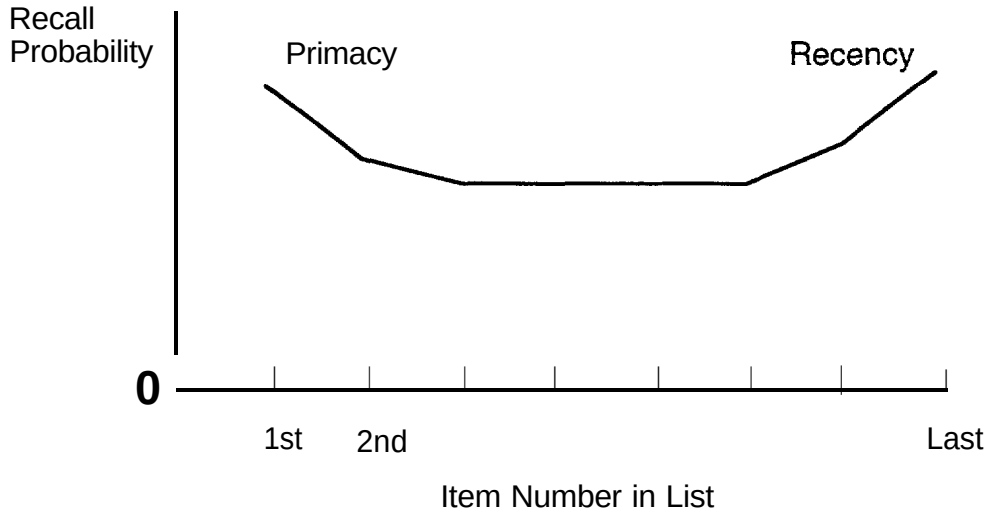


Figure 2. Recall probability for items in a list.

Research Issues Investigated

In selecting issues to examine, the following were considered:

1. What issues were not addressed in the literature?
2. Did the issue pertain to multiple systems?
3. Would the results have applicability beyond the DIRECT project?
4. Would the results have a practical impact?
5. Would the results aid in the selection of a particular implementation?
6. Was the issue reasonably inexpensive to evaluate?
7. Could the issue be included in a clean experimental design?

The literature contained considerable information on message recall. However, the effects of message quality have not been considered nor have the effects of listening to messages on driving performance. These issues were common to all of the systems of interest and, potentially, had major impacts on safety, usability, and usefulness of auditory traffic information systems. Accordingly, those topics were the focus of this research. Specifically:

1. How does the recall of real messages vary

   (a) with the total amount of relevant information presented (number of terms)
   (b) if one is relevant
   (c & d) with the total amount of information presented: number of messages (c) and the number of terms (d)
   (e) with message quality
   (f) with driver differences (age and sex)?

Of these issues, message quality has never been addressed in the literature.

2. How does driving performance (mean speed, speed variance, mean headway, headway variance, lateral position variance) and driver control inputs (throttle variance, steering wheel angle variance) vary with

   (a-e) the message characteristics named above (message length, relevance, and quality), as well as
   (f) driver age and sex?

   The impact of auditory message characteristics on driving performance has not been examined in the literature.

3. Relative to other in-vehicle tasks, how easy and safe to use do drivers consider auditory traffic information systems to be?

   To date, the literature has considered recall and route following, but not driver impressions.

4. How useful do drivers consider such systems to be, and what is the relative usefulness of each information element? Would drivers use such systems if they had them?

   Usefulness provides another perspective of what should be in traffic messages.

5. How much are drivers willing to pay for such systems?

   Willingness to pay is a primary measure of the perceived value of a product or service.

In addition, other issues were considered (e.g., (1) if messages were automatically presented and “never” terminated or were manually retrieved and self terminating, (2) the nature of the signal quality loss (diminished signal- to-noise ratio versus clips and drops)) but not examined due to budget constraints.

TEST PLAN

Test Participants

A total of 32 licensed drivers participated in the experiment. They were either friends of the experimenters, respondents to a newspaper advertisement, or participants in previous UMTRI studies. As shown in Table 1, subjects were divided equally by gender, age group, and time of day. The traffic levels were approximately equivalent for the two test times. Young subjects ranged from age 18 - 29 (mean = 24) and older subjects ranged from age 65 - 81 (mean = 71). Subjects were paid a total of $30 upon completion of the 2-hour experiment.

Table 1. Subjects.

	Young		Older	
Time	Male	Female	Male	Female
10 a.m.-12 p.m.	4	4	4	4
1:30 p.m.-3:30 p.m.	4	4	4	4

Participants reported that they drove between 1,500 and 86,000 miles per year (mean = 17,300). Older drivers accumulated about 7600 more miles per year than younger drivers. Subjects rated their familiarity with the Detroit area highway system as 5.8 (mean) on a scale from 1 to 10 (1 = unfamiliar and 10 = very familiar). On the same scale, subjects rated their familiarity with the DIRECT traffic information project as 2.2. Additionally, subjects were asked how often they drive in the Detroit area. Responses to these questions can be found in Table 2.

Table 2. Subject responses to driving questions.

	Number of subjects responses to question (n = 32)	
Subject Response	How often do you drive on the Detroit highway system?	How often do you drive on I-94 eastbound between I-275 and I-75?
Daily	2	0
Few times/week	3	2
Once/week	4	2
Few times/month	7	5
Once/month	6	3
Less	10	20
Total	32	32

To determine how familiar subjects were with the route to which the traffic messages referred, subjects matched the names of major roads in metro Detroit with roads shown on an unlabeled map (Appendix A). The matching task was completed at three separate times before driving the test route. By the third time, subjects correctly matched all roads on the route for which the audio traffic messages would be provided. (See Table 3.) By the third map-matching task, almost all subjects correctly identified the relevant roads.

Table 3. Number of correct matches.

		Number of correct matches (n = 32)		
		Task repetition		
Road	**Relevant**	1	2	3
I-75	no	13	19	19
Michigan Ave.	no	23	29	29
I-275	yes	21	28	31
Telegraph Rd.	yes	8	26	29
Davison Fwy.	no	9	21	24
I-96	no	25	28	27
Southfield Fwy.	yes	8	24	28
Ford Rd.	no	19	25	25
I-94	yes	30	32	32
Lodge Fwy.	no	13	17	19

Participants were also asked a series of questions about their radio listening preferences while driving. Twenty-three subjects indicated that they listen to music most frequently while driving, 8 subjects listen to news, and 7 subjects listen to talk shows. (Note: these numbers do not add up to 32 because subjects were allowed to respond with more than one answer.) Nine subjects change/tune the radio/CD/tape once every 10 minutes while driving, 7 subjects change/tune once every 30 minutes, 5 subjects change/tune once every hour, and the remaining 11 subjects change/tune less often. Six subjects indicated that they listen to radio broadcasts concerning local or area traffic daily, 7 listen to reports a few times a week, 1 subject listens once a week, 5 listen a few times a month, 3 listen once a month, and 7 subjects listen to traffic reports less often.

Instrumented Car

The test vehicle was the UMTRI Driver Interface Research Vehicle, a highly instrumented, 1991 Honda Accord station wagon. This car has sensors for all major driver control inputs (steering wheel angle, throttle and brake position, turn signal, cruise), vehicle responses (speed, lateral position), and cameras for recording the forward scene and driver.

The video recording system consisted of two bullet (lipstick) cameras (one to record the forward scene mounted below the inside rearview mirror, and a second aimed at the driver and mounted on the A-pillar) and two small cameras located in the outside mirrors to record the lane markings on either side of the vehicle (lane trackers). Camera outputs were combined, along with a summary of the data collected by the computer by a quad splitter, displayed on a monitor, and recorded on a VCR. The two lane tracker images were combined by a two-image splitter and fill one quadrant of the quad splitter image. (See Figure 3.)

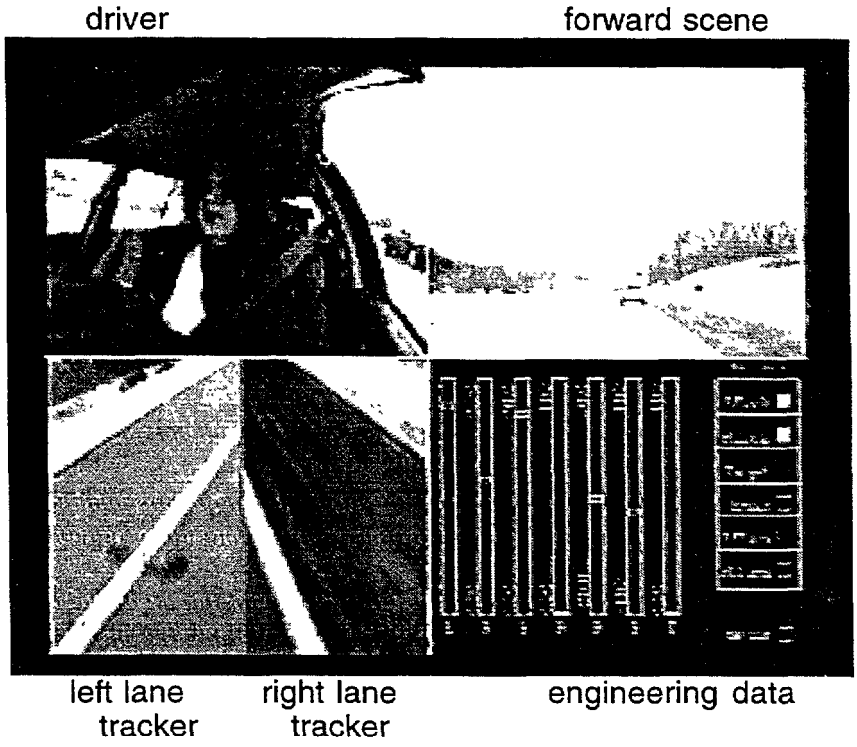


Figure 3. Typical quad-screen image.

Note: The capital letters L,R,S,R,R,S,T at the bottom of the screen are labels for the bar graphs for the left lane tracker, right lane tracker, speed, range (headway), range rate, steering, and throttle.

Sound was picked up by two miniature lavolier microphones, one mounted on the A-pillar, a second mounted on the inside rearview mirror. An audio mixer combined the two microphone outputs for recording on one of the VCR's audio channels.

Engineering data was collected by a 486 computer via a custom-made signal conditioner (both located in the cargo section of the car). Sensors included a potentiometer mounted below the steering wheel (to measure steering-wheel angle), and a headway sensor mounted to the front bumper. The engine computer located under the passenger's feet provided the speed, throttle, and brake signals. Lane position was determined in real time by the 486 from video images provided by the lane tracking cameras. The 486 received the majority of its data from the custom-built signal conditioner that received the signals from both the engine controller and the steering wheel angle sensor. The data were stored on an external hard drive and then copied to a Bernoulli drive for analysis.

The data collection and video equipment can be either powered by the car, or when stationary, by a 1 IO-volt AC wall outlet source. During on-road tests, a 400-watt, 1 10-volt AC power converter connected to the car's electrical system to supplement the 12-volt supply drawn from the car's battery. There were no supplemental batteries to power the equipment. Figure 4 shows most of the engineering data equipment and the power supplies in the rear of the test vehicle.

All equipment was operated by an experimenter seated in the right rear passenger seat. Using the video display showing the quad splitter output (Figure 3), the experimenter monitored the camera output, making adjustments as necessary and checked the proper operation of all engineering data sensors. A keyboard was in the equipment rack next to the experimenter (and behind the driver). Figure 5 shows the equipment available to the experimenter in the back seat. Appendix B shows a plan view of the test vehicle and the model numbers of all equipment in the vehicle. The measures collected by the instrumented car are summarized in Table 4.

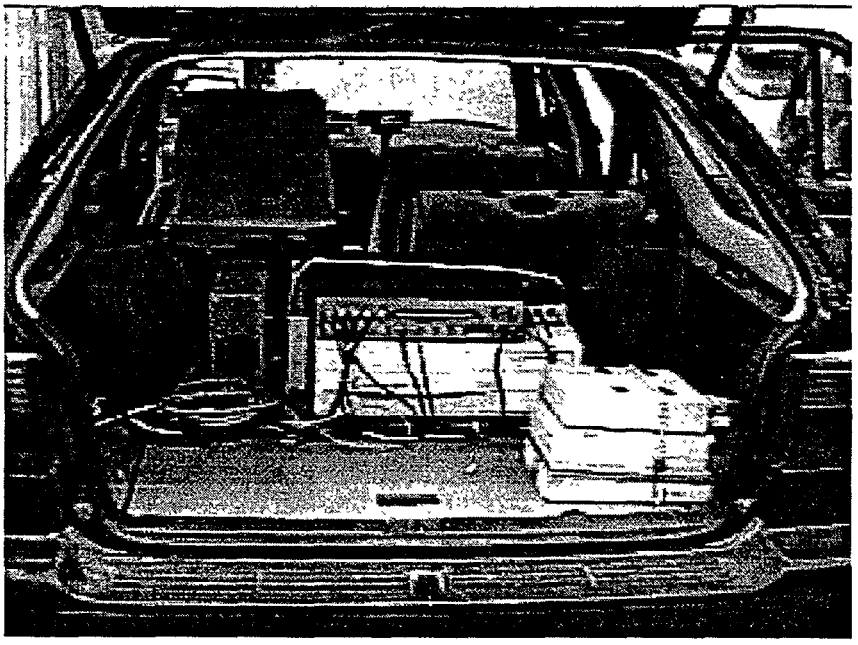


Figure 4. Data collection equipment and power supplies.

Figure 5. Some of the equipment operated by the experimenter.

During this experiment, a portable compact disc player was installed in the test vehicle. A Fisher EES Opti Trac Personal Compact Disc Player (Model PCD-5) was attached using Velcro to the car next to the rear passenger seat. The CD player was powered through the cigarette lighter and was connected to the tape deck of the vehicle via an adapter cassette.

Table 4. Measures collected.

Measure	Units	Sampling Rate (Hz)	Type	Comments
left lane distance	ft	10	vehicle response	from center of edge marking to centerline of vehicle
right lane distance	ft	10	vehicle response	from center of edge marking to centerline of vehicle
speed	ft./s	10	vehicle response	from wheel pulser, later converted to mi/hr
range (headway)	m	30	vehicle response	to lead vehicle (if detected)
range rate	m/s	30	vehicle response	change in range to lead vehicle (rate of closure)
steering wheel angle	deg	30	driver input	position of steering wheel, plus is clockwise
throttle position	percent	30	driver input	angle of accelerator, 0 is no depression, floored is 100
left lock	none	10	status	does the left lane tracker see a lane marking (yes/no)
right lock	none	10	status	does the right lane tracker see a lane marking (yes/no)
headway target	none	10	status	does the headway sensor see a target (yes/no)
brake	none	10	driver input	is the brake pedal being pressed (yes/no)
left turn	none	10	driver input	is the left turn signal on (yes/no)
right turn	none	10	driver input	is the right turn signal on (yes/no)

Traffic Messages

Traffic messages were designed to cover a large portion of the metropolitan Detroit roads. The structure and content was designed to mimic messages actually used by the MDOT (Michigan Department of Transportation) traffic control center in Detroit, Michigan. Messages reported traffic incidents at major intersections. Each intersection was described first by the road that the incident occurred on, second by the direction (north, south, east, or west) of the road where the incident occurred, and third by the closest intersecting street to the incident. Seven main roads were used as sites of traffic incidents Some of these seven major thoroughfares intersected with each other, and seven additional roads were used as intersecting streets. The 7 roads chosen as locations for traffic incidents and the directions of traffic flow on these

roads are shown in the first and second columns of Table 7. The roads used as intersections for each of the seven main roads are listed in the third column of Table 5.

Table 5. Roads used as sites of traffic incidents and their intersections.

Road name	Traffic flow direction	Intersecting roads	
I-94	East/West	*I-275 *Southfield Freeway *Lodge Freeway	*Telegraph Road *I-96 *I-75
‘I-275	North/South	*Michigan Avenue *I-96	*Ford Road
Telegraph Road	North/South	*Michigan Avenue *I-96	*Ford Road
Southfield Freeway	North/South	*Michigan Avenue *I-96	*Ford Road *I-75
I-96	East/West	*Southfield Freeway *I-275	*Telegraph Road
Lodge Freeway	North/South	*Davison Freeway *6 Mile Road	*Livernois Road
‘I-75	North/South	* Davison Freeway *I-696	*7 Mile Road

It was important that the path chosen as relevant for this experiment included roads with which the majority of drivers were familiar. Therefore, the route chosen as the “relevant” route was one commonly used by those in Ann Arbor (the test sample) to travel to Detroit. (See Figure 6.) Subjects were informed that their destination was off of I-94 after Southfield Freeway, but before Michigan Avenue. Relevant messages pertain to the driven road, I-94 eastbound before Michigan Avenue, as well as roads intersecting I-94 before Michigan Avenue. Messages pertaining to the following intersections were relevant to the route: I-275 and I-94, Telegraph Road and I-94, and Southfield Freeway and I-94. Because the route takes I-94 east, any message pertaining to I-94 west was not relevant. Messages pertaining to any of the intersecting roads (I-275, Telegraph Road, or Southfield Freeway) at I-94 were also considered relevant because they could spill over and cause congestion on I-94. A typical traffic message heard by subjects was: “I-94 eastbound at Southfield freeway, continuing construction, right lane blocked, 3 mile backup.”

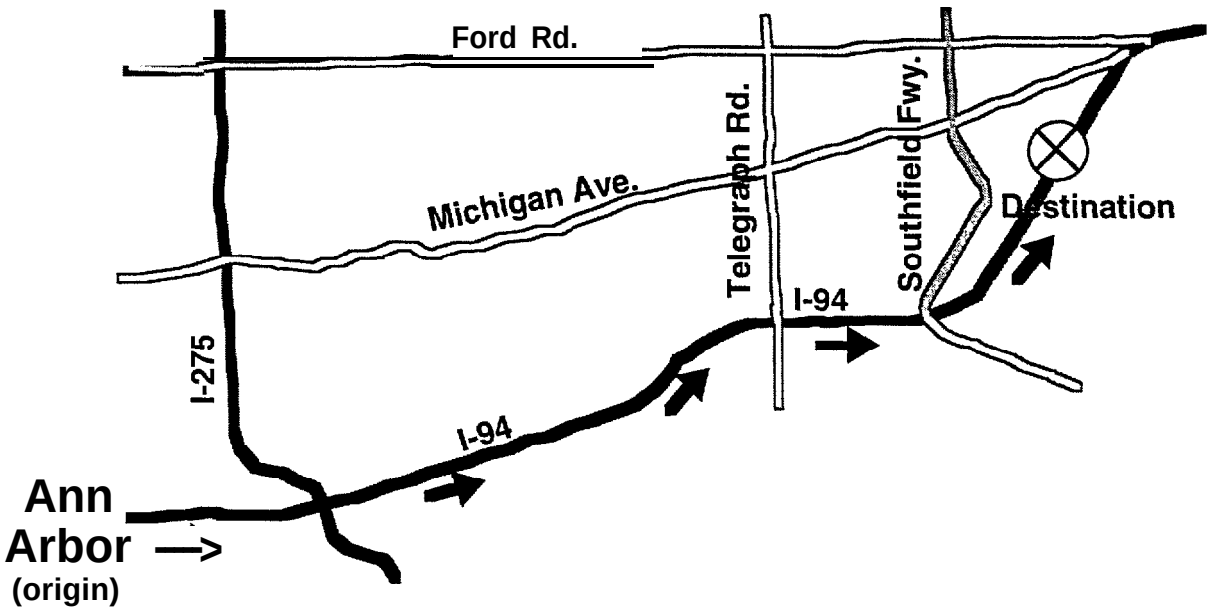


Figure 6. Relevant route for traffic messages.

Each of the traffic messages had a cause of congestion (e.g., accident, construction, etc.). Seven causes were provided, four of which had to do with differing accident severity. There were a variety of incident descriptions depending on the cause of congestion. An incident description was possibly accompanied by blockage information, service vehicle information, and traffic backup information. Table 6 shows the possible components of the traffic message based on the cause of congestion.

Table 6. Components of traffic messages.

Cause of Congestion	Incident Description	Blockage	Additional information	Back-up
minor accident	*1 car minor accident *2 car minor accident			
accident	*2 car accident *multiple car accident	*right lane blocked *2 right lanes blocked	*police en route *police on scene	*1/2 mile back-up
injury accident	*multiple car injury accident *2 car injury accidenl		*police and EMS en route *police and EMS on scene	*3 mile back-up
serious injury accident	*multiple car serious injury accident *2 car serious injury accident	*right lane blocked *2 right lanes blocked *all lanes blocked	*police and EMS en route *police and EMS on scene	*5 mile back-up
construction	*continuing construction *shoulder work	*right lane blocked		*3 mile back-up
hazardous material spill	*hazardous material spill	*right lane blocked *2 right lanes blocked	*clean up crew en route *clean up crew on site	*2 mile back-up *4 mile back-up
railroad derailing	*railroad derailing	*right lane blocked *2 right lanes blocked	*police and EMS en route *police and EMS on scene	*4 mile back-up *5 mile back-up
sporting event/ concert	*congestion due to event			*1 mile back-up *2 mile back-up /*4 mile back-up

Several aspects of the traffic messages were studied. One issue was whether the number of traffic messages presented to the driver affects either driving performance or driver response to the information in the traffic message. To determine the effect of more than one traffic message at a time, traffic messages were presented in groups of either 1, 2, or 3 messages per trial.

There are times when poor quality audio signals are received by the vehicle. Because the vehicle is distant from the receiver, there is physical interference (buildings) or electrical interference. This is particularly important for the low-power systems being studied. Therefore, a second issue to consider is whether or not the quality of the audio signal affects driving performance and/or driver response to traffic information. To study the effect of message audio quality, half of the traffic messages were degraded to simulate the effect of static.

Finally, the issue of whether or not the traffic message was relevant to the driver's route was studied. Therefore, half of the traffic message trials contained, at least, one message that was relevant to the route the driver was taking. The other half of the

trials contained no messages that were relevant to the driver's route. Table 7 shows the matrix of the experimental design. Eight replications for each condition (message quality * number of messages * relevance) were presented to the driver, for a total of 96 trials per subject.

Table 7. Design of the experiment.

	Message relevance			
	Relevant		Irrelevant	
Number of messages	Message quality		Message quality	
	good	poor	good	poor
1	8	8	8	8
2	8	8	8	8
3	8	8	8	8

Audio Compact Disc

To assure consistent message quality, an audio CD was custom-made for this experiment. A male UMTRI employee recited each traffic message using Adobe Premiere version 4.0 running on a Power Macintosh 8500/120 (and its microphone) to record each message. A separate audio track was used for each message. For trials of 2 or 3 traffic messages, a 2-second pause (silence) was inserted between the traffic messages of that trial. Between each trial, an 8-second pause of silence was inserted.

Half of the traffic messages were degraded using Adobe Premiere to simulate the poor audio quality. To degrade messages, three audio tracks were played simultaneously. The first audio track, the foreground, used the boost filter to amplify weak sounds while leaving loud sounds intact, the echo filter (creating an echo effect) was used with a 13-second delay and soft intensity, an adjusted gain of 50 percent, and varying levels (volume) were created throughout. The varying levels were produced using the audio fade control of Adobe Premiere. The second track of the poor quality messages, the first background track, used the backwards audio filter to play the message backwards, the boost filter, the echo filter at a 13-second delay and soft intensity, and a gain of 5 percent. For the third audio track, the second background track, static was recorded from the radio (in between radio stations) and was set to a gain of 120 percent. Together, all three of these audio tracks produced one degraded traffic message.

Once all traffic messages were recorded and properly manipulated, each trial of messages (1, 2, or 3 messages in a trial) was separately saved as a .wav file using SoundEdit Pro version 1.05. All 96 audio .wav files were saved to disc, and were then sent to a contractor who burned them onto a compact disc as 96 separate audio tracks. The CD was used during the experiment to play traffic messages to the drivers.

Test Route

Unfortunately, the route chosen for relevant traffic messages (I-94 east of Ann Arbor) was under construction for the duration of experimental testing. Because driving was simply a loading task, an alternate route (I-94 west of Ann Arbor) was driven while drivers listened to messages and imagined they were driving towards Detroit. The

task of reading signs, a minor additional load, was ignored. I-94 west is a fairly straight, flat highway with a speed limit of 70 mi/hr. Drivers were informed when they should exit from the highway and return to Ann Arbor via I-94 east. (See Figure 7.)

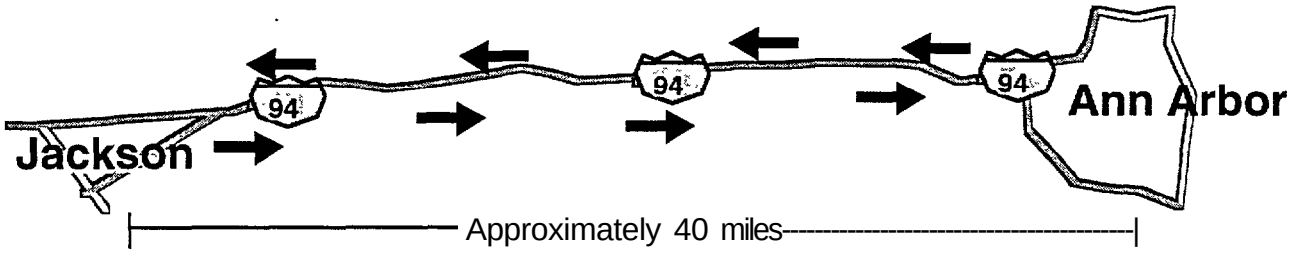


Figure 7. Test route driven.

Test Activities and Their Sequence

After listening to an introduction of the experiment (see Appendix C for a copy of the instructions), subjects matched roads on a map (identified by letter) with a list of names (as described earlier). Subjects were informed if they had made any mistakes when the task was completed. Next, a labeled map was shown to the subjects to familiarize them with the road names.

Subjects were then told to pretend they were driving on I-94 east of Ann Arbor to Detroit (the "relevant" route) even though they actually would be driving on I-94 west (to avoid construction). Subjects then completed the matching task a second time. Again, subjects were informed if they had made any mistakes.

Each participant then completed a biographical form and a consent form (Appendices D and E, respectively).

In the test vehicle, while the subject adjusted the seat, vehicle mirrors, etc., the experimenter, seated in the right back seat of the vehicle, initialized the test equipment, and then reinforced key points concerning the relevant route. Each subject chose a music CD to listen to while driving to the beginning of the test route, adjusting the volume to a comfortable level. The same volume level was later used for traffic messages. Before heading out of the parking lot, the subject was asked to complete the road-matching task a third and final time. Upon completion of the task, the subject was guided by experimenter to the entrance to I-94 westbound.

Once the subject had merged onto I-94, the experimenter replaced the music CD with the traffic-message audio CD. The CD player displayed the number of the audio track currently playing. Precisely when this number changed (e.g., from track 1 to track 2), the experimenter would enter the number of that trial into the computer. Three seconds of silence were present between the time that the track number changed on the CD player and the time that the traffic messages began playing. When the traffic messages for a given trial were completed, the experimenter paused the CD and the subject responded to a series of questions. Once the subject had responded to all questions, the experimenter played the CD again. Five seconds of silence were present after the traffic messages were completed, but before the track number

advanced on the CD player. The experimenter entered the next trial number into the computer precisely as the track number advanced on the CD player. This system of coding was later used to parse the driving data into 96 separate sections, one for each trial.

Following each audio track, the experimenter asked the subject questions about the traffic messages that had just been played. If the trial consisted of two or three traffic messages, then the experimenter would simply ask if any of the traffic messages were "relevant" to the route described to them previously. If the subject responded yes to this question, the experimenter then asked the subject to indicate which message or messages were "relevant". This data was entered into a worksheet (Appendix D) by the experimenter. If only one traffic message was present during the trial, the subject was first asked if the traffic message was "relevant," and was then asked to repeat as much information as he/she could remember from the traffic message. A flow chart of the experimenter-subject interaction is shown in Figure 8.

The first 50 trials of audio traffic messages were played to the subjects during the drive on I-94 west. Once the 50th trial was completed, subjects were asked to exit the highway and to pull over and park in a lot immediately off the exit ramp. The first half of the driving data was then saved, and a new file name was entered into the computer. Subjects were then asked to get onto I-94 eastbound. Trials 51 through 96 were completed during the return trip to Ann Arbor. Once all of the trials were complete, the subject was given directions to return to UMTRI.

In the UMTRI parking lot, subjects completed a post-driving survey. (See Appendix E.) Questions concerned safety, the usefulness of various aspects of the traffic messages, system use, and the difficulty of common driving tasks. Additionally, subjects were asked about the audio quality of the traffic messages, and how much they would be willing to spend on a traffic information system.

After the paperwork was completed, subjects were paid and thanked for their participation in the experiment.

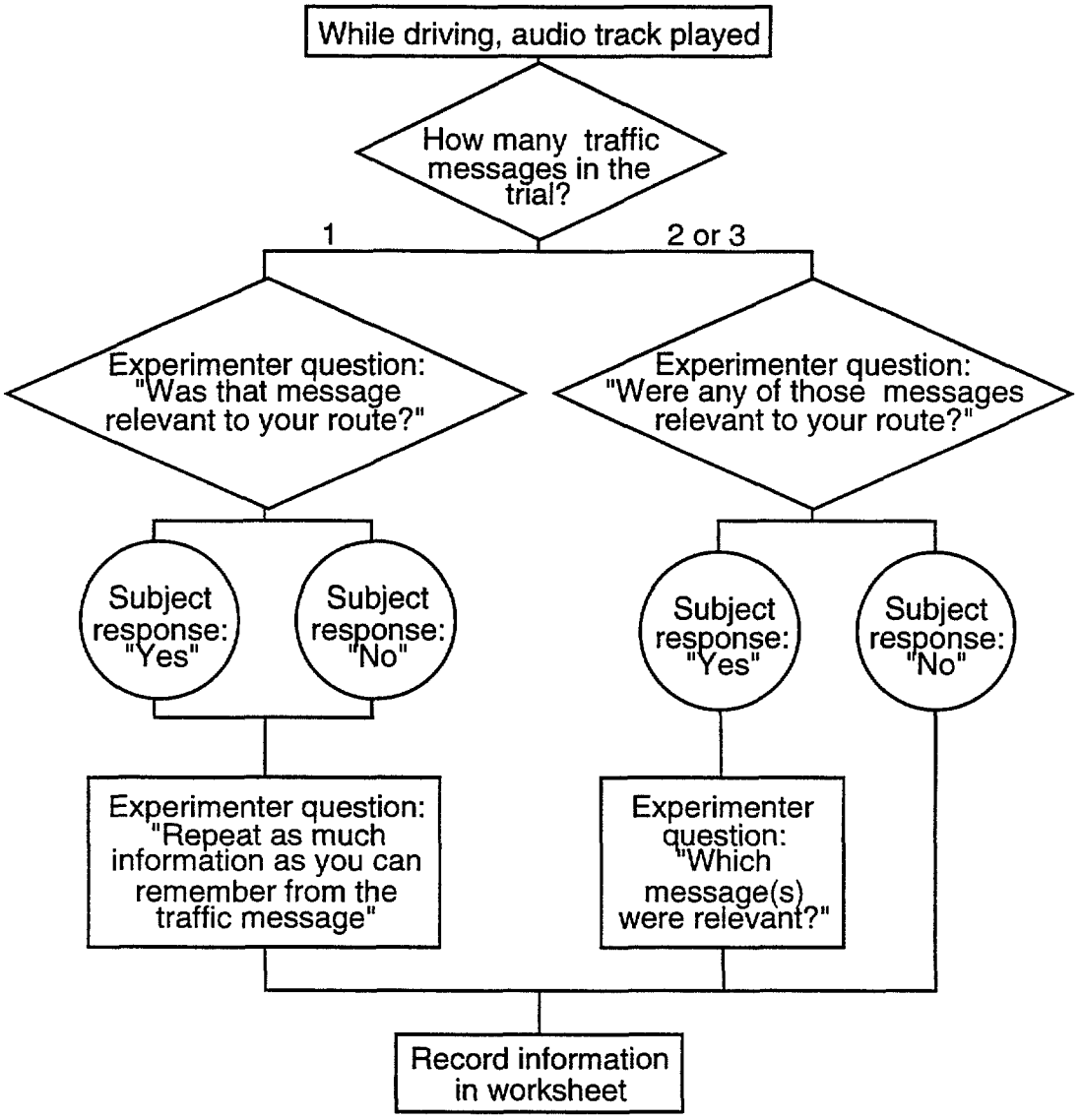


Figure 8. Flow chart of questions asked by experimenter.

RESULTS

Recognition of Relevant Messages

Question: How does the recognition that traffic messages are relevant vary with (1) message relevance, (2) the number of messages/trial, (3) audio quality, (4) time of day, and (5) driver differences (age and sex).

Subjects were explicitly told which intersections were considered to be relevant to the route, as described in the test plan and checked using the map-matching task. After each message sequence, subjects identified if each traffic message was relevant or irrelevant to the route. For the most part, subjects identified relevant messages as instructed by the experimenter. A few subjects identified messages that contained incidents on the major road traveled (I-94 eastbound) as relevant, but did not consider messages with incidents on one of the cross roads (e.g., I-275, Telegraph, or Southfield) at the major thoroughfare (I-94) to be relevant. Therefore, these subjects identified fewer relevant messages. In contrast, a few subjects considered messages containing incidents a couple of miles away from the route to be relevant because they may have caused a large enough back-up to affect the route. These subjects, therefore, identified a greater number of relevant messages.

To determine what factors affected the number of messages recognized by each subject, a repeated analysis of variance (ANOVA) was computed as shown in Table 8. The dependent measure in the model was the number of messages recognized. If the trial contained three traffic messages where the first message was relevant, and the last two messages were irrelevant, and the subject recognized such, then the subject would be scored as identifying three messages correctly. Since 1, 2, or 3 messages occurred in each trial, each equally often, the mean number of messages was 2.

Table 8. ANOVA independent variables for message recognition.

Factor	Levels	Comments
age group	2	young, old
gender	2	male, female
time of day	2	morning, afternoon
message characteristics	8	compact variable-combined message relevance (relevant, irrelevant), audio quality (good, poor), number of messages/trial (1,2,3)
repetitions	8	of message characteristics compact variable

As is usually the case, individual differences (age and, secondarily, sex) were important factors. Younger subjects recognized significantly more messages correctly ($p = 0.0074$, mean = 1.8 for young drivers, mean = 1.6 for old drivers), and the time of day was also significant ($p = 0.04$, morning mean = 1.7 and afternoon mean = 1.8). Also significant, was the unexplained interaction between sex and time of day (Figure 9).

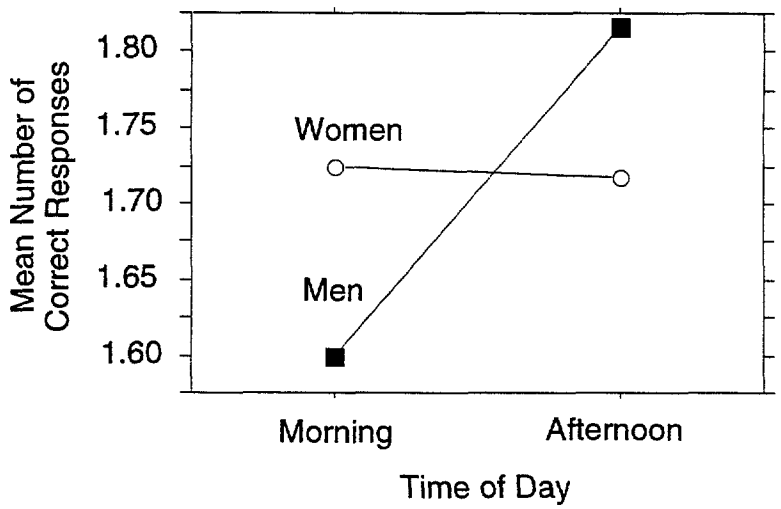


Figure 9. Mean number of correct responses for gender by time of day.

In general, message characteristics had small impacts on performance, as did some of their interactions. The number of correctly identified messages per trial increased as the total number per trial increased ($p = 0.0001$), though the absolute percentage decreased slightly (87.2 %, 86.9 %, and 84.4% correct). The number of messages and age also interacted ($p = 0.0002$), with the age gap increasing with the number of messages (Figure 10).

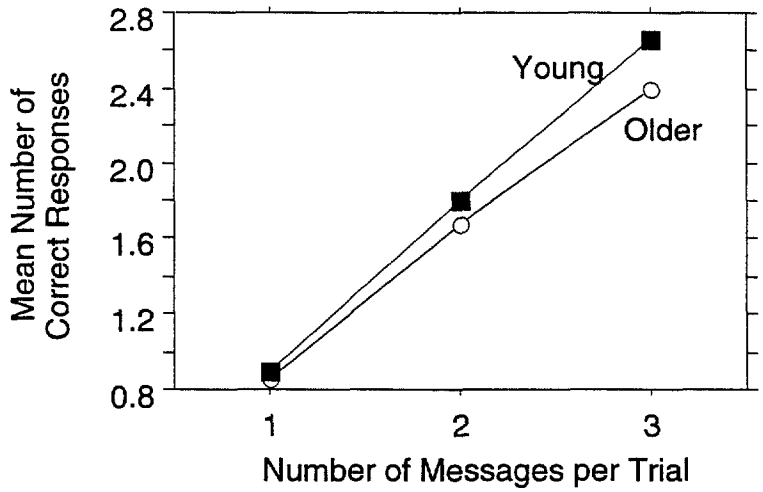


Figure 10. Mean number of correct responses for the number of messages by age group interaction.

Also significant was message quality ($p = 0.0016$) and its interaction with relevance (irrelevant versus relevant messages), but not the main effect of audio quality (Figure 11). The lack of a difference when audio quality is poor makes sense. If a message is difficult to decipher, relevance should not matter. However, the interaction does not make sense. If anything, the number of correct responses should be greater for relevant messages than for irrelevant messages when audio quality is good because drivers will pay more attention to relevant messages.

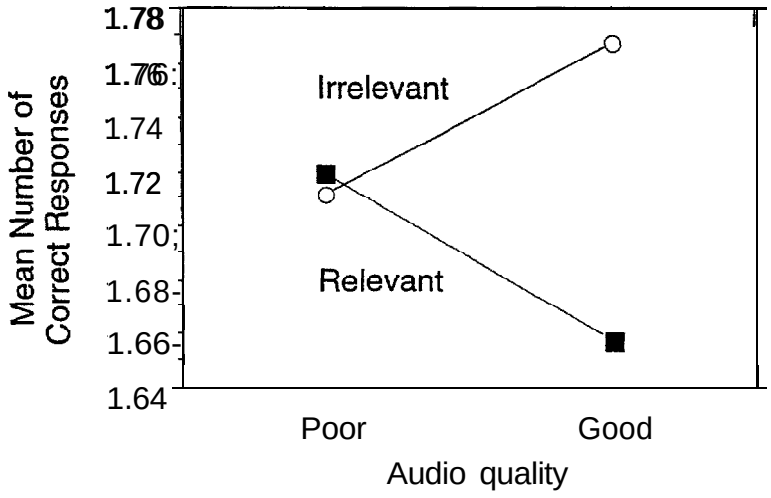


Figure 11. Mean number of correct responses for the relevance by audio quality interaction.

Thus, most factors seemed to have small impacts on the number of messages recalled, with driver age being the most important factor.

Message Content Recall

Question: Which terms in a message were **most likely to be remembered?**

For single message trials, subjects recalled the message. Each piece of information that was repeated correctly was counted as a correct term. (See Table 9.) Other partitioning strategies could also make sense. False reports (additional terms) were tallied as well. In this experiment, messages contained between 6 and 14 terms (mean of 9.7, Figure 11).

Table 9. Example of a 14-term traffic message.
("Telegraph Road northbound at I-94, two car serious injury accident, right lane blocked, police and EMS en route, three mile back-up")

Terms	Piece of information/Term	Example
1	Road incident occurred on	Telegraph Rd.
1	Direction of travel incident occurred on	northbound
1	Cross road that incident occurred near	at I-94
1	Piece of information that describes incident	2 car
1	Additional piece of information that describes incident	serious
1	Additional piece of information that describes incident	injury
1	Type of incident	accident
1	Description of lane blockage	right
1	Existence of lane blockage	lane blocked
1	Type of services involved	police
1	Additional type of service involved	and EMS
1	Location of service involved	en route
1	Length of back-up	3 miles
1	Existence of back-up	3 mile backup

As suggested by the literature on limited capacity models of human short-term memory, subjects consistently recalled about 4 terms (Figure 12), no matter how many were in the message, except for 14-term messages (better recall), and 8 terms (worse recall). The poor recall of 2 terms may reflect statistical variability. (There were only 2 messages of that length.) For additional details, see Appendix F.

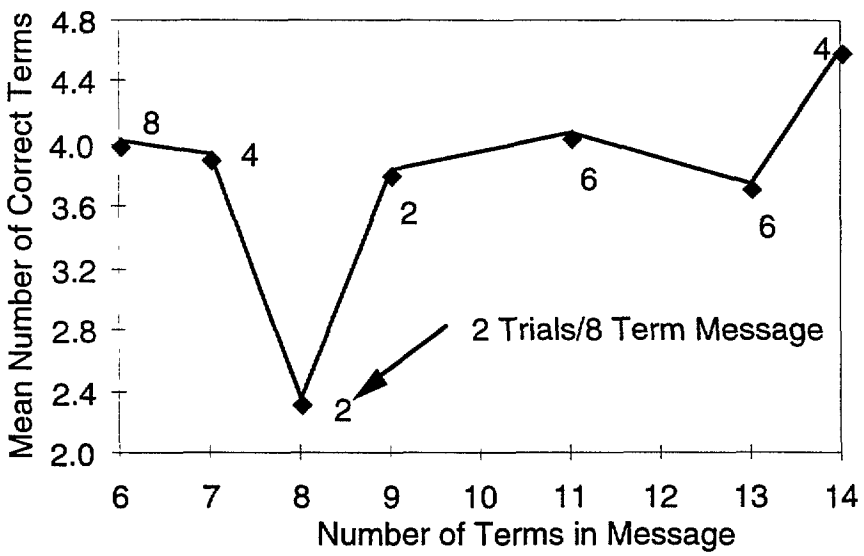


Figure 12. Mean number of terms recalled for each message length.

Table 10 shows the probability that each term was recalled correctly. Because a limited set of real messages were evaluated, term type and location in each sequence were not counterbalanced. This is important in considering recall for specific terms as

location effects were important. Nonetheless, the road, crossroad, incident/cause, and terms “mile back-up” and “hazardous material spill” were more often recalled than other elements of a message, indicating their relative importance.

Table 10. Recall of various message terms.

	Message Length							
Item	14	13	11	9	8	7	6	Mean
road	69.5	57.8	61.3	62.5	73.4	84.4	85.9	70.7
direction	31.3	35.4	39.4	29.7	34.4	44.5	56.3	38.7
crossroad	59.4	50.5	59.4	65.6	45.3	75.8	82.4	62.6
number of cars	22.7	15.6						19.1
"serious"	22.7							22.7
"injury"	16.4	08.3						12.4
"accident"	49.2	43.2						46.2
incident/cause			63.1				55.1	59.1
first service			33.8					33.8
second service			30.0					30.0
"hazardous material spill"				64.1				64.1
"cleanup"				35.9				35.9
"crew"				25.0				25.0
"shoulder work"					25.0			25.0
"continuing"						25.0		25.0
"construction"						25.0		25.0
number of lanes						25.0		25.0
lane description	17.2	11.5	11.9		06.3			11.7
"lanes blocked"	19.5	16.7	16.9		09.4			15.6
"police"	28.9	26.6						27.7
"EMS"	21.1	21.9						21.5
location	25.0	20.8	23.1	26.6				23.9
number of miles	28.9	24.0	37.5	21.9	10.9		42.2	27.6
"mile back-up"	46.9	38.5	54.4	48.4	26.6	60.2	75.4	50.0

Question: What factors affected the recall of terms?

To examine the terms recalled in a single traffic message, a repeated-measures analysis of variance was computed (Table 11). The type of information repeated by subjects was found to be significant ($p = 0.0001$) with subjects providing a mean of 0.3 additional (or incorrect) terms versus 3.9 correct terms.

Table 11. ANOVA independent variables for message relevance.

Factor	Levels	Comments
age group	2	young, old
gender	2	male, female
time of day	2	morning, afternoon
message characteristics	8	compact variable-combined message relevance (relevant, irrelevant), audio quality (good, poor audio), number of messages/trial (1.2.3)
response type	2	of message characteristics compact variables
type of information	2	additional, correct

Older subjects (mean = 1.7 terms) repeated significantly less information (p = 0.0065) than younger subjects (mean = 2.5 terms), a 40 percent difference. Additionally, the age by information type interaction was significant (p = 0.0001). Younger subjects provided more correct information than older subjects while older subjects falsely provided slightly more additional (incorrect) terms (see Figure 13).

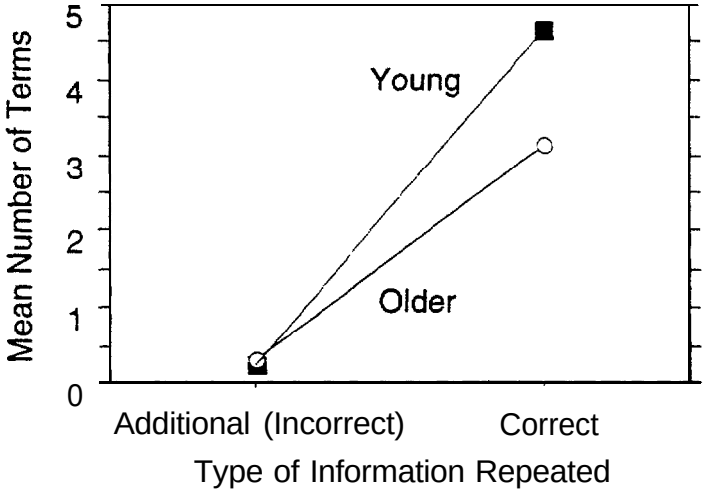


Figure 13. Mean number of terms provided by young and older subjects.

Subjects generally recalled significantly (p = 0.0002) more correct terms for good quality messages (2.2) than for poor messages (2.0) but the opposite for poor quality messages. (See Figure 14.) This interaction was significant (p = 0.0001).

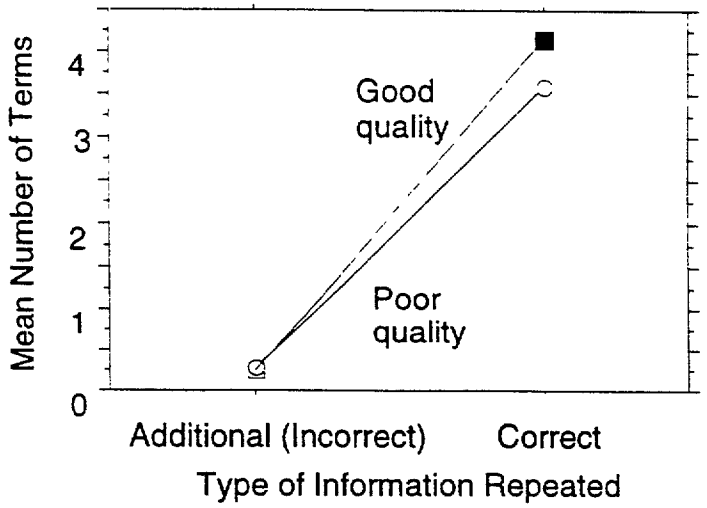


Figure 14. Mean number of terms provided during poor and good quality messages.

The same pattern occurred for message relevance. Recall of relevant messages was significantly greater (by about 32 percent) than recall of irrelevant messages for correct information but not for additional (incorrect) information, with the interaction ($p = 0.0001$) being the predominant result (Figure 15).

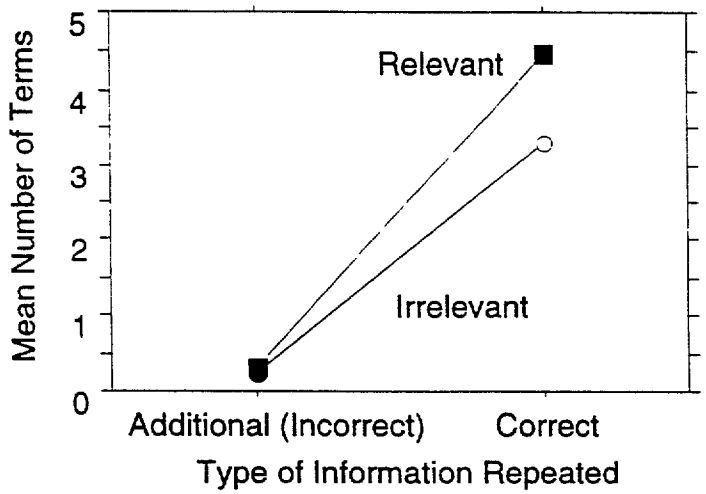


Figure 15. Mean number of terms provided during irrelevant and relevant messages.

Hence, the order of effects (by size) affecting recall were driver age, message relevance, and audio quality, all of which had practical impacts on recall. In addition, the particular message element and the location of the message in the sequence were practically important.

Ratings of Safety, Ease of Use, Usefulness, and Willingness to Pay

Question: Relative to other in-vehicle tasks, how easy and safe to use do drivers consider auditory traffic information systems to be?

Table 12 shows baseline post-test survey data for the difficulty of various driving tasks from this experiment along with a previous on-road experiment concerning the Ali-Scout navigation system (Katz, Fleming, Green, Hunter, and Damouth, 1997) and an on-road experiment concerning the UMTRI prototype navigation system (auditory, instrument panel, and HUD versions). Notice that subjects in this experiment tended to rate baseline tasks as more difficult than subjects in previous experiments. The previous study contained a smaller proportion of older drivers (3 age groups, not 2) and older drivers tended to rate tasks as more difficult.

With regards to the safety ratings (Table 13), there were no significant differences (in an ANOVA) due to age, gender, or time of day, but there were differences between questions ($p = 0.0047$). In general, participants did not consider the traffic messages to be very distracting. Notice that the auditory traffic information system was rated as much safer than either the Ali-Scout or UMTRI interfaces, even when the shift in the baseline is considered.

Table 12. Mean difficulty ratings for common (baseline) driving tasks.

Common driving task **Not difficult 1 -> 10 Extremely difficult**	**This experiment (n=32)**	**Katz, et al., 1997 (n = 54)**	**Green, et. al., 1993 (n=33)**
reading the speedometer	1.7	1.7	1.7
drinking a beverage	5.0	3.5	3.3
reading street names	5.0	3.2	--
adjusting the fan speed on the car heater or air conditioner	2.9	2.3	1.6
Mean	**3.7**	**2.7**	**--**

Table 13. Mean safety ratings.

Safety rating **Strongly disagree 1 -> 10 Strongly agree**	**This experiment (n = 32)**	**Katz, et al., 1997 (n= 54)**	**Green, et al., 1993 (aud/IP/HUD) (n=13/14/16)**
1. It is safe for me to use this system while driving.	8.7	**4.0**	**5.0/4.7/4.5**
2. It is safe for an inexperienced driver to use this system while driving.	5.9	**2.8**	**3.7/3.0/2.7**
3. The traffic messages/Ali Scout were/was not distracting.	**6.8**	**3.9**	**--**

Question: How useful do drivers consider such systems to be and what is the relative usefulness of each information element? Would drivers use such systems if they had them?

Information usefulness ratings concerned the utility of the data provided for choosing an alternate route. The most highly rated pieces of information included the intersection, the direction of travel, the number of lanes blocked, and the mile back-up information (Table 13). Information concerning police, EMS, or cleanup crews was rated as least useful. Older subjects tended to provide higher ratings (p = 0.000l), with an ANOVA of these data also showing an unexplainable age by time of day interaction (p = 0.0187) and a subject by question interaction (p=0.0053).

Table 14. Mean usefulness ratings for traffic information.

Information usefulness rating Not at all useful 1 ---> 10 Extremely useful	**Mean young rating (n = 16)**	**Mean older rating (n = 16)**
intersection	9.5	9.1
direction (N, S, E, or Westbound)	9.0	9.8
accident/incident type	4.9	6.8
number of lanes blocked	7.3	9.6
police, EMS, or clean up crew information	3.6	5.6
mile back-up information	8.6	9.3
Mean	7.2	8.4

As shown in Table 14, subjects felt that the traffic information would be helpful when driving in a familiar area, but less so in an unfamiliar area. Hence, the primary benefit will be to commuters. As with the usefulness ratings, older subjects gave higher ratings for use (p = O.OOOl), with an ANOVA again showing an unexplained interaction with time of day that drivers were tested (p = 0.0311). (See Figure 16.)

Table 15. Mean system use ratings.

System use rating Strongly disagree 1 ---> 10 Strongly agree	**Mean overall rating (n = 32)**
1. I would likely use this system for my daily travel.	7.3
2. I would use this system if I was in a hurry.	8.3
3. The traffic information provided is useful.	8.9
4. I would rather use this system than a traffic news report.	7.6
5. The traffic information would be useful when driving in an unfamiliar area.	5.8
6. The traffic information would be helpful when driving in a familiar area.	9.2

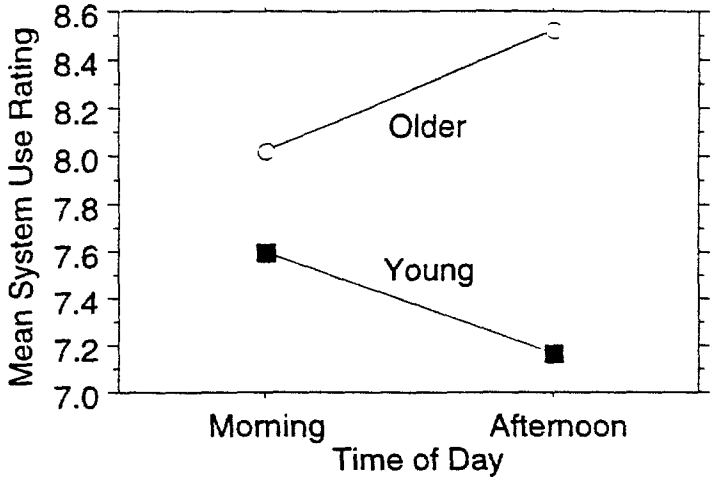


Figure 16. Mean system use ratings for age group by time of day.

In terms of audio quality (Table 15) a majority of subjects agreed that it affected their understanding of the traffic information. More drivers had trouble distinguishing I-75 from I-275 than I-96 from I-94, all real interstate highways in metropolitan Detroit.

Question: Did drivers think audio quality affected message understandability?

Three questions in the post-test survey concerned audio quality (Table 15). As a reminder, the conditions examined were (1) high quality, professionally spoken messages and (2) messages containing static and background noise representative of a reasonable worst case while driving. The majority of subjects agreed that the audio quality affected their understanding of the traffic messages (16 yes (50%), 14 somewhat (44%), 2 no (6%)). More subjects had trouble distinguishing I-75 from I-275 than I-94 from I-96. However, the majority of subjects did not have trouble distinguishing between either pairs of route numbers.

Table 16. Number of driver responses to audio quality questions.

Question	Yes	Somewhat	No
Did the audio quality affect your ability to understand the information provided by the traffic messages?	16	13	2
Did you have trouble distinguishing between I-75 and I-275?	5	9	18
Did you have trouble distinguishing between I-96 and I-94?	1	3	27

Question: How much were drivers willing to pay for such systems?

Figure 17 shows how much subjects would pay for a traffic information system similar to that evaluated. For the 24 subjects that responded to this question, the mean value was $117.71. The most common response was $0.

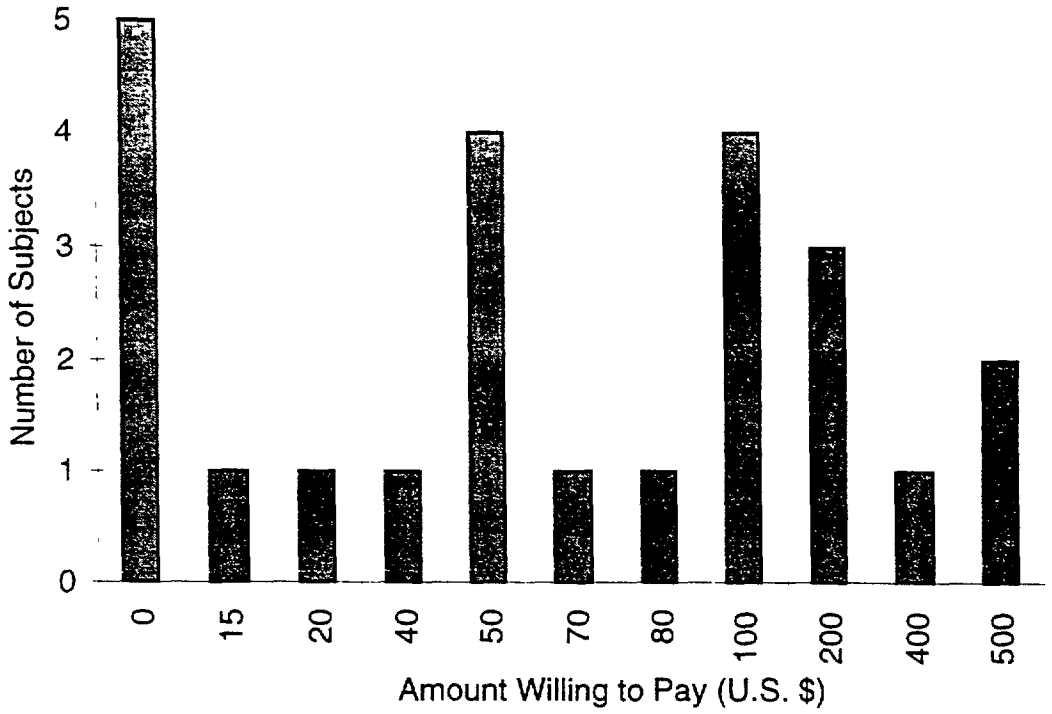


Figure 17. Amount subjects were willing to pay for a traffic information system.

Driving Performance and Message Characteristics

Question: How does driving performance (mean speed, speed variance, mean headway, headway variance, lateral position variance) and driver control inputs (throttle variance, steering wheel angle variance) vary with (1) message characteristics (message length, relevance, and quality) and (2) driver characteristics (age and sex).

Data Reduction

All data channels were sampled at a fixed rate, either 10 or 30 Hz. Hence, each data point describes the status of the vehicle for 1/10 or 1/30 of a second. For channels sampled at 10 Hz (all low bandwidth signals), the "missing" data was filled in by assuming that every three successive samples were identical.

Each of the two test runs (I-94 west, I-94 east) lasted about 30 minutes, and the data file for the 13 measures for each run averaged just over 4 MB. There were 96 trials for each of the 32 subjects. Because the saved files were larger than can be accommodated by Excel, means and standard deviations for each measure of interest were computed for subsequent statistical analyses using custom software.

Each file was split by trial and then by segment (baseline, message, response) as shown in Figure 18 for a sample audio message. Using baseline data from seconds before the message provides a control condition that minimizes differences due to time of day, subjects, traffic, and other factors.

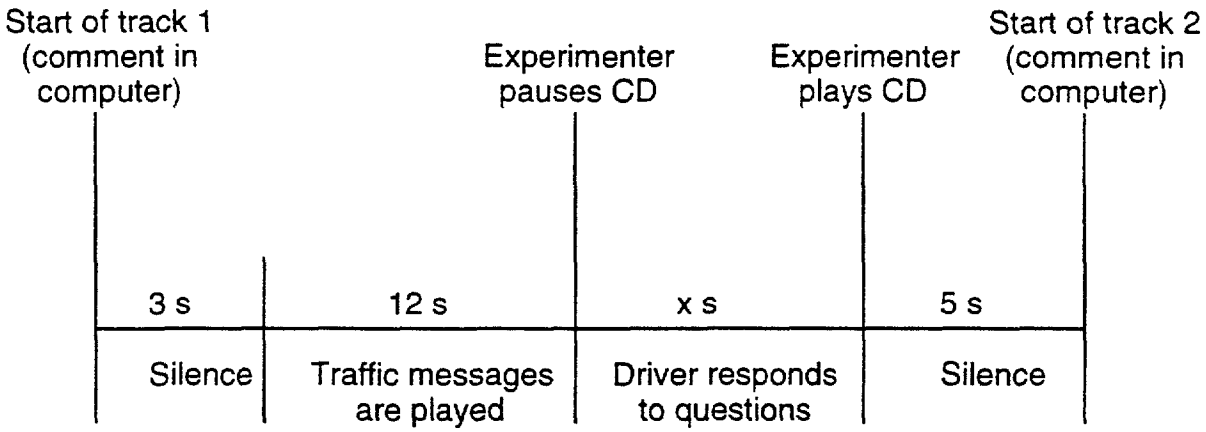


Figure 18. Example of how trials were sectioned for analysis.

Since both the start and end time were recorded into the computer file by the experimenter and, the exact time for each trial's audio messages was known, the response time for each trial could be computed. For example, if the start time for track one was at data point 2300 and the end point was marked at data point 3500 (1200 data points later), the time difference would be 1200/30 Hz or 40 s. Allowing 8 s for silence and 12 s for traffic messages, the actual response time would be 20 s.

Prior to detailed analysis, the speed signal was smoothed by the data parser to remove occasional signal drop outs from the wheel sensor (indicating instantaneous 2 mi/hr decreases and increases in speed). The analysis could be improved by equalizing the duration of the 3 intervals of interest.

Measures of driving performance were divided into 2 categories: longitudinal control (the position of the vehicle along the length of the road) and lateral control (lane position). Longitudinal measures include speed, throttle position, and headway with means and standard deviations for each. Lateral control measures included the mean and standard deviation of lateral position and steering wheel angle variance. Each dependent measure was analyzed in a manner similar to the dependent measures described previously using a repeated measures ANOVA though the variable trial segment was added. (See Table 16.) Although many 3- and 4-way interactions were found to be significant, only significant main effects and 2-way interactions are discussed in the results section of this report.

Table 17. ANOVA independent variables for driving data

Factor	Levels	Comments
age group	2	young, old
gender	2	male, female
time of day	2	morning, afternoon
message characteristics	8	compact variable-combined message relevance (relevant, irrelevant), audio quality (good, poor), number of messages/trial (1,2,3)
response type	2	of message characteristics compact variable
trial segment	3	baseline, message, response

Summary of the driving results

Table 17 summarizes the driving data results, with a more detailed discussion of the results and a complete ANOVA table appearing in Appendices G and H. The results are grouped by independent variable, with the Appendix containing them grouped by dependent variable. In general, the pattern was for the main effects of message characteristics to be significant as well some interactions of message characteristics with each other. However, there were few instances in which subject-related variables (age, sex) were significant or subject-related variables interacted with message characteristics. The major individual differences were that younger subjects drove faster than older subjects (68.5 versus 64.7 mi/hr) and men drove faster than women (67.6 versus 65.5 mi/hr).

In this experiment, manipulations of messages were expected to have a small impact on driving performance. That proved to be the case. Message relevance had significant impacts on the standard deviation of the range (headway), the standard deviation of lane position, and the associated driver measure, the standard deviation of steering wheel angle. Figure 19 is indicative of the results obtained, with relevant messages leading to less lane variability, the opposite of what one would expect, since relevant messages demand attention and distract one from driving.

Table 18. P-values for all driving performance measures.

Independent Variables	Dependent Driving Performance Measures						
	Mean speed	SD of speed	SD of throttle	Mean range	SD of range	SD of lane position	SD of steering wheel angle
Age Group	+++						+
Gender	+						
Time of Day							
Relevance			+		+	+++	++++
Quality		++++	++	++		++++	++++
Number			++++	++	++++	++	++++
Repetitions	++++		++		++++	++++	++++
Segment (within trial)		++++					++
Age Group * Gender							
Age Group * Time of Day							
Gender * Time of Day							
Relevance * Age Group							
Relevance * Gender							
Relevance * Time of Day						+	
Quality * Age Group							
Quality * Gender							
Quality * Time of Day							
Repetitions * Age Group							
Repetitions * Gender							
Repetitions * Time of Day	+						
Number * Age Group							+
Number * Gender							
Number * Time of Day							
Segment * Age Group	++	+					
Segment * Gender							
Segment * Time of Day							
Relevance * Quality		++	++++	++++	++++	++++	++++
Relevance * Repetitions	++++	++++	++++		++++	++++	++++
Quality * Repetitions	++++	++++	++++	+++	++++	++++	++++
Relevance * Number	++	+	++++	+	++++	++++	++++
Quality * Number	+++	++	++++	+	++++	++++	++++
Number * Repetitions	++++	++++	++++	++++	++++	++++	++++
Relevance * Segment							+
Quality * Segment							
Number * Segment			+++		+++		+++
Repetitions * Segment			+				

	+	++	+++	++++
p ≤	0.05	0.01	0.001	.0001

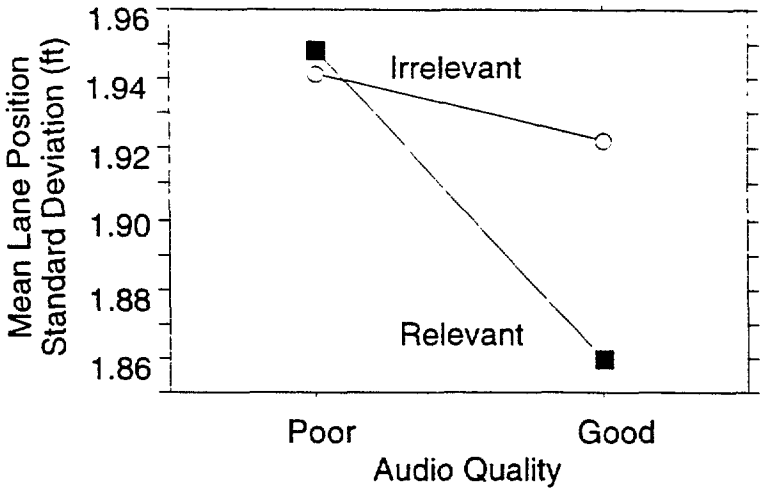


Figure 19. Mean lane position standard deviation for message quality by relevance interaction.

The differences in message quality generally had significant effects on driving performance. As shown in Figure 19 previously, lane variance was reduced when good quality messages were provided. Also, subjects tended to drive at a less variable speed (Figure 20).

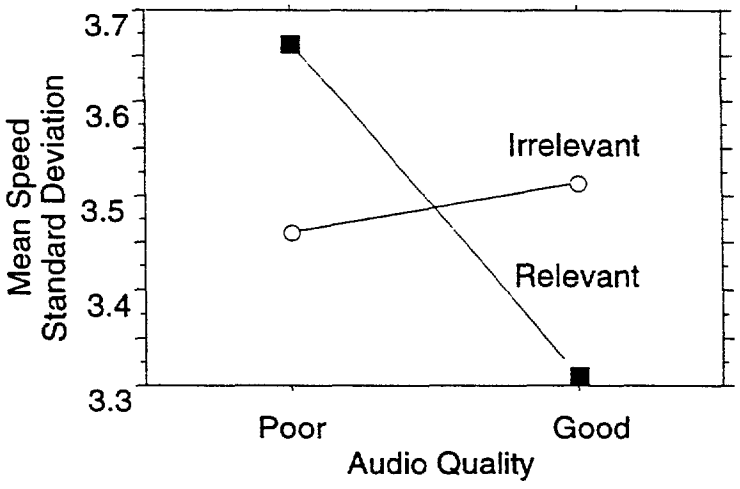


Figure 20. Mean speed standard deviation for message quality by relevance interaction.

The effect of the number of messages on performance were difficult to explain. The number of messages was generally significant, but the results for 2 messages usually did not fall in the range between 1 and 3. Figure 21 is an example.

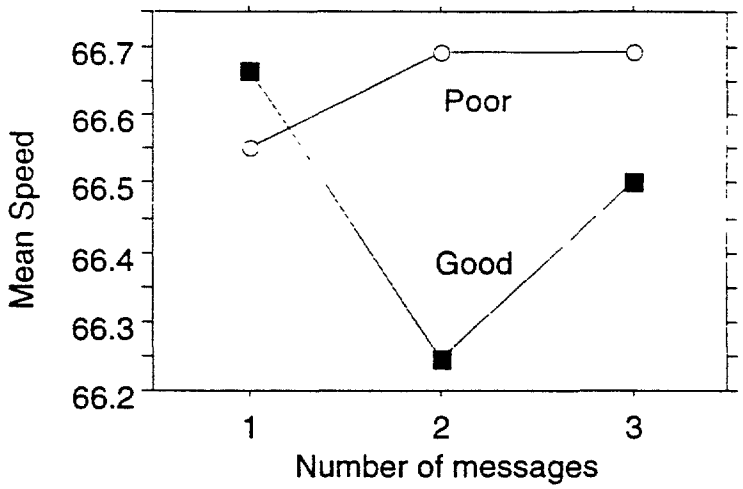


Figure 21. Mean speed for the audio quality by number of messages interaction.

In this experiment, there were consistent effects for the number of repetitions on all measures of driving performance except the standard deviation of speed and mean range. However, there were few interactions of repetitions with other variables, suggesting that learning occurred quickly in the first few trials. Figure 22 shows one of the few instances (for mean speed) where interactions with repetitions occurred.

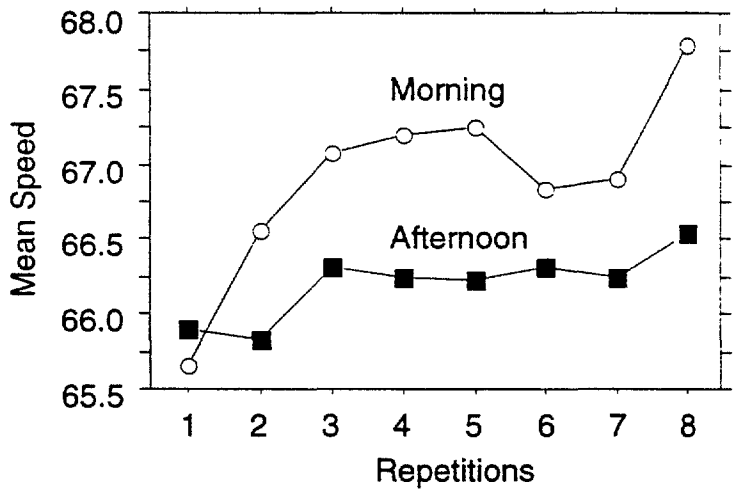


Figure 22. Mean speed for the repetition by time of day interaction.

Finally, there were two instances in which the segment had an effect on driving performance. When drivers became overloaded, say by an in-vehicle task, one potential response is to slow down. In this experiment, the result would be a drop in mean speed (which did not occur) and change in speed variance. Speed variance could decrease when the mean speed dropped since the mean and variance are correlated. An increase would be expected when behavior varies considerably from trial to trial. Here, notice that just listening to a message increased speed variability (though only slightly), and speaking to the experimenter, even only for a few seconds, further increased speed variability (Figure 23).

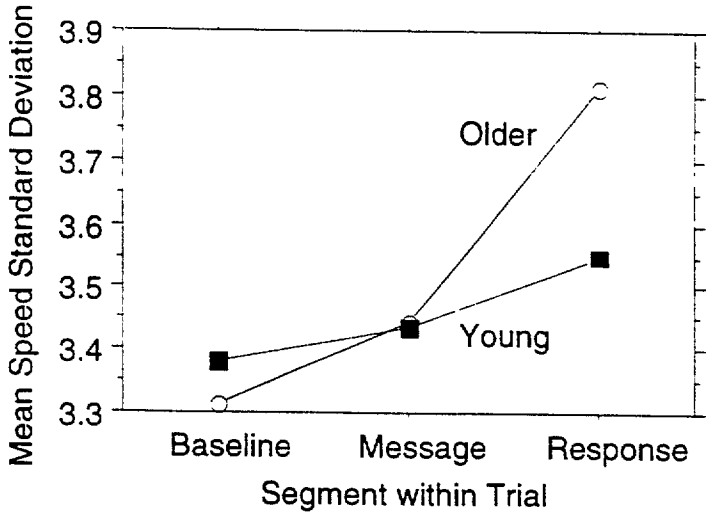


Figure 23. Mean speed standard deviation of segment by age group interaction.

CONCLUSIONS

How Well Did Drivers Recognize Relevant Messages?

As a reminder, drivers were presented with a representative set of messages that they might receive on a trip to Detroit. The experiment was conducted while driving on a real road. The first step in responding to a traffic message was to determine if the message was relevant. Message characteristics that were manipulated included the number of messages per trial, their relevance, and audio quality.

In this experiment, the number of messages (up to 3) had only a small impact on identification of relevant messages, with just over 85 % of the messages being correctly recognized as relevant. However, in many metropolitan areas, including Detroit, drivers will often hear more than 3 messages, and often more than 1 is relevant.

Also having an impact on the recognition of messages was the message quality-relevance interaction. The fewest number of correct responses was for good audio quality relevant messages, though the size of the effect was small. The effects of message relevance or quality were not significant. The authors have no explanation for this result.

Subject differences were primarily age-related. Younger subjects identified about 89 percent of the messages correctly versus 82 % for older subjects. The difference between age groups became more apparent as more messages were presented (9 % for 3 messages). Driver age was one of the largest sources of variability identified.

Recognition was affected by the time of day that subjects were tested, with better performance in the afternoon. If anything, traffic may have been slightly heavier during the afternoon session. It could be this reflects a difference in alertness, though by late morning (10 a.m. - noon), drivers should be fully alert.

Thus, the overall impact of message and driver characteristics on the recognition of messages were small at best, with driver age being the primary factor. Although large message sets were not explored, it does not appear that recognition is a concern for drivers familiar with a metropolitan area.

How Well Did Drivers Recall Information from Traffic Messages?

For single message trials, drivers were asked to recall the message content. If traffic messages are to be useful, drivers must be able to remember relevant messages. Overall, drivers recalled about 4 terms per message with a few exceptions (4.5 for 14-term messages, fewer terms for B-term messages). Recall rates were 43 % for good quality messages and 37 % for poor quality. This level of performance is consistent with both the human performance literature in general and prior studies of traffic-message recall reported in the introduction to this report. Consistent with the recall data, almost all drivers thought audio quality influenced their performance at least to some degree as expressed in the post-test survey.

For almost all messages, the same items were usually recalled with drivers behaving rationally to recall the most important items within the limits of their capabilities. Some 60-70 % of the time, drivers remembered the intersection (road and crossroad) of a message, but often not the direction of traffic that was affected (39 %). The incident or cause of traffic (e.g., accident, construction, etc.) was recalled with similar accuracy. Drivers recalled if there was a back-up with approximately 50% accuracy. Except for hazardous material spills, other information (the exact length of the back-up, if lanes were blocked, the lane description, or descriptions of emergency services and cleanup crews) were recalled one-third to one-fifth of the time. This failure to recall the full content of a message reflects the limits of human short-term memory. Those limits were also reflected in the pattern of recall, with drivers remembering the first few items presented (primacy) and the last few items (recency) more often than items in the middle of the message.

During recall, drivers sometimes provided information that was not in the original message (on average, one-fourth of an item), with poor audio quality elevating the false report rate slightly for messages of 6 to 14 items. While not explicitly evaluated, comments from drivers suggested that they were selective in what they remembered, tuning out messages not relevant to their route as soon as they realized such.

Even more than recognition, recall performance was influenced by driver age. While false reports of additional information were comparable, young subjects recalled about 16% more correct information than their older counterparts. Driver sex or time of day, did not affect the recall of traffic information.

This research reinforces the idea that people have significant limitations in their ability to remember dynamic information, in this case about 4 terms pertaining to traffic. While improving message quality helped, its impact was small. If drivers are expected to be able to recall and effectively utilize traffic information, presentation modalities other than auditory should be considered. If it is necessary to use the auditory modality, presenting lengthy, detailed messages, while comforting to drivers, may be of limited value in making driving decisions because drivers can only recall a limited portion of each message. Accordingly, presenting great detail may warrant reconsideration.

Did Drivers Believe It Was Safe for People to Listen to Traffic Messages While Driving?

Drivers generally felt that the system was safe for themselves to use while driving (8.7 on 10 point scale where 1 = strongly disagree and 10 = strongly agree). There was some agreement that the messages were not distracting (6.8 on the same 1 to 10 scale). The system was rated as much safer for themselves or others to use than existing or proposed navigation systems. Subjects did not believe that it was as safe for inexperienced drivers to use the traffic information system. This suggests that the traffic information interface investigated was safe for experienced drivers, The situation for novices is uncertain and may be worth investigating.

How Useful Was the Traffic Information System?

Overall, subjects thought the traffic information was quite useful (mean = 8.9) and they would use the system when in a hurry (mean = 8.3). To a lesser extent, subjects moderately agreed that they would use the system for daily travel (mean = 7.3), and that they would rather use the system than a traffic news report (mean = 7.6). Subjects felt that the traffic information would be useful when driving in a familiar area (mean = 9.2 on a 10 point scale where 1 = strongly disagree and 10 = strongly agree). Subjects did not feel that the traffic information would be as useful, however, in an unfamiliar area, emphasizing the need for local area knowledge (something commuters have but visitors lack).

How Useful Were the Traffic Information Elements?

Subjects were asked to rate how useful each piece of information provided in the traffic messages would be in choosing an alternative route. Each piece of information was rated on a 10 point scale where 1 = not at all useful, and 10 = extremely useful. The highest rated terms in the traffic message were the intersection (road and crossroad), the direction of traffic affected, and the mile back-up information, all with mean ratings of approximately 9.0 or above. The number of lanes blocked was also rated as useful (8.4). The accident/incident type was rated as somewhat useful (5.8), and the emergency or cleanup crew information was rated as least useful (4.6). These ratings reflect findings from the recall task described earlier. Items receiving high ratings were more likely to be recalled, though there were several exceptions. The lanes blocked and the direction of travel affecting traffic were rarely recalled by subjects, but were rated as useful. Additionally, subjects rated the incident/accident type as only moderately useful, however, it was recalled with great frequency. This indicates that drivers may have a different sense of what is important and what is actually recalled.

How Much Were Drivers Willing to Pay for a Traffic System?

One of the best measures of the usefulness of a system is how much people are willing to pay for it, a question appearing in previous UMTRI studies concerning driver information systems (Green, Williams, Hoekstra, George, and Wen, 1993; Katz, Fleming, Green, Hunter, and Damouth, 1997). An alternative approach, asking drivers what they would give up instead, was not explored to save time.

The mean amount drivers were willing to pay for a traffic information system similar to the one used in the experiment was $117 with responses ranging from $0 to $500. The most common response observed was $0 (free), indicating that many subjects did not find the system useful enough to pay for it. This may be because information of this type is already available for free on the radio.

Did Use of the Traffic Information System Affect Driving Performance and What Were the Effects of Various Message Characteristics?

Audio Quality

For the most part, poor audio quality degraded driving performance. The speed standard deviation (3.6 mi/hr) was slightly greater for poor quality audio messages than during good audio (3.4 mi/hr). Similarly, throttle variance was also greater for poor messages. In this case, differences in the standard deviation of speed were not the result of unique, individual driving patterns as there were no aggressive drivers, few lane changes, large headways, and no stop-and-go driving. Lateral control was also degraded while listening to poor quality messages (standard deviation of lane position = 1.95 ft for poor quality messages and 1.89 ft for good quality messages). Paralleling the speed-throttle data, increased lateral variability was associated with increased steering wheel angle variability. Speed and lateral variability reflect erratic driving and can be the cause of crashes.

These findings are supported by the hypothesis that the poor quality messages increased driving workload, causing drivers to maneuver slightly more erratically. Further support for this hypothesis comes from the headway data, where mean headway increased from 204 to 211 feet when poor quality messages were presented, an attempt by drivers to increase the cushion around themselves when they were less able to deal with external demands. Curiously, speed actually increased (but not significantly) when poor quality messages were presented. That may be a result of individual differences and the fact that the mean and standard deviation of speed are correlated. Normally, when workload increases, drivers slow down. Nonetheless, the authors still support the workload explanation of the poorer driving observed as a result of just presenting messages that were more difficult to hear. The authors believe this is the first time evidence has appeared in the literature demonstrating the safety implications of auditory message quality.

Number of Messages

The results for the effects of the number of messages do not make sense. One would expect that either (1) workload would increase and driving performance would decrease in proportion to the number of messages, or (2) that increasing the number of messages would have no impact on driving workload and driving performance. Given the effect of message quality, the degraded driving explanation would be the better a priori choice. However, in this experiment, driving performance was generally poorest for 2 messages. That condition had significantly more throttle variance, more (but not significantly more) speed variance, slightly more lateral variance, and slightly more steering wheel angle variance. Headway was also greater for this condition, but only by 4 feet. Given the very small differences found, the authors attribute these differences to chance variation.

Message Relevance

One would expect that driving performance would be worse when relevant messages are presented than irrelevant messages because relevant messages need attention.

In fact, poorer driving performance was observed during irrelevant messages. Irrelevant trials had greater lane variance, more steering wheel angle variance (both significant), more throttle variance, and a greater headway variance (all differences that were slight but significant). There were no significant differences in mean speed, the standard deviation of speed. Explanations for this contraindicated finding are desired.

Segment Differences

Readers are reminded that each trial was divided into three consecutive sections each several seconds long: (1) baseline driving, (2) message presentation, and (3) message recognition/recall (during which the subject and experimenter talked). In the baseline segment, subjects just drove, that is they engaged in “plain old driving.” Differences between segments had less of an impact on driving performance than any of the message characteristics examined. Significant increases were noted in the standard deviation of speed and the standard deviation of steering wheel angle. These differences were primarily due to degradation in the response segment, though some degradation appeared in the message presentation segment as well.

In other words, assuming there were no differences due to segment length, just listening to messages sometimes degraded driving, though only slightly, with speaking having a slightly greater impact. Engineers developing new information systems for motor vehicles have assumed that voice-based systems (e.g., Auto-PC, Car-PC) have no safety impact. The evidence here suggests otherwise. If anything, the effects of listening to messages were underestimated. When driving, failing to steer results in yaw errors, which in turn leads to lateral position errors. The passage of some time is required for lateral error to accumulate. Thus, failing to steer while listening to messages (segment 2) could result in errors not appearing until segment 3. This explanation, also viable for speed corrections, could be checked by varying the time between message presentation and recall.

Thus, of the message characteristics examined, message audio quality had the largest adverse impact on performance, with message length and the number of messages having effects on a few occasions, an outcome the authors would not have suggested prior to the study. The specific findings reported ((a) relevant messages improving driving and (b) driving being worst when 2, not 3 messages were presented) defy explanation, Finally, the segment differences reported suggest that the proposal that auditory information system information exchange, both drivers listening to system messages and speaking commands all have some impact on driving. They are not zero workload additions as some have suggested. The extent of the safety impact needs further examination.

Individual Differences

Relative to previous UMTRI studies (Green, Hoekstra, and Williams, 1993; Green, Williams, Hoekstra, George, and Wen, 1993; Katz, Fleming, Green, Hunter, and Damouth, 1997; Green, Fleming, and Katz, 1998), there were very few individual differences in driving performance. The primary individual differences were of age, and those differences were reflected in mean speed, as was the case in previous

studies. (Young subjects drove faster on average (68.5 mi/hr versus 64.7 mi/hr). In addition, men drove faster (67.6 mi/hr versus 65.6), especially the young men. The only other individual difference noted was that young subjects had a slightly greater steering wheel angle standard deviation (2.7 degrees versus 2.4 degrees). Given the few times passing occurred, this difference is probably not due to differences in lane-changing behavior.

What Should Be Done in Future Studies?

1. improve the measurement of audio quality

In this experiment, audio quality was either "good" or "poor." There exists a significant literature concerning the specification and assessment of speech using various word lists and other techniques. (See, for example, Sanders and McCormick, 1993.) In future studies, resources should be provided so audio quality can be quantified using those methods. Also, additional levels of variation should be explored. This will allow for a precise specification of audio quality necessary to achieve various levels of driving performance.

2. Record and analyze lane-change behavior

In this experiment, changing lanes was uncommon. Accordingly, while driver maneuvering was recorded on videotape, lane changes were not specifically examined to keep the analysis effort within the resources available. Changing lanes adds to the driver's workload and alters measurements of lane variance and steering wheel angle variance. It may be possible to detect and filter out lane changes by performing additional analysis of the lane position data.

3. Explain why driving was poorest with 2 messages

Driving was consistently poorer when drivers were presented 2 messages, as opposed to 1 or 3. There is no explanation for this outcome.

4. Examine other modes of conveying traffic information

The research reported here has broad applicability to all of the traffic information systems under consideration -- AHAR, LPHAR, cellular call in, and RBDS. Of these systems, however, cellular call in has other unique safety concerns in that subjects will be dialing in and potentially interacting with menu systems while driving (Manes and Green, 1997), by either using a phone keypad or a voice recognition system While there has been some research on the use of phones while driving (Goodman, Bents, Tijerina, Wierwille, Lemer, and Benel, 1997), that topic, especially in this context, is in urgent need of additional research. Both driving simulator and on-the-road experiments using instrumented vehicles should be considered.

5. **Develop improved explanations of how drivers react to workload and measures sensitive to workload.**

In designing an experiment, important decisions are (1) which dependent measures to emphasize, either driver inputs (standard deviation of steering wheel angle, standard deviation of throttle) or outputs (standard deviation of lane position, standard deviation of speed), and (2) which measures in each category to select.

First, the input measures are most responsive. When the driver modulates the throttle, only some time later is the vehicle response noted. However, when events of interest (listening and responding to messages) and their corresponding baseline data are in close temporal proximity, such as in the case here, even the benefits of responsiveness of the input measures can be lost if the actions in one time period only become apparent in another period. This may suggest misleading conclusions about cause and effect. In this experiment, introducing delays between events would have permitted that hypothesis to be checked.

Second, the vehicle acts as a filter, removing moment-to-moment subject variability in the data. When the conditions being compared are sessions on different days, output measures are more likely to be statistically significant than input measures because excess variability has been removed. In this case, the benefits of filtering were not realized, and in fact, higher levels of significance were achieved for input measures.

Third, when overloaded by a secondary task, drivers either try to abandon or reschedule the added task. If that is not possible, then drivers slow down to reduce the workload of driving. Staying in the lane remains a protected task. Only when the workload is very high will diminished attention be paid to steering. Accordingly, speed should be affected first by workload and only when workload is high should lane variance increase. Therefore, mean speed and speed variance (and the associated throttle measures) should be more sensitive indicators of workload than lane variance. In this experiment, the opposite was true, lane position variability was significantly affected more often than speed variability. Explanations of how drivers react to workload may need further consideration. Such explanations are critical to the fielding of safe and easy-to-use driver interfaces for ITS products.

REFERENCES

Davis, J.R. (1989). Back Seat Driver: Voice Assisted Automobile Navigation. Unpublished doctoral dissertation, Cambridge, MA: Massachusetts Institute of Technology.

Davis, J.R. and Schmandt, C.M. (1989). The Back Seat Driver: Real Time Spoken Driving Instructions. In D.H.M. Reekie, E.R. Case, and J. Tsai (Eds.), First Vehicle Navigation and Information Systems Conference (VNIS'89), New York: Institute of Electrical and Electronics Engineers, pp. 146-150.

Dudek, C.L., Huchingson, R.D., and Brackett, R.Q. (1983). Studies of Highway Advisory Radio Messages for Route Diversion, Transportation Research Record #904, 4-9.

Dudek, C.L., Huchingson, R.D., and Stockton, W.R. (1981). Human Factors Design of Dynamic Visual and Audio Displays for Metropolitan Traffic Management. Volume 3, Highway Advisory Radio (Technical Report FHWA/RD 81-041), Washington, D.C.: U.S. Department of Transportation, Federal Highway Administration, (available from NTIS as PB 83 133611).

Gatling, F.P. (1975). Auditory Message Studies for Route Diversion (Report FHWA-RD-75-73), Washington, D.C.: U.S. Department of Transportation, Federal Highway Administration, (available from NTIS as PB 246 12519).

Gatling, F.P. (1976). The Effect of Auditory and Visual Presentation of Navigational Messages on Message Retention (Technical Report FHWA-RD-76-94), Washington, D.C.: U.S. Department of Transportation, Federal Highway Administration, (available from NTIS as PB 256 599/2).

Gatling, F.P. (1977). Further Studies of Auditory Messages (Technical Report FHWA-RD-77-130), Washington, D.C.: U.S. Department of Transportation, Federal Highway Administration, (available from NTIS as PB 279 858/5).

Goodman, M., Bents, F.D., Tijerina, L., Wierwille, W., Lemer, N., and Benel, D. (1997). An Investigation of the Safety Implications of Wireless Communications in Vehicles (Technical Report DOT HS 808-635), Washington, D.C.: U.S. Department of Transportation, National Highway Traffic Safety Administration,

Graham, R. and Mitchell, V.A. (1994). An Experimental Study into the Ability of Drivers to Assimilate and Retain In-Vehicle Traffic Information, 1994 Vehicle Navigation and Information Systems Conference Proceedings, New York: Institute of Electrical and Electronics Engineers, 463-468.

Green, P. (1992). American Human Factors Research On In-Vehicle Navigation Systems (Technical Report UMTRI-92-47), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Green, P., Fleming, J., and Katz, S. (1998). Driving Performance Evaluation of the Ali-Scout Navigation System: A Preliminary Analysis, Intelligent Transportation Society of America Eighth Annual Meeting Conference Proceedings [CD-ROM].

Green, P., Hoekstra, E., and Williams, M. (1993). On-The-Road Tests of Driver Interfaces: Examination of a Navigation System and a Car Phone (Technical Report UMTRI-93-35), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Green, P., Williams, M., Hoekstra, E., George, K., and Wen, C. (1993). Initial On-the-Road Tests of Driver Information System Interfaces: Examination of Navigation, Traffic Information, IVSAWS, and Vehicle Monitoring (Technical report UMTRI-93-32), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Katz, S., Fleming, J., Green, P., Hunter, D.R., and Damouth, D. (1997). On-the-Road Human Factors Evaluation of the Ali-Scout Navigation System (Technical Report UMTRI-96-32), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Manes, D. and Green, P. (1997). Evaluation of a Driver Interfaces: Effects of Control Type (Knob Versus Buttons) and Menu Structure (Depth Versus Breadth) (Technical Report UMTRI-97-42), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Reed, T., Hanafi, T., Richeson, R., and Underwood, S, (1998). Natural Use Study (Technical Report EECS-ITS LAB-DT98-001), Ann Arbor, MI: Intelligent Transportation Systems Laboratory, University of Michigan.

Richeson, R., Underwood, S., and Waldman, Y. (1998). Institutional Issues (Technical Report EECS-ITS LAB-DT98-005), Ann Arbor, MI: Intelligent Transportation Systems Laboratory, University of Michigan.

Ristenbatt, M. and Shahine, G. (1998). Technical Performance and Cost (Technical Report EECS-ITS LAB-DT98-002), Ann Arbor, MI: Intelligent Transportation Systems Laboratory, University of Michigan.

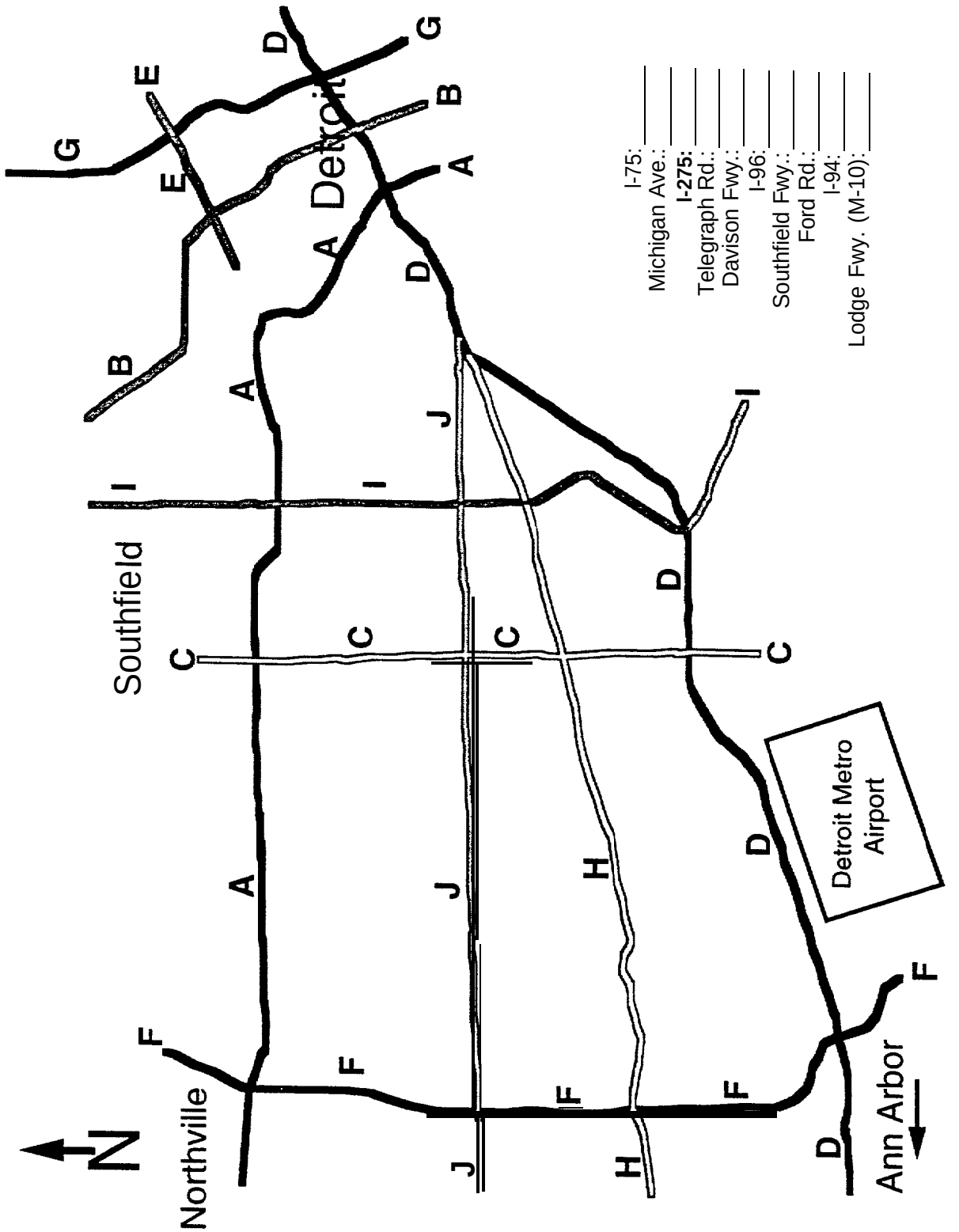
Sanders, M.S. and McCormick, E.J. (1993). Human Factors in Engineering and Design (7th ed.), New York: McGraw-Hill.

Turnage, H. C. (1980). Highway Advisory Radio, IEEE Transactions on Vehicular Technology, VT-29,(2), 183-191.

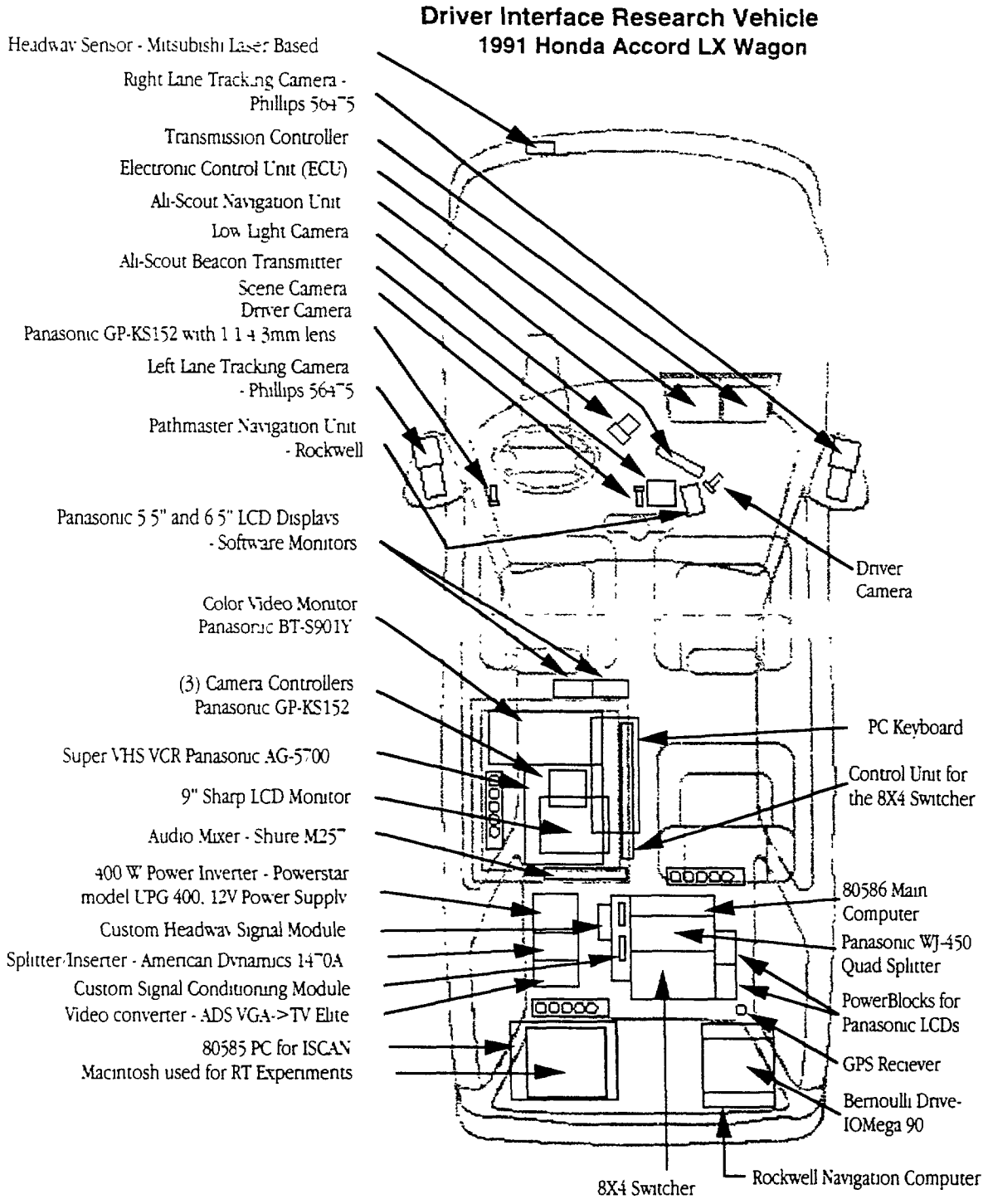
Turnage, H. C. (1981). Systems Analysis and design Guidelines for Highway Advisory Radio, Executive Summary (Technical Report FHWA-RD-80-176), Washington, D.C. U.S. Department of Transportation, Federal Highway Administration,

Underwood, S., Juma, O., Gurusamy, S., Hadj-Alouane, A., and Hadj-Alouane, N. (1998). Simulation and Modeling (Technical Report EECS-ITS LAB-DT98-002), Ann Arbor, MI: Intelligent Transportation Systems Laboratory, University of Michigan.

APPENDIX A - MATCHING TASK



APPENDIX B - TEST VEHICLE LAYOUT



APPENDIX C - INSTRUCTIONS

Hi, My name is (experimenter's name). Thank you for coming today. Let's go to the office and get started.

This study will consist of one on-the-road session that will take about two hours. You will be paid a total of $30 for your time. You will be driving an automatic transmission Honda Accord station wagon on major highways near the Ann Arbor/Dexter/Jackson area. Please follow ail traffic laws and speed limits while driving. if you do not comply with these safety measures, the experiment may be stopped by the experimenter.

Before we begin, I would like to give you a little background information concerning traffic information systems. Some new systems are being designed to interrupt radio broadcasts to provide up-to-the-minute information regarding traffic in the area. Messages may or may not be relevant to the particular streets that you will travel, however traffic at intersecting roads may cause you to experience congestion. Therefore, it is important that you are familiar with the route which you are driving and any related highways.

To determine how familiar you are with the Detroit area, please match the freeway names with the letters on this Detroit map.

During this experiment you will be driving on I-94 west towards Jackson. However, for the sake of experimentation, imagine driving on I-94 <u>east towards Detroit.</u> imagine that your destination is off of I-94 after the Southfield Fwy, but before Michigan Ave. it is important that you are familiar with the highway system in the Detroit area and especially i-94 so that you will recognize major intersections pertaining to your route.

During the experiment, traffic messages will be played. Following the last message in each set, you will be asked which, if any, of the messages are relevant to the route that you are pretending to drive. Relevant messages will concern the hypothetically driven road (e.g., I-94 <u>east</u> before Michigan Ave), or describe congestion on a road intersecting any part of the imaginary route (e.g., Telegraph Rd. and I-94) that might spill over traffic into the intersection. *Messages pertaining to the following* intersections *will be relevant to the route that you are imagining to drive: I-275 and I-94, Telegraph Rd and I-94, and Southfield Fwy and I-94.* Because you are imagining to drive on i-94 east, any messages pertaining to I-94 west are not relevant to your route. You should assume that while driving you never pass any of the relevant intersections and they will remain relevant throughout the entire experiment.

At various points during the experiment, I may ask you to repeat as much information as you can remember from the traffic message that you just

heard. So, pay attention to all information provided by the traffic messages, not just the names of intersecting highways.

Messages will not be repeated for you. If you can not hear a message due to outside traffic, or miss a message because you are concentrating on driving, make your best guess, or tell the experimenter and we will simply move on to the next trial.

Please remember that driving is your main priority, and recall of the messages and/or their relevance to the hypothetical route is secondary. Again, obey all traffic laws and speed limits while driving. It is also important that you do not use the cruise control option on the car while driving. You will be videotaped throughout the entire experiment, for analysis purposes. A camera and microphone are mounted to the windshield frame on your left.

Once again, complete this map of the Detroit freeway system by matching the highways/roads with their corresponding letter. Feel free to ask questions at any time.

Consent and Bio Forms

First, please read and sign this consent form, and then fill out the biographical form. If you have any questions feel free to ask them at any time.

At the test vehicle

Let me reiterate a few important points from the consent form. First of all, driving safely is your main priority. If you feel unsafe or unable to answer my questions during the experiment, please wait. Second, if you are uncomfortable or wish to stop at any time, please let me know right away. You will be paid regardless. You are expected to obey all speed limits and driving laws.

Let me stress a couple of points about driving with the traffic information system. Sets of 1, 2, or 3 traffic messages will be played through the car radio at various points during the experiment. Immediately following each set of messages, I will ask you which (if any) messages are relevant to the imaginary route that you are traveling. As a reminder, any message pertaining to I-94 eastbound and I-275, Telegraph Rd, or Southfield Fwy is relevant to your route. Bear in mind that this also includes traffic on any of these intersecting roads (I-275, Telegraph, or Southfield) at I-94. In addition, at various points during the experiment, I will ask you to repeat as much information as you can remember from the traffic message that you just heard.

I will play a music CD for you on the way to the starting point of the experiment. You are free to choose a CD from our collection. While you

are driving to your destination, feel free to adjust the volume to a level that you are comfortable with. Volume should only be adjusted during the experiment if absolutely necessary. But remember that your first priority is to drive safely and your second priority is to get information from the traffic messages. No music will be played during experimentation.

Because it is important for you to hear the whole message, and the messages are randomly timed, we ask that you not talk during the experiment unless it is necessary or when asked questions by the experimenter. If during the experiment you have any questions, please raise your right hand, and then I will pause the experiment at a convenient point.

For the last time, I will ask you to complete this highway map of Detroit by matching the highway names with the corresponding letter on the map.

Please fasten your seat belt, adjust the seat, mirrors, steering wheel height, as you feel necessary.

- Adjust the car seat, steering wheel height, and side and rearview mirrors.
- Fasten seat belt.
- Point out climate controls, the radio may not be operating during experimentation.
- Remind about following speed limit.

APPENDIX D - SAMPLE OF RECALL AND RECOGNITION WORKSHEET

Trial	Message number	Relevant	Not Relevant	Message Relevance
1	1			relevant
1	2			irrelevant
1	3			irrelevant
* 2	1			relevant
Telegraph N I-94 spill clean up crew en route 4 mile back-up				
3	1			irrelevant
3	2			irrelevant
3	3			irrelevant
4	1			irrelevant
4	2			irrelevant
5	1			irrelevant
5	2			irrelevant
* 6	1			irrelevant
I-94 W I-96 2 car injury police/EMS on scene 3 mile back-up				
7	1			irrelevant
7	2			irrelevant
7	3			relevant
* 8	1			relevant
I-275 S I-94 1 car minor accident.				
9	1			irrelevant
9	2			irrelevant
9	3			irrelevant
10	1			irrelevant
10	2			irrelevant
10	3			irrelevant
11	1			irrelevant
11	2			relevant
* 12	1			irrelevant
I-275 N I-96 multiple car injury police/EMS en route 3 mile back-up				
* 13	1			irrelevant
Lodge N I-94 continuous construction 3 mile back-up				
* 14	1			relevant
I-94 E Telegraph railroad derailment police/EMS on scene 5 mile back-up				
15	1			relevant
15	2			relevant
16	1			irrelevant
16	2			relevant

APPENDIX E - SUBJECT SURVEY

Subject # _____ Name _____

Post-Drive DIRECT Project Usability Survey

Information usefulness for choosing alternative routes

If you used this system to determine if an alternate route is necessary due to congestion, which of the following pieces of information would be useful?

	Not at all useful									Extremely useful
Intersection	1	2	3	4	5	6	7	8	**9**	10
Direction	1	2	3	4	5	6	7	8	9	10
Accident/incident type	1	2	3	4	5	6	7	8	9	10
Number of lanes blocked	1	2	3	4	5	6	7	8	9	10
Police, EMS, or cleanup crew information	1	2	3	4	5	6	7	8	9	10
Mile back-up information	1	2	3	4	5	6	7	8	9	10

System Use

	Strongly Disagree									Strongly Agree
It is safe for me to use this system while driving.	1	2	3	4	5	6	7	8	9	10
I would Iikely use this system for my daily travel.	1	2	3	4	5	6	7	8	9	10
I would use this system if I was in a hurry.	1	2	3	4	5	6	7	8	9	**10**
It is safe for an inexperienced driver to use this system while driving.	1.2		3	4	5	6	7	8	9	10
The traffic information provided is useful.	1	2	3	4	5	6	7	8	9	10
The traffic messages were not distracting.	1	2	3	4	5	6	7	8	9	10
I would rather use this system than a traffic news report.	1	2	3	4	5	6	7	8	9	10
The traffic information would be useful when driving in an unfamiliar area.	1	2	3	4	5	6	7	8	9	10
The traffic information would be helpful when driving in a familiar area.	1	2	3	4	5	6	7	8	**9**	**10**

Familiarity

	Unfamiliar									Very Familiar
How familiar are you with the Detroit area highway system (in general)?	1	2	3	4	5	6	7	8	**9**	**10**
Previous to today, how familiar were you with the DIRECT traffic information project?	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	9	10

How often do you take I-94 from I-275 into Detroit?

Daily A few times a week Once a week A few times a month Once a month Less

Do you commute to Detroit?

Daily A few times a week Once a week A few times a month Once a month Less

How often do you listen to radio broadcasts concerning local or area traffic'

Daily A few times a week Once a week A few times a month Once a month Less

Message Quality

Did the audio quality affect your ability to understand the information provided by the traffic messages? Yes Somewhat No

Was there anything about the message/system that led to confusion'

_____

Did you have trouble distinguishing between:

I-75 and I-275	Yes	Somewhat	No
I-96 and I-94	Yes	Somewhat	No

Task Difficulty

Using all of your driving experience (not just what you used today), please rate the difficulty o performing each of these tasks while driving, using the scale below.

	Not difficult									Very difficult
Reading the speedometer	1	2	3	4	5	6	7	8	9	10
Drinking a beverage	1	2	3	4	5	6	7	8	9	10
Reading street names	1	2	3	4	5	6	7	8	9	10
Adjusting the fan speed on the car heater or air conditioner	1	2	3	4	5	6	7	8	9	10

Other

Not including today, have you ever driven a vehicle with a traffic information system installed (not radio broadcasts): Yes No

When do you plan on buying your next new or used vehicle?

less than a month within 6 months within 2 years more than 2 years

How much do you plan on spending? _____

How much would you pay to purchase (no monthly fee) for a traffic information system like tht one you used today? $_____

Additional Comments:

_____

_____

_____

_____

APPENDIX F - RECALL AS A FUNCTION OF MESSAGE LENGTH

Figures 24 through 30 show the percent correct by term for each message length. The first, second, and third term type (road, direction of travel affected, crossroad) were the same for each message length. The road and crossroad terms consistently had a higher percent of correct responses than the second term, and the "mile back-up" element (last term) had recall superior to the back-up distance.

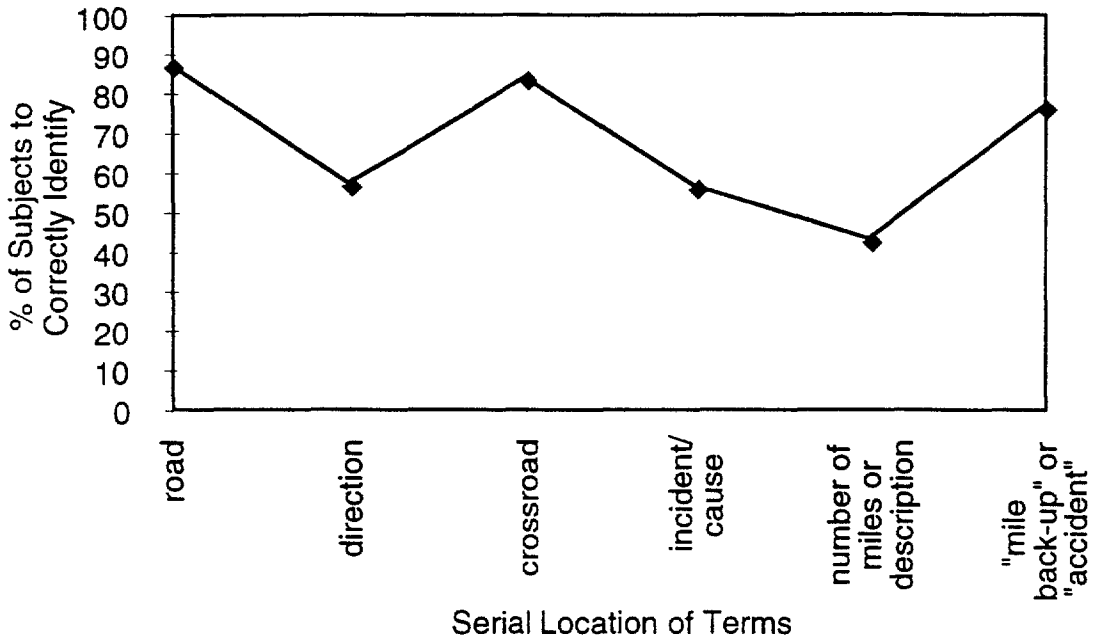


Figure 24. Percent correct responses for 6-term messages.

For 7-term messages (Figure 25), the term "construction" was recalled by a greater percentage of subjects than the construction description, "continuing". It is assumed that subjects remember the type of incident/cause of traffic, but not any additional descriptive information concerning the incident.

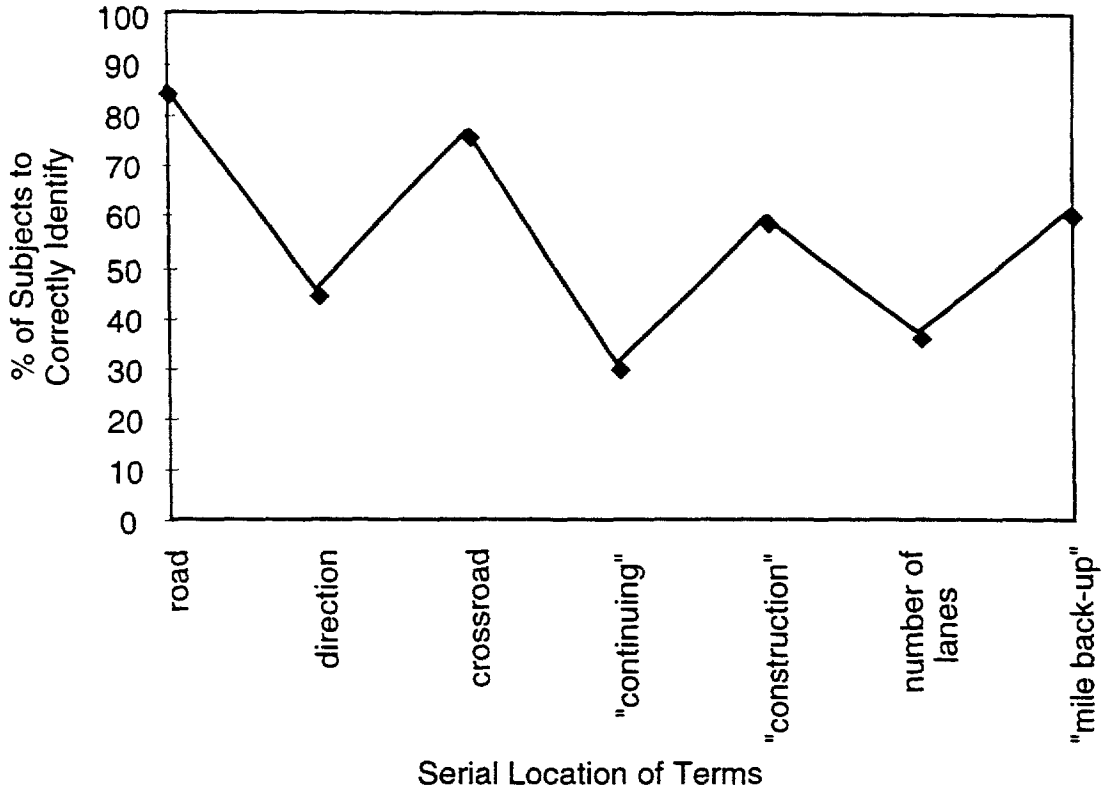


Figure 25. Percentage of correct responses for 7-term messages.

For 8-term messages (Figure 26) recall was poor overall except for the road in question.

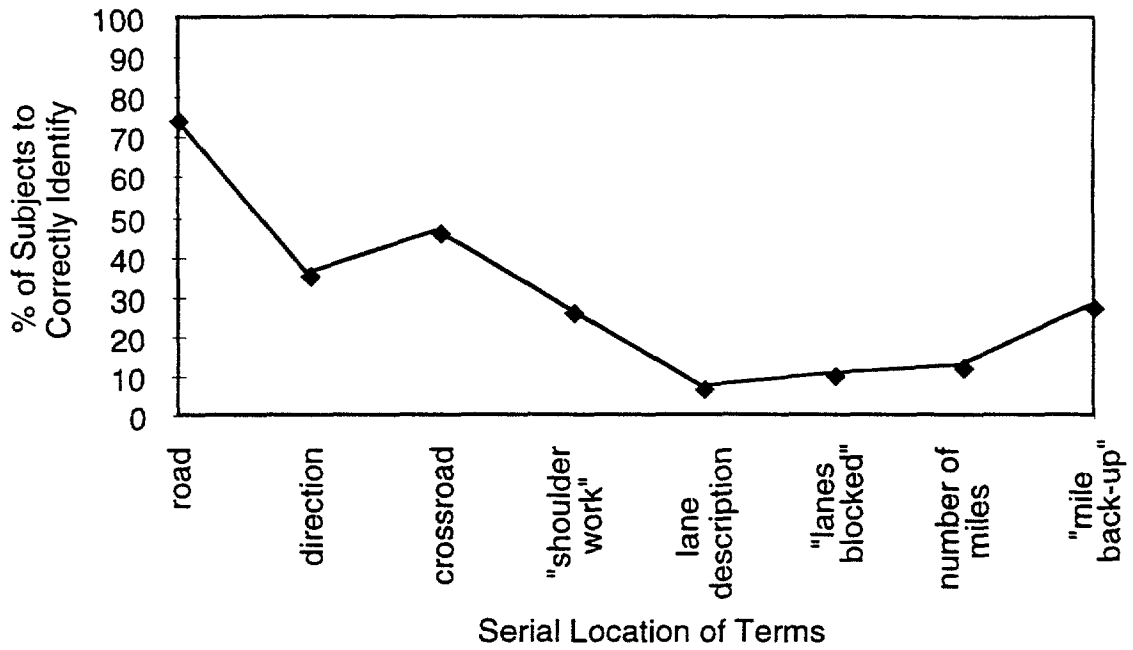


Figure 26. Percentage of correct responses for 8-term messages.

For messages that contain 9 terms (Figure 27), recall of the incident / cause of traffic was relatively good. The cause, "hazardous material spill" was recalled much more often than any of the other information describing the situation.

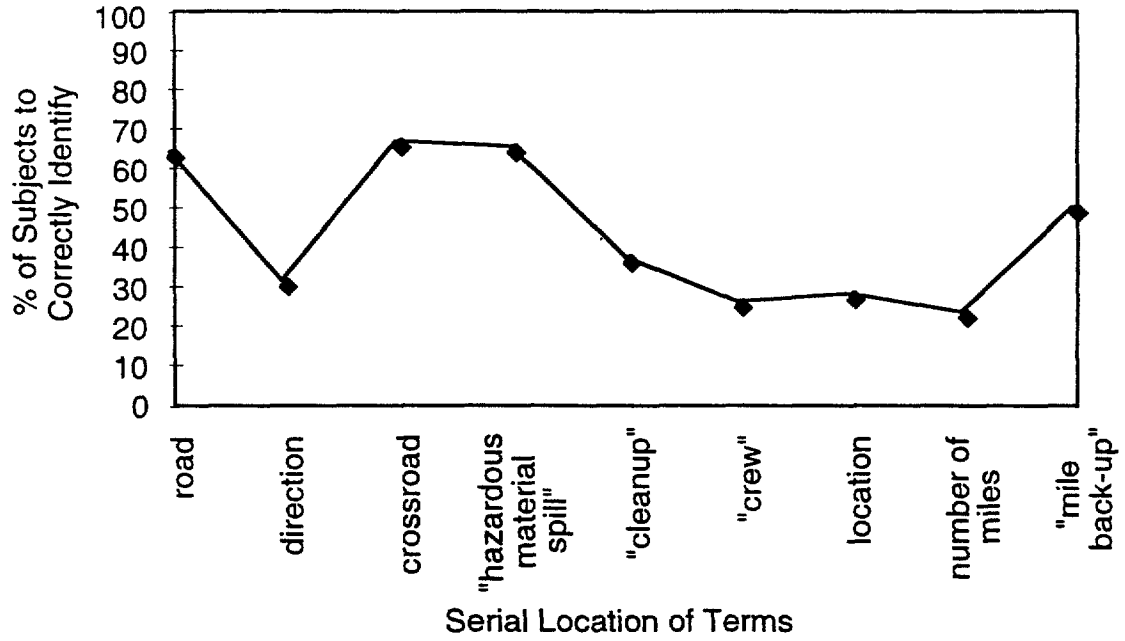


Figure 27. Percentage of correct responses for 9-term messages.

Eleven-term messages concerned either a "railroad derailing" or a "hazardous material spill" incident / cause of traffic. Figure 28 shows that the incident term was recalled much more often than other information provided in the messages.

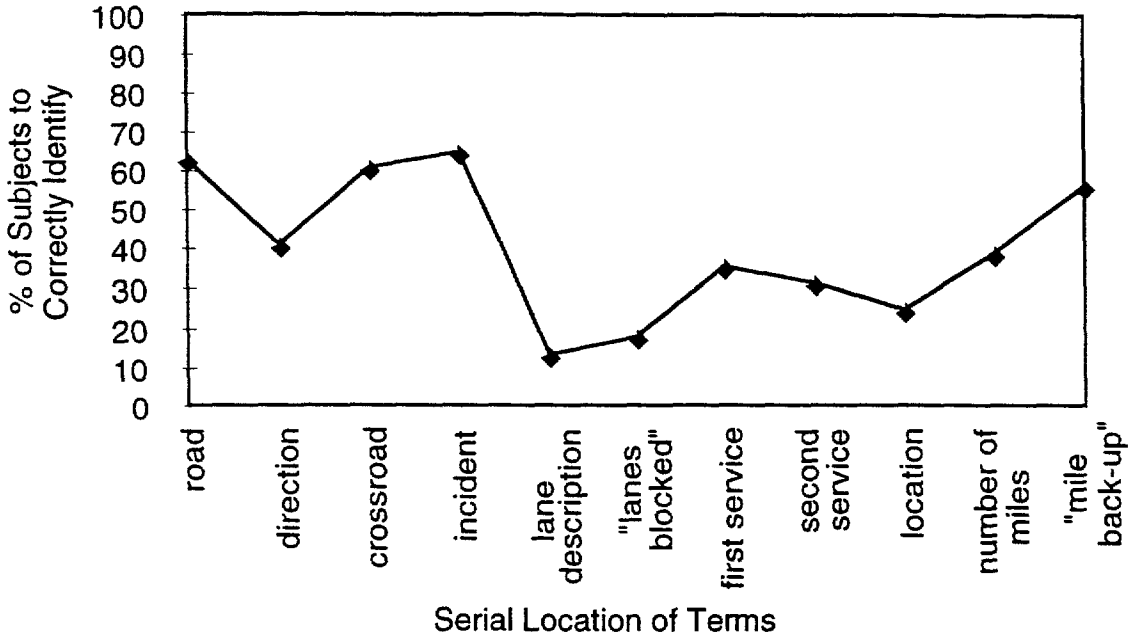


Figure 28. Percentage of correct responses for 11-term messages.

Figure 29 shows that the term "accident" was recalled by more subjects than any of the other terms describing the accident or the handling of the accident. A similar pattern is present in Figure 30 for 14-term messages.

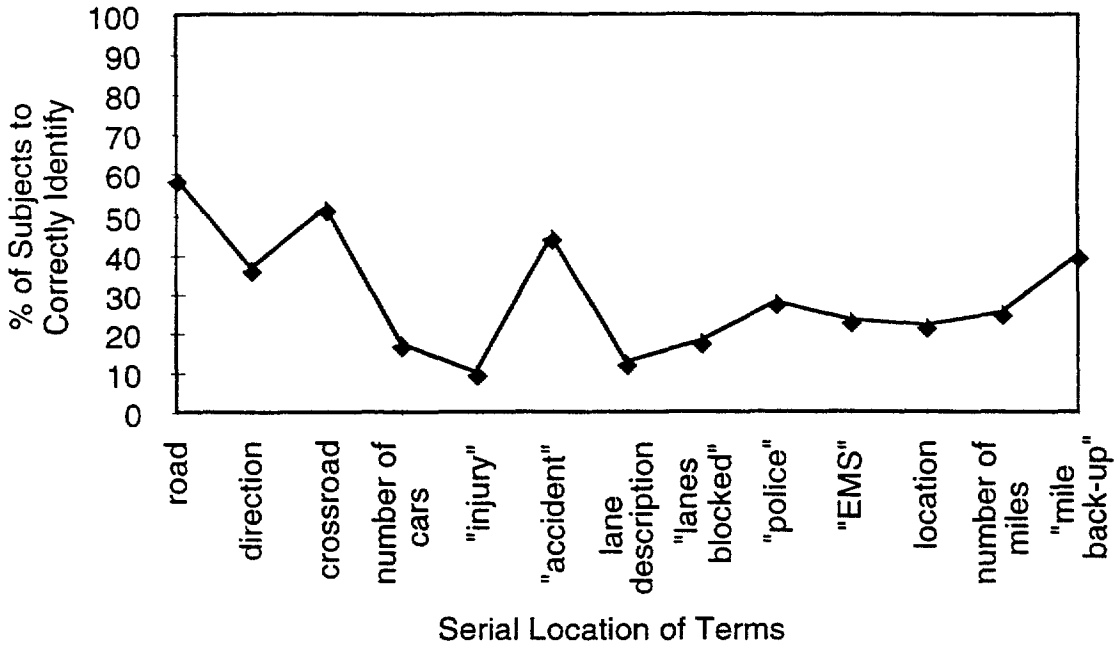


Figure 29. Percentage of correct responses for 13-term messages.

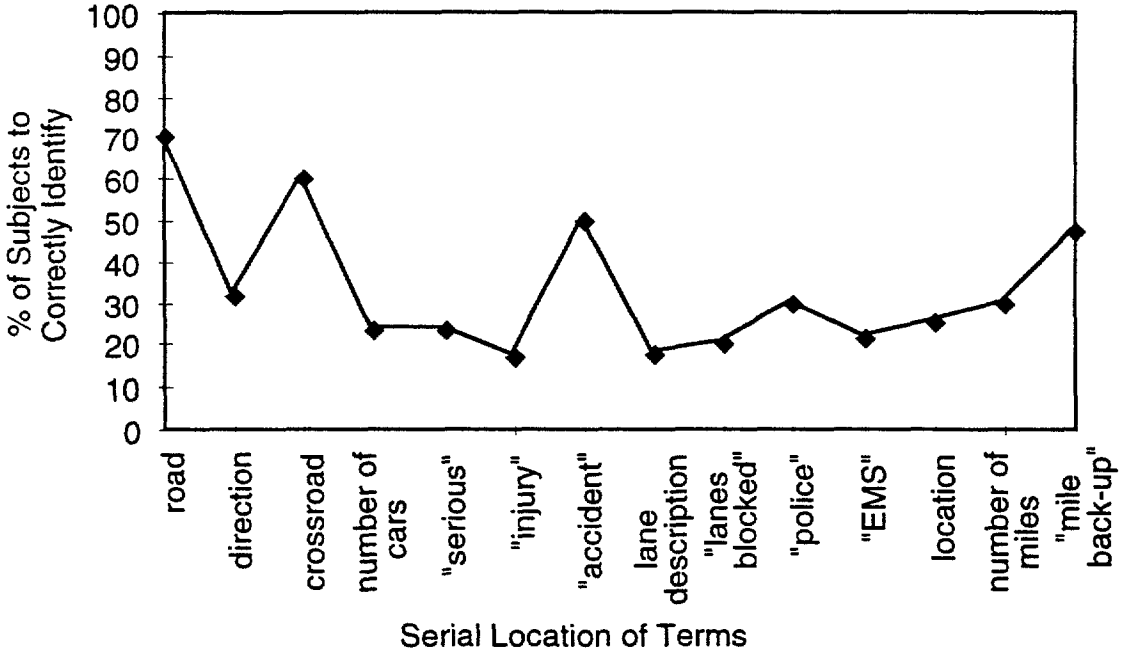


Figure 30. Percentage of correct responses for 14-term messages.

APPENDIX G - ANOVA OF DRIVING DATA

Independent Variables	Dependent Driving Performance Measures						
	Mean speed	Standard deviation of speed	Standard deviation of throttle	Mean range	Standard deviation of range	Standard deviation of lane position	std deviation of steering wheel angle
Age Group	0.0002	0.8784	0.7705	0 0748	0.7410	0.9176	0.0133
Gender	0.0283	0 4727	0.564 1	0.623	0.4072	0.8343	0.9384
Time Of Day	0.4278	0.6467	0.5073	0.4179	0.5134	0.8821	0.489
Relevance	0.6763	0.9863	0.0268	0.4332	0.0177	0.0006	0.0001
Message Quality	0.0919	0.0001	0.0042	0.0036	0.7769	0.0001	0.0001
Number Of Messages	0.2308	0.1464	0.0001	0.005	0.0001	0.0018	0.0001
Repetitions	0.0001	0.0805	0.0012	0.1985	0.0001	0.0001	0.0001
Segment (Within Trial)	0.5363	0.0001	0.0526	0.3182	0.0539	0.2128	0.0025
Age Group *Gender	0.432	0.1006	0.5367	0.5717	0.5614	0.867	0.4867
Age Group . Time Of Day	0.7693	0.7096	0.8447	0.9859	0.8290	0.3484	0.4046
Gender 'Time Of Day	0.9736	0.8531	0.7996	0.2716	0.7447	0.2000	0.9395
Relevance . Age Group	0.2628	0.9830	0.2334	0.9845	0.4031	0.3189	0.9322
Relevance . Gender	0.1956	0.4284	0.7198	0.7567	0.6050	0.8002	0 617
Relevance * Time Of Day	0.5714	0.4502	0.9665	0.7308	0.2360	0.0400	0.7696
Message Quality . Age Group	0.9746	0.1797	0.4507	0.2828	0.9087	0.6843	0 6872
Message Quality * Gender	0.6778	0.1518	0.9668	0.5889	0.3369	0.9132	0.0677
Message Quality . Time Of Day	0.761	0.2813	0.0689	0.8923	0.7363	0.3863	0.944 1
Repetitions* Age Group	0,.9189	0.9646	0.9694	0.8321	0 9898	0.0617	0.3126
Repetitions . Gender	0.1925	0.8181	0.9654	0.9287	0.9679	0.3019	0.9901
Repetitions ' Time Of Day .	0.0183	0.2426	0.4905	0.3341	0.4633	0.4913	0.3247
Number Of Messages . Age Group	0.1944	0.7508	0.6305	0.5305	0.8002	0.783	0.0362
Number Of Messages . Gender	0.2469	0.4523	0.4370	0.4774	0.8107	0.8668	0.3263
Number Of Messages . Time Of Day	0.492	0.2776	0.9000	0.4604	0.3277	0.2071	0.4294

Independent Variables	Mean speed	Standard deviation of speed	Standard deviation of throttle	Mean range	Standard deviation of range	Standard deviation of lane position	std deviation of steering wheel angle
Segment * Age Group	**0.0034**	**0.0328**	0.9901	0.9977	0.4662	0.8516	0.9308
Segment * Gender	0.7558	0.6027	0.7788	0.1553	0.1475	0.8921	0.7180
Segment * Time Of Day	0.1059	0.2133	0.3044	0.5695	0.8115	0.6018	0.0866
Relevance* Message Quality	0.1816	**0.0054**	**0.0001**	**0.0001**	**0.0001**	**0.0001**	**0.0001**
Relevance * Repetitions	**0.0001**	**0.0001**	**0.0001**	0.0578	**0.0001**	**0.0001**	**0.0001**
Message Quality * Repetitions	**0.0001**	**0.0001**	**0.0001**	**0.0007**	**0.0001**	**0.0001**	**0.0001**
Relevance * Number Of Messages	**0.0023**	**0.032**	**0.0001**	**0.0284**	**0.0001**	**0.0001**	**0.0001**
Message Quality * # Of Messages	**0.001**	**0.0024**	**0.0001**	**0.0119**	**0.0001**	**0.0001**	**0.0001**
Number Of Messages * Repetitions	**0.0001**	**0.0001**	**0.0001**	**0.0001**	**0.0001**	**0.0001**	**0.0001**
Relevance * Segment	0.7991	0.3438	0.0815	0.8117	0.8779	0.5945	**0.0146**
Message Quality * Segment	0.1878	0.2797	0.8018	0.3653	0.7341	0.2067	0.2837
Number Of Messages * Segment	0 1710	0.6592	**0.0003**	0.7358	**0.0008**	0.1405	**0.0006**
Repetitions * Segment	0.8251	0.1650	**0.0213**	0.4863	0.5765	0.2456	0.0859

APPENDIX H - DETAILED DISCUSSION OF THE DRIVING DATA

Mean Speed

Mean speed is the measure of a driver's speed while driving the test route, and is reported in mi/hr. Young drivers tended to drive faster than older drivers (mean = 68.5 versus 64.7 mi/hr, p = 0.0002) and men drove faster than women (67.6 versus 65.5 mi/hr, p = 0.0283). Except for the standard deviation of steering wheel angle, mean speed was the only driving measure for which there were significant subject effects, an unusual situation.

Each relevance * message quality * number of messages combination of the experiment was repeated to each subject 8 times. The mean speed increased significantly with repetitions (p = 0.0001) probably indicating greater comfort with the task with practice. In addition, there was an interaction of repetitions with time of day (p = 0.0183), with less of an effect of practice in the afternoon, most likely due to a difference in traffic levels (Figure 31).

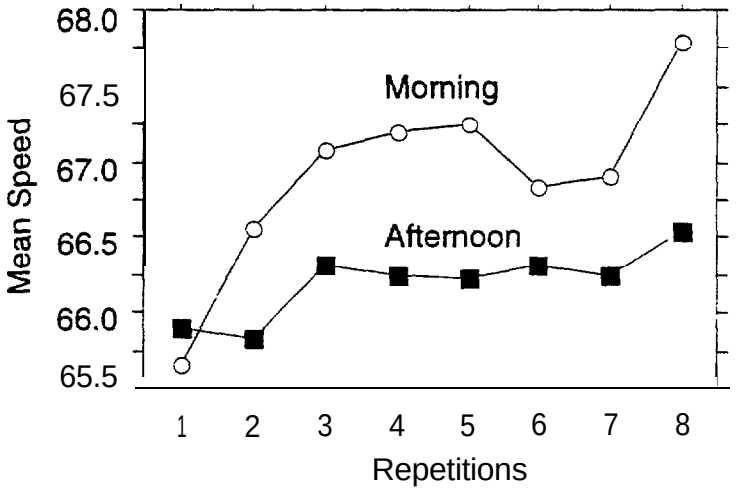


Figure 31, Mean speed for the repetition by time of day interaction.

Within each trial, there was a significant but very slight interaction with speed, with younger subjects driving faster while responding and older subjects driving slower (p= 0.0034). (See Figure 32). One explanation is that the act of speaking with the experimenter added some mental load for older drivers, and they compensated by slowing down to reduce driving workload.

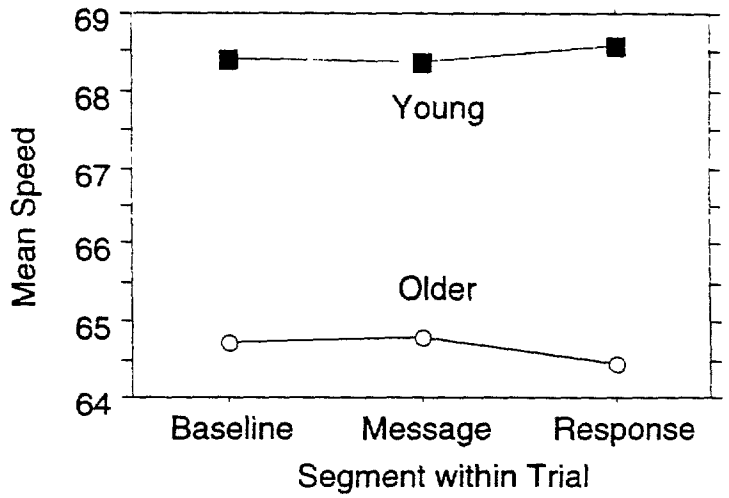


Figure 32. Mean speed for the segment within trial by age group interaction.

The significant (p = 0.0001) interaction between message audio quality and number of messages and the lack of a significant difference due to audio quality does not make sense (Figure 33). If anything, one would expect people to slow down when listening to poor quality messages to be able to pay more attention to them and reduce road noise.

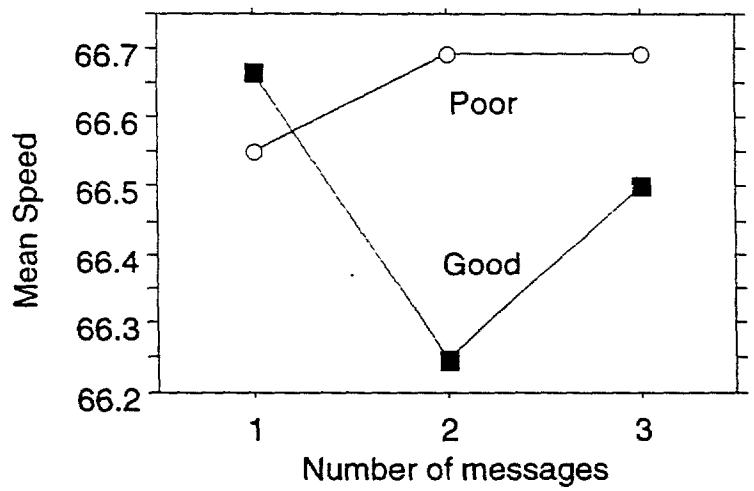


Figure 33. Mean speed for the message quality by number of messages interaction.

Standard Deviation of Speed

The standard deviation of speed was significantly affected by the audio quality of traffic messages (p = 0.0001, 3.4 mi/hr for good quality messages, 3.6 mi/hr for poor quality messages). Most likely this was because the mean speed was slightly less for good messages, and speed and speed variability are correlated.

The ANOVA also reveal a significant difference (p = 0.0001) between segments (baseline = 3.35 mi/hr, message = 3.44 mi/hr, response = 3.68 mi/hr) in the standard

deviation of speed. The baseline, message, and response segments of the trial have mean standard deviations of 3.35 mi/hr, 3.44 mi/hr, and 3.68 mi/hr, respectively (Figure 34). Thus, just listening to traffic messages led to a tiny increase in speed variability, and talking with the experimenter increased variability further. The interaction between segment and age group was also significant (p = 0.0328).

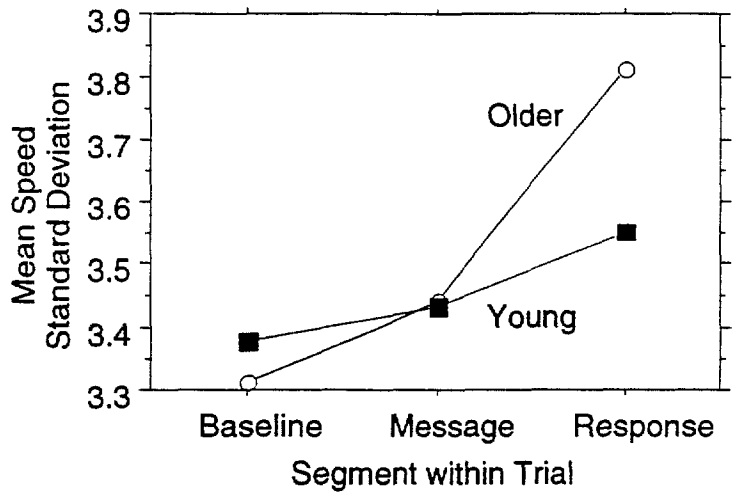


Figure 34. Mean speed standard deviation of segment by age group interaction.

The interaction between relevance and message quality (Figure 35) was significant (p = 0.0054). In brief, audio quality only mattered (influenced speed variability) for relevant messages. Improving audio quality led people to drive at a steadier pace.

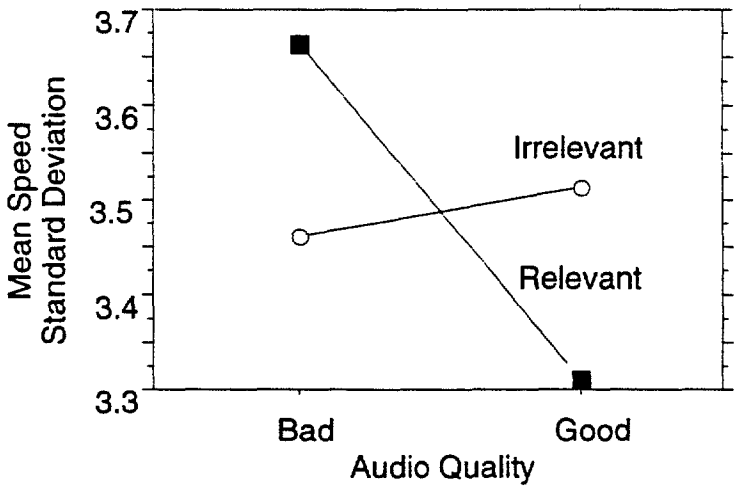


Figure 35. Mean speed standard deviation for message quality by relevance interaction.

The significant interaction (p = 0.0024) between audio quality and number of messages is most likely chance (Figure 36). There is no reason why people should wander less in the lane when 3 messages are presented rather than when a smaller number were presented.

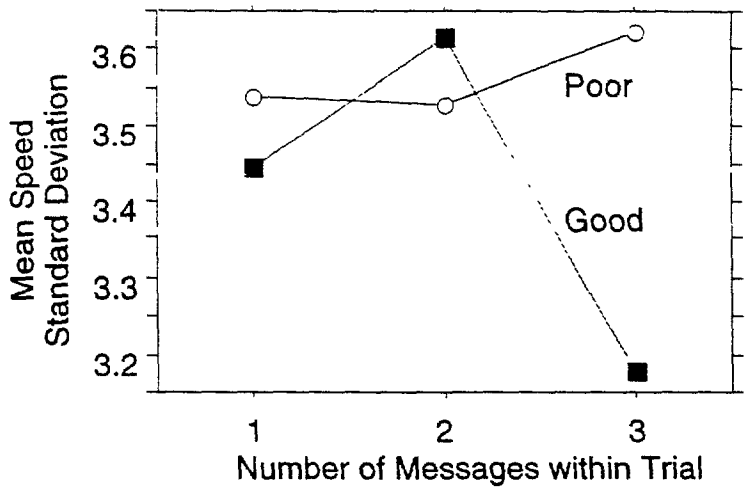


Figure 36. Mean speed standard deviation for audio quality by number of messages interaction.

Standard Deviation of Throttle

The standard deviation of throttle position is the measure of the variance of throttle position, how much the driver changes the position of the gas pedal, here measured in percentages. As the input to speed, throttle variance is sometimes more sensitive than speed variance in some situations that affect speed. Here, that was the case for message relevance ($p = 0.0268$), with irrelevant messages leading to a 0.1 percent increase in throttle variability, a small amount (Figure 37). Similarly, throttle position was slightly more variable with poor messages ($p = 0.0042$) by an identical amount. The two interacted as well ($p = 0.0001$). The effect of the number of messages was significant ($p = 0.0001$) but for an unknown reason, it was greatest for 2 messages (Figure 38). There was a similar interaction ($p = 0.0001$) with message quality (Figure 39).

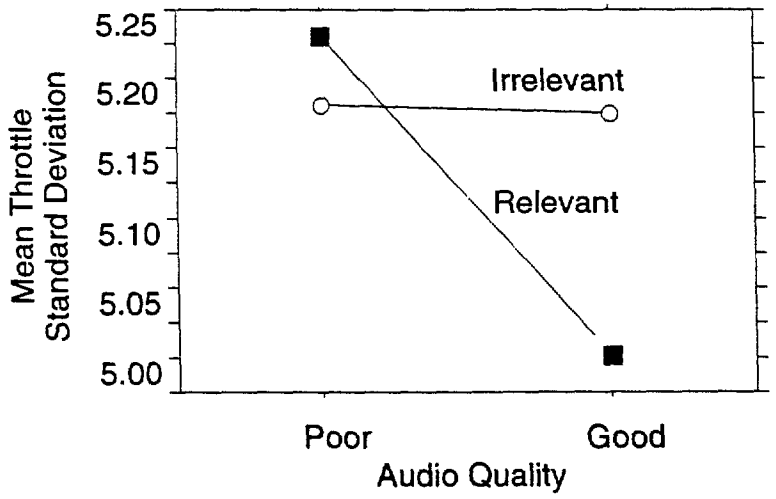


Figure 37. Mean throttle position for audio quality by relevance interaction.

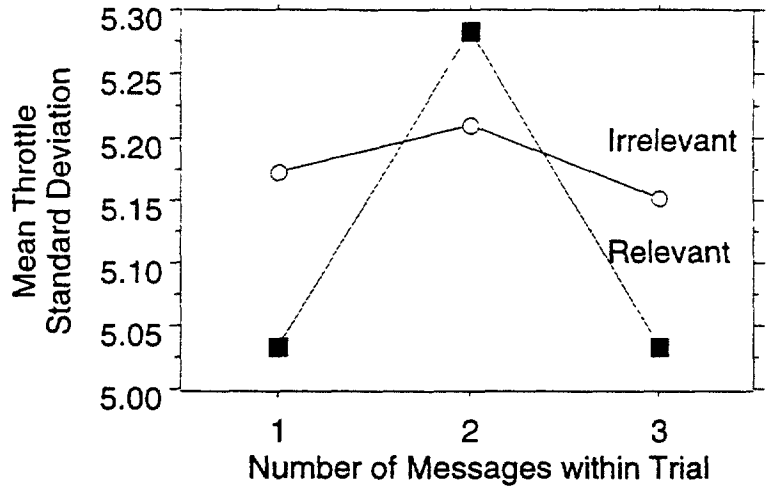


Figure 38. Mean throttle standard deviation for number of messages by relevance interaction.

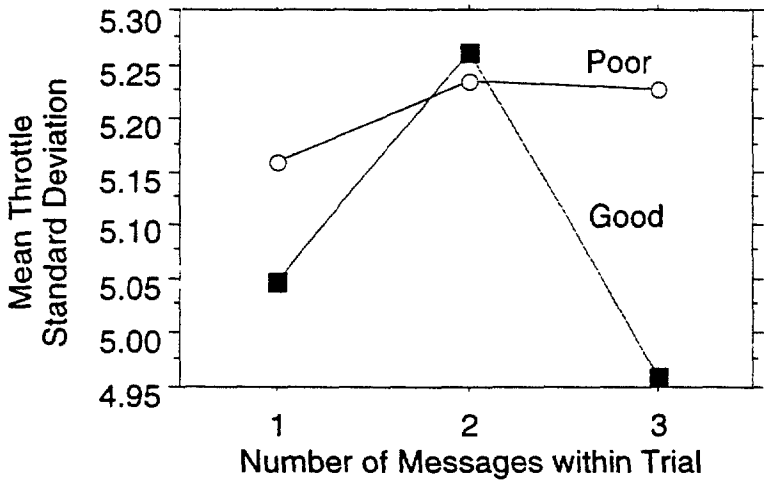
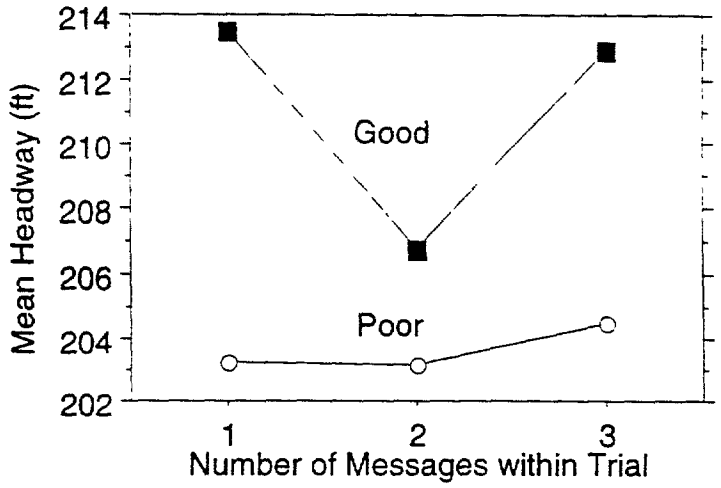


Figure 39. Mean throttle standard deviation for number of messages by audio quality interaction.

Mean Headway

Mean headway (or range) is the average distance between the test vehicle and the first vehicle ahead (measured in feet). Mean headway was significantly affected by message quality ($p = 0.0036$), with greater headways for good than poor (211 versus 204 ft) (Figure 40). That is, people drove slightly closer to cars ahead of them when presented with poor messages, possibly indicating that attending to messages distracted them from following as they should. There was also a slight effect of the number of messages ($p = 0.005$) and an interaction with audio quality ($p = 0.0119$).



For message relevance, only its interaction with audio quality (Figure 41) was significant (p=0.0001).

Figure 40. Mean headway for number of messages by mean headway interaction.

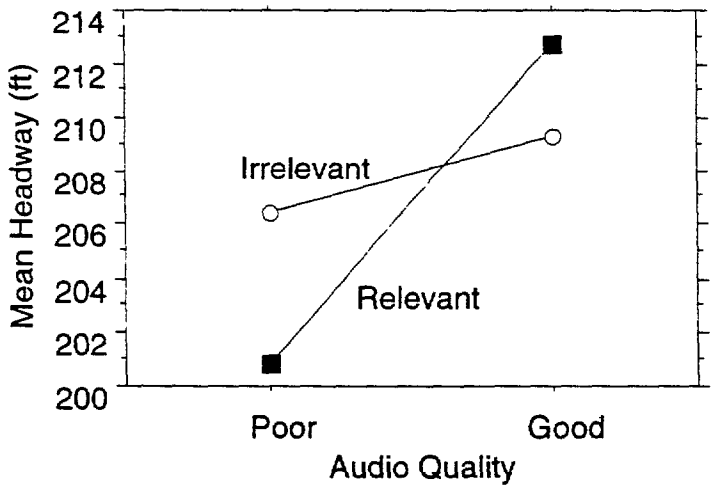


Figure 41. Mean headway for audio quality by message relevance interaction.

Standard Deviation of Headway

The standard deviation of headway is the variability in distance between the test vehicle and a vehicle in front of it. Significant differences were found due to message relevance (p = 0.0177, 1 foot less for relevant messages) and the number of messages per trial (p = 0.0001, 4 foot range), and their interaction (p = 0.0001), all slight differences. (See Figure 42.)

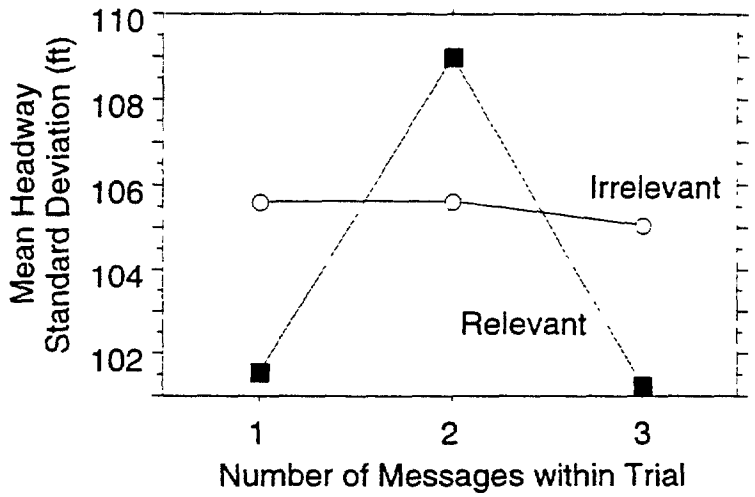


Figure 42. Mean headway standard deviation for number of messages by relevance interaction.

Additionally, there were significant interactions (p = 0.0001) between audio quality and the number of messages (Figure 43), and the number of messages and the segment within the trial (p = 0.0008, Figure 44).

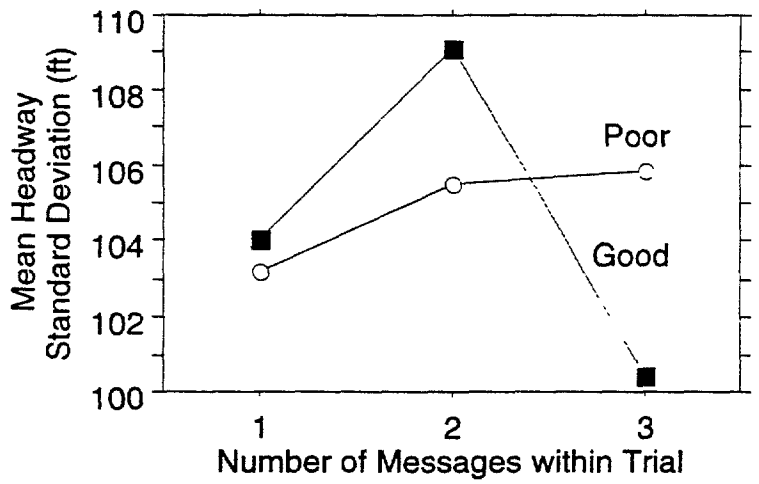


Figure 43. Mean headway standard deviation for number of messages by audio quality interaction.

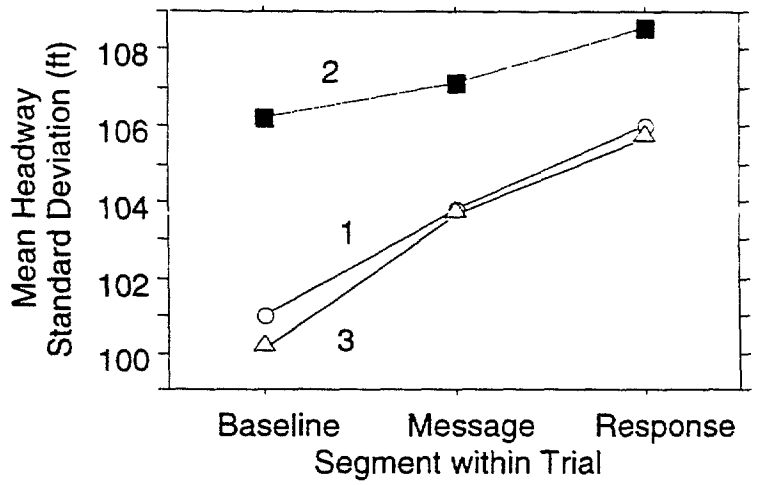


Figure 44. Mean headway standard deviation for number of messages by segment within trial interaction.

Standard Deviation of Lateral Position

Standard deviation of lateral position (measured in feet) is the variance of the test vehicle's position within the lane of the road. Greater lane variability reflects poorer driving and may lead to crashes. The first significant variable with respect to standard deviation of lateral position was message relevance ($p = 0.0006$). The mean standard deviation of lateral position for irrelevant trials was 1.93 ft, and the mean standard deviation for relevant trials was 1.9 ft. There was slightly greater lane position variability during irrelevant trials.

Message audio quality also significantly affected the standard deviation of lateral position ($p = 0.0001$). Poor and good audio quality trials had mean standard deviations of 2.0 ft, and 1.9 ft, respectively. Greater lane position variability was seen during poor audio quality messages.

The number of messages within a trial has a p-value of 0.0018, hence significantly affecting standard deviation of lateral position. Trials with 1, 2, or 3 traffic messages per trial had mean lane-position standard deviations of 1.92 ft, 1.94 ft, and 1.9 ft, respectively.

The interaction between relevance and message quality was found to significantly affect standard deviation of lane position ($p = 0.0001$). Figure 45 shows that irrelevant and relevant trials had similar lane variability during bad quality trials. However, during good quality trials, irrelevant trials had greater lane position variability.

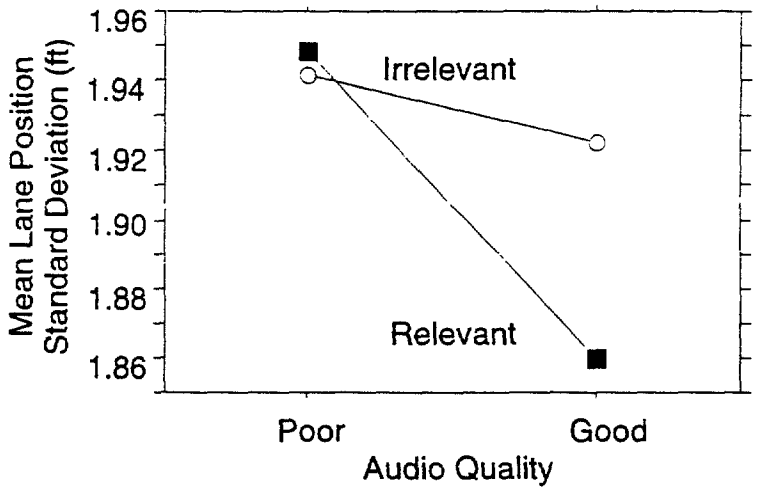


Figure 45. Mean lane position standard deviation for message quality by relevance interaction.

Additionally, the interaction between message quality and the number of messages within the trial has a p-value of 0.0001. Poor quality trials had a greater lane position variability for trials with 2 and 3 traffic messages. (See Figure 46.) Both poor and good quality audio trials had similar lane variability when one traffic message was present.

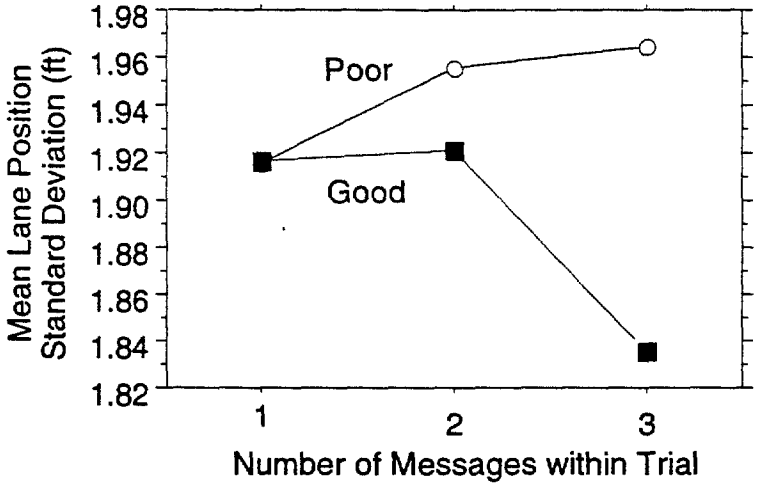


Figure 46. Mean lane position standard deviation for number of messages by audio quality interaction.

Standard Deviation of Steering Wheel Angle

The standard deviation of the steering wheel angle (measured in degrees) is the variance in the angular position of the steering wheel. Due to imperfect mounting of the sensor, when the steering wheel was centered, a measurement of -18.8 degrees was recorded.

Age did lead to significant differences in the standard deviation of steering wheel angle (p = 0.0133), with the variability being greater for young drivers (mean = 2.7 versus 2.4 degrees). This could indicate that younger drivers make larger corrections or more lane changes than older drivers.

Message relevance was also significant with respect to the standard deviation of steering wheel angle (p = 0.0001). Variability was 0.1 degrees greater for relevant trials (2.6 degrees versus 2.5 degrees), possibly reflecting the slightly additional demand of attending to relevant messages and delays in making corrections. Similarly, good quality messages led to reduced steering wheel angle variability (p = 0.0001, 2.5 versus 2.6 degrees), again reflecting diminished attentional demand. The two also interacted (p = 0.0001, Figure 47).

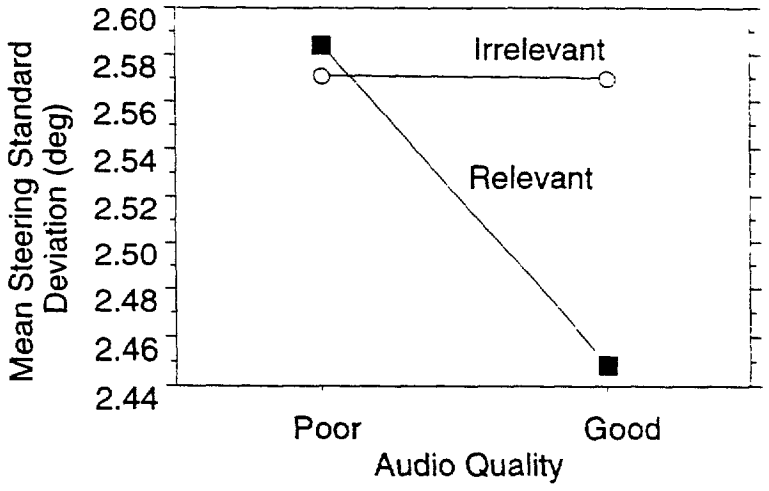


Figure 47. Mean steering wheel standard deviation for audio quality by message relevance interaction.

Consistent with measures, there were significant differences due to the number of messages present in a trial (p = 0.0001), with a maximum for 2 messages.

The segment within each trial also significantly affected the standard deviation of the steering wheel angle (p = 0.0025), during the response segment of the trial when the subject was interacting with the experimenter (Figure 48). The segment effect also interacted with message relevance (p = 0.0146) and with the number of trials (p = 0.0006, Figure 49).

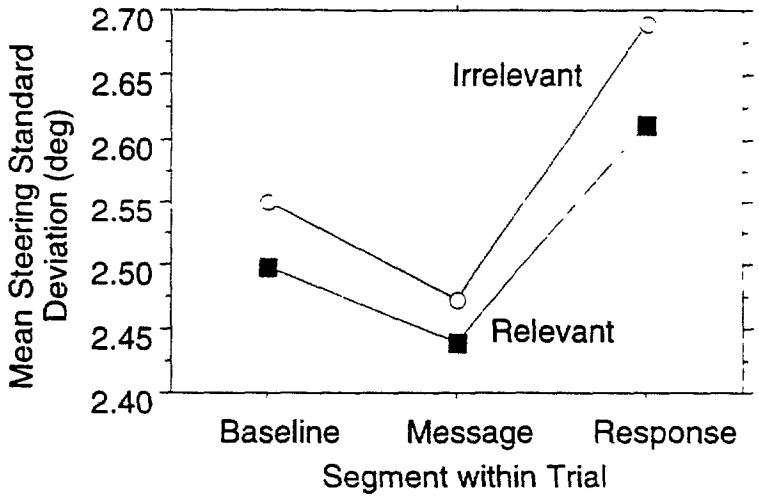


Figure 48. Mean steering wheel angle standard deviation for relevance by segment interaction.

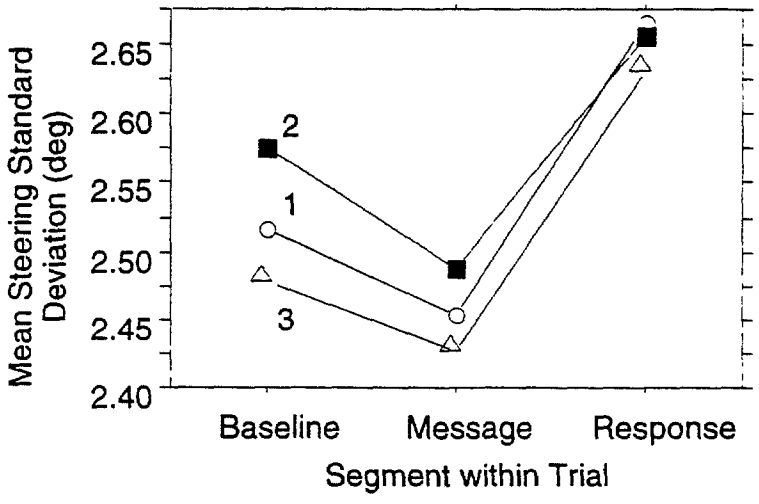


Figure 49. Mean steering wheel angle standard deviation for number of messages by segment interaction.