U.S. Department
of Transportation

**Federal Highway
Administration**

Publication No. FHWA-RD-94-088
August 1995

Measures and Methods Used to Assess the Safety and Usability of Driver Information Systems

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

FOREWORD

This report concerns in-car systems that may be used to present navigation, hazard warning, vehicle monitoring, traffic, and other information to drivers in cars of the future. It describes in detail measurements researchers have made to determine if those systems are safe and easy to use. This report also touches upon issues relating to comfort, convenience, and confidence.

This report will be useful to the designers of driver information systems and scientists conducting driving-related human factors research in academia, industry, or government agencies.

Lyle Saxton, Director
Office of Safety and Traffic
Operations Research and Development

NOTICE

Technical Report Documentation Page

1. Report No. FHWA-RD-94-088	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle
MEASURES AND METHODS USED TO ASSESS THE SAFETY AND USABILITY OF DRIVER INFORMATION SYSTEMS

5. Report Date
August 1995

6. Performing Organization Code
Account 080066

7. Author(s)
Paul Green

8. Performing Organization Report No.
UMTRI-93-12

9. Performing Organization Name and Address
The University of Michigan
Transportation Research Institute
2901 Baxter Road, Ann Arbor, Michigan 48109-2150

10. Work Unit No. (TRAIS)
3B3B-2032

11. Contract or Grant No.
DTFH61-89-C-00044

12. Sponsoring Agency Name and Address
Office of Safety and Traffic Operations R&D
Federal Highway Administration
6300 Georgetown Pike, McLean, Virginia 22101-2296

13. Type of Report and Period Covered
Final Report
September 1991 - May 1993

14. Sponsoring Agency Code

15. Supplementary Notes
This research was jointly funded by the Federal Highway Administration and the National Highway Traffic Safety Administration. Contracting Officer's Technical Representative (COTR): Nazemeh Sobhi, HSR-30.

16. Abstract

This report concerns in-car systems that may be used to present navigation, hazard warning, vehicle monitoring, traffic, and other information to drivers in cars of the future. It describes in detail measurements researchers have made to determine if those systems are safe and easy to use.

Measures that appear most promising for safety and usability tests of driver information systems include the standard deviation of lane position, speed, speed variance, and the mean and frequency of driver eye fixations to displays and mirrors. In some cases, laboratory measures (errors, etc.) may also be useful. Also of interest are time-to-collision and time-to-line crossing, although hardware for readily measuring them in real time is not available. Of lesser utility are workload estimates (SWAT, TLX). Secondary task measures and physiological measures are very weak predictors of safety and usability.

To assess usability, application-specific measures (e.g., the number of wrong turns made in using a navigation system) should be collected.

17. Key Words IVHS, human factors engineering, ergonomics, navigation, driver information systems, test methods	18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
unclassified	unclassified	115	

Form DOT F 1700.7 (8-72)
(PF V2.1, 12/13/93)

Reproduction of completed page authorized

PREFACE

PROJECT OVERVIEW

The United States Department of Transportation (DOT), through its Intelligent Vehicle-Highway Systems (IVHS) program, is aiming to develop solutions to the most pressing problems of highway travel. The goal is to reduce congestion and improve traffic operations, reduce accidents, and reduce air pollution from vehicles by applying computer and communications technology to highway transportation. If these systems are to succeed in solving the nation's transportation problems, they must be safe and easy to use, with features that enhance the experience of driving. The University of Michigan Transportation Research Institute (UMTRI), under contract to DOT, carried out a project to help develop IVHS-related driver information systems for cars of the future. This project concerns the driver interface, the controls and displays that the driver interacts with, as well as their presentation logic and sequencing.

The project had three objectives:

- Provide human factors guidelines for the design of in-vehicle information systems.
- Provide methods for testing the safety and ease of use of those systems.
- Develop a model that predicts driver performance in using those systems.

Although only passenger cars were considered in the study, the results apply to light trucks, minivans, and vans as well because the driver population and likely use are similar to cars. Another significant constraint was that only able-bodied drivers were considered. Disabled and impaired drivers are likely to be the focus of future DOT research.

A complete list of the driver interface project reports and other publications is included in the final overview report, 1 of 16 reports that document the project.[1] (See also Green, Serafin, Williams, and Paelke, 1991 for an overview.)[2] To put this report in context, the project began with a literature review and focus groups examining driver reactions to advanced instrumentation.[3,4,5] Subsequently, the extent to which various driver information systems might reduce accidents, improve traffic operations, and satisfy driver needs and wants, was analyzed.[6,7] That analysis led to the selection of two systems for detailed examination (traffic information and car phones) and contractual requirements stipulated three others (route guidance, road hazard warning, and vehicle monitoring) likely to appear in future vehicles.

Each of the five systems selected was examined separately in a sequence of experiments. In a typical sequence, patrons at a local driver licensing office were shown mockups of interfaces, and driver understanding of the interfaces and preferences for them was investigated. Interface alternatives were then compared in laboratory experiments involving response time, performance on driving simulators, and part-task simulations. The results for each system are described in a separate report. (See references 8, through 14.) To check the validity of those results, several on-road experiments were conducted in which performance and preference data for the various interface designs were obtained.[15,16]

Concurrently, UMTRI developed test methods and evaluation protocols, UMTRI and Bolt Beranek and Newman (BBN) developed design guidelines, and BBN worked on the development of a model to predict driver performance while using in-vehicle information systems. (See references 17 through 20.)

Many of the reports from this project were originally dated May, 1993, the contractual end date of the project whereby reports were to be delivered. However, the reports were actually drafted when the research was conducted, more than two years earlier for the literature review and feature evaluation, and a year earlier for the laboratory research and methodological evaluations. While some effort was made to reflect knowledge gained as part of this project, the contract plan did not call for rewriting reports to reflect recent findings.

THIS REPORT

This report is one of two concerning the testing of the safety and ease of use of driver interfaces. It also touches upon issues relating to comfort, convenience, and confidence, but there is very little information in the literature on those issues as they relate to driver interfaces.

The bulk of this report is devoted to a review of the methods and measures for assessing safety and ease of use of IVHS-related driver information systems. Because of dissemination constraints, it is quite likely that coverage of the DRIVE and PROMETHEUS programs is incomplete. Very little is known in the U.S. about work in Japan. Participants in programs in Europe and Japan are encouraged to send the author copies of reports and papers that are pertinent to this review.

This report is written for scientists conducting automotive human factors research, though some practitioners interested in evaluation may find this report to be of interest. Those scientists may be working in academia, industry, or for government agencies. Within the Department of Transportation, both the Federal Highway Administration (FHWA) and the National Highway Traffic Safety Administration (NHTSA) are interested in this research, with NHTSA having expressed the greatest interest. Accordingly, this report emphasizes safety issues.

A secondary audience is human factors scientists with expertise in defense applications, but little knowledge of automotive applications. Their interest is the result of federal policy decisions to foster defense conversion to civilian applications. To provide context for defense scientists, reviews of key studies have been included in this report, as well as a detailed tabular summary of all the studies on navigation systems and related topics.

Serving as a companion to this report is a subsequent report that describes suggested assessment protocols (Green, 1993).[18] That subsequent report is written primarily for practitioners interested in conducting assessments of driver interfaces, many of whom will be automotive human factors engineers. It is likely that human factors scientists will have a keen interest in that report as well. These two reports were produced as separate documents because the audiences were different and to expedite release of the material.

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

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

APPROXIMATE CONVERSIONS FROM SI UNITS

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

---

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

(Revised September 1993)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

BACKGROUND

The car of the future could be quite different from that on the road today. While it will still have a steering wheel, brake, and accelerator, and controls for secondary functions (radio, climate, etc.), a host of new systems will either be introduced or see expanded use. These systems include navigation, traffic information, collision avoidance, etc.

In the recent past, electronic technologies have been added to cars in the unrealized hope that such technologies would see widespread use [electronic displays in general, voice output, HUDs (Head-Up Displays, etc.)]. In implementing these technologies, significant human factors problems have arisen, in addition to problems with cost and reliability. If Intelligent Vehicle-Highway Systems (IVHS) are to succeed, those systems must be safe and easy to use, and provide useful information.

PURPOSE AND SCOPE OF THIS REPORT

To develop safe and easy to use systems one needs:

1. A set of research-based human factors design guidelines/requirements.
2. Analytical and simulation procedures that can be used to predict driver performance with alternative designs.
3. Methods for testing and evaluating alternatives.

This report explores the third point, test methods. The human factors test that is most appropriate depends upon the intended use of the desired information. Potential uses include:

- Identifying the strengths and weaknesses (problems) associated with a design.
- Exploring alternative designs.
- Determining how problems might be solved.
- Determining how common and severe the problems are.
- Determining if a system is fit (safe) for use.

The perception is that the DOT, and especially NHTSA, has traditionally emphasized safety and system effectiveness, but given less attention to ease of use. Evaluations should assess both the current level of performance and problems that need to be corrected. It is important to provide incentives to improve systems, not just determine minimum acceptability. In recognition of those broader needs, the contract called for quantifying "the influence of in-vehicle systems on driver safety...the effectiveness with which information is transferred...to drivers," and "assess[ing] driver comfort, convenience, and confidence when using these systems." A further goal was to select measures and test procedures "to assess the safety of drivers' performance while using in-vehicle systems." Further, it was desired to determine which levels of performance are unsafe. Establishing these levels is extremely difficult to do.

Therefore, this report examines the following questions:

1. How have basic and applied studies been conducted that relate to the safety and usability of driver information systems?
2. What are some of the key studies and what has been learned from them?
3. Which measures have been used to assess the safety and usability of driver information systems?
4. Which measures should be considered for future assessment protocols?

Selecting the appropriate test protocol and integrating measures of interest into test protocols is covered in a subsequent report.[18] Details of how measurements are to be collected, because they depend on the protocol selected, are also covered in that report.

To address these four questions, a section has been included describing how people drive cars and how that process has been modeled as a part of this project. Use of models is one method for identifying key measures and their relationships. Following that section are reviews of previous reviews. Several are quite insightful and provide schemes for grouping measures and approaches.

Most of the report is an indepth review of the literature, in particular 15 key papers, reports, and programs. Finally, to provide breadth to the review, a larger set of references is summarized in several tables. The focus is on general methodological studies and specific interface evaluations.

This report ends with a summary emphasizing those measures that reflect safety and ease of use.

For the most part, the discussion of measures is fairly general. However, it is recognized that a key application of this report is to five functions that have been chosen for further evaluation—(1) navigation, (2) traffic information, (3) cellular phone, (4) vehicle monitoring, and (5) the In-Vehicle Safety Advisory Warning System (IVSAWS), with navigation receiving greater attention that the other functions. In IVSAWS, radio transmitters are attached to road hazards (vehicles involved in an accident, police cars in a chase, etc.). Drivers nearby receive either visual or auditory warnings from an in-vehicle receiver.

This review of existing work does not specifically discuss other driver information systems [e.g., in-car signing, motorist services, and entertainment (radio, CD, cassette tape player, TV)], though many of the ideas are germane to the evaluation of those systems. Furthermore, systems for vehicle control (braking, steering, headway/speed maintenance, performance limits, such as rollover and traction, collision avoidance (back up, blind spot, long range obstacle detection, etc.), and driver performance monitoring are not considered, as they are outside of the scope of this project. Those systems all have interfaces that will communicate information to drivers, and, hence, their development and evaluation should benefit from the ideas presented here.

Lastly, readers are reminded that the ideas presented here focus on human performance, not crash biomechanics, or other topics. So, the failure of a central traffic control computer, or the consequences of electrical shorts, or problems associated with occupant impact during a crash are beyond the bounds of this research.

HOW DO PEOPLE DRIVE CARS?

To evaluate the effectiveness of driver information systems, one must understand the context in which they operate. Considering driving's significant impact on society (the typical adult probably spends an hour driving daily; motor vehicles are the leading killer of young adults), how people actually drive is not well understood. Most of the research has focused on what happens to people when they are involved in accidents and other matters pertaining to crashworthiness, not what happens beforehand (pre-crash). Further, very little is known about what behavior constitutes normal driving.

Driving consists of a set of tasks and activities requiring perception, cognition, motor response, planning, and task selection. The latter activity is particularly important, as the driver must often choose between attending to the roadway cues needed for vehicle control and other information sources competing for visual attention (e.g., rearview mirror, climate control, advanced in-vehicle display).

These activities are organized in a hierarchical manner as shown in the simplified conceptual model of the driving task presented in figure 1.[19] The top-level activity consists of setting overall goals (e.g., drive from point A to point B in the shortest time). A variety of subgoals, or "maneuvers," are formulated and satisfied over time in order to achieve the top-level goals. Maneuvers relating to automobile control include the relatively high-level tasks that determine the intended path of the automobile (e.g., pull into traffic, drive in the current lane, change lanes, turn right at the next intersection). In general, maneuver selection is a rule-based process, whereas maneuver execution is skill-based.

Having defined the maneuvers to be performed, the driver must then select the lower-level task to be attended to at any given instant. The task that is always competing for attention is that of vehicle control. It consists of maintaining lateral position and either speed or headway in a manner that allows the intended maneuver to be carried out safely and efficiently. Other tasks, which may or may not be adjuncts to the vehicle control task, will compete on an intermittent basis at frequencies that vary widely from task to task.

In general, each low-level task includes perceptual ("obtain information"), cognitive ("process and plan"), and motor ("execute response") components. These processes may be considered and performed concurrently for some tasks—especially the task of continuous vehicle control. For other tasks, such as reading a message on an advanced in-vehicle monitor and then turning it off once read, the perceptual and response activities are separated in time sufficiently to be considered sequential activities.

At the very least, task selection involves determining which low-level tasks need attention. For tasks that are interruptable (such as talking on a phone), task selection may also include the selection of the appropriate process (perception, cognition, or execution) as determined by the status of the task when last attended.

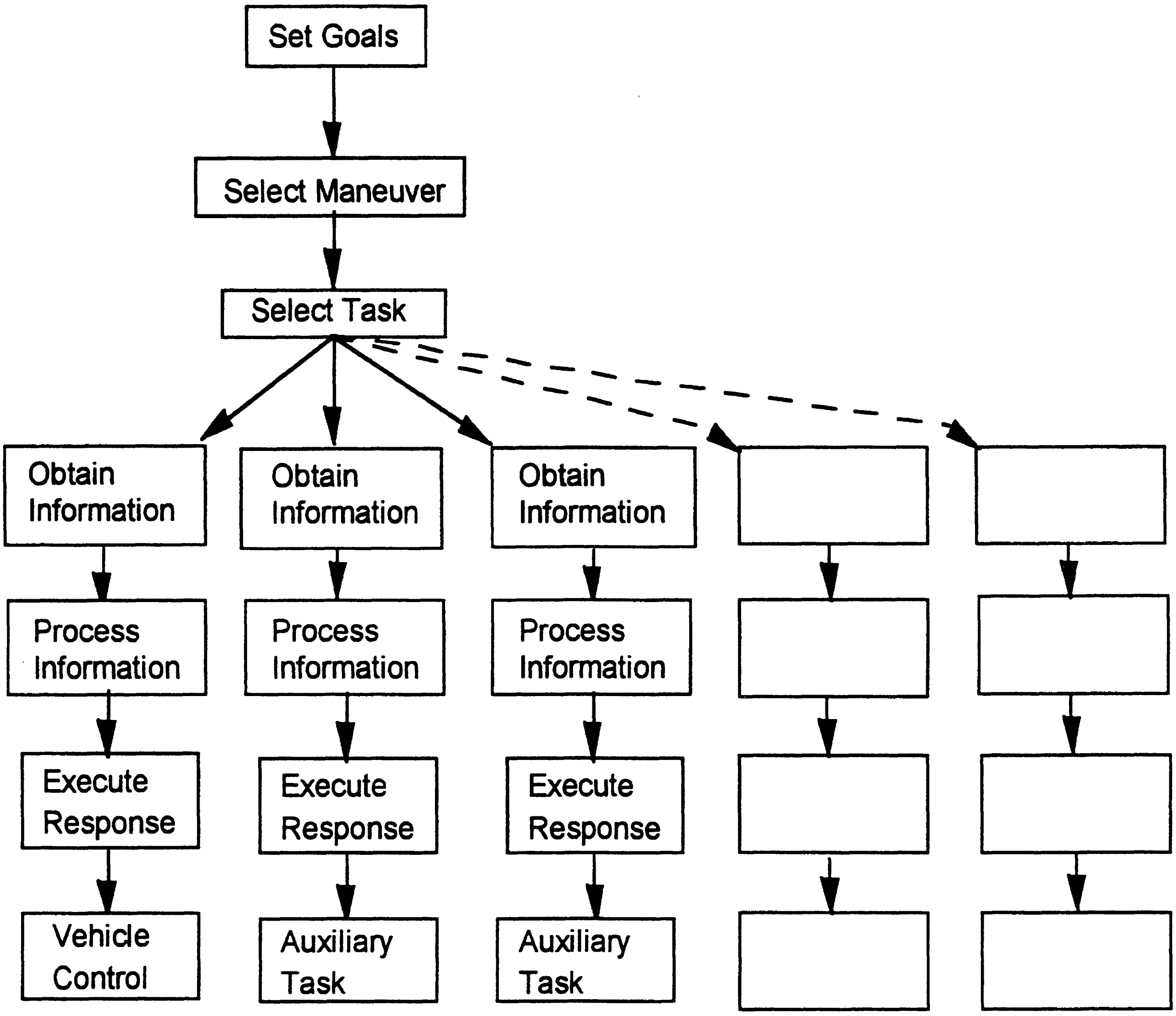


Figure 1. Simplified conceptual model of driving.

The task to be selected at any given instant is presumed to be based on a number of factors, including the perceived criticality to safety, the times to complete the various competing tasks, the penalties for not immediately attending to the tasks, driver preferences, and so on. The perceived priorities of the competing tasks will vary with time.

It is often assumed that because most tasks compete for visual attention, they can only be attended to sequentially. There is some empirical evidence, backed by the multiple-resource theory of Wickens, that certain tasks can be performed concurrently.[21] The competition among two or more tasks has the potential to cause performance degradation in one or more tasks, either because (1) performance of one or more tasks is delayed, (2) one or more tasks are dropped from the task queue, or (3) cognitive resources must be shared among tasks performed concurrently.

Nevertheless, when competing tasks cause an unacceptable degradation in performance—in particular, when safety is noticeably compromised—the driver is considered "overloaded."

Whether or not task overload occurs depends on the number of tasks competing for attention, the nature of these tasks, and the instantaneous state of the world. The driver may be able to deal with a number of items competing for attention if the driving task is relatively benign and if other tasks are easily rescheduled or are otherwise relatively undemanding. On the other hand, a single task in addition to the driving task may impose an unacceptable workload if the driving task is inherently difficult (e.g., driving a mountain road, driving in poor visibility) and if the auxiliary task is relatively frequent, of long duration, and not easily rescheduled. It is important to remember that given sufficient time, drivers make rational decisions. Drivers consider the extent to which attending to competing tasks and task components compromise safety.

This discussion makes several key points about driving that are critical to the analysis of the safety and ease of use of in-vehicle information systems. First, the extent to which a task interferes with driving depends primarily on the extent to which it is visual, because driving is primarily a visual task, though the motor demands are significant. Interference can also be caused by competition for cognitive and motor resources as well, or by their aggregate effect. Hence, interference depends on the degree to which a multiplicity of conflicts occur. This also complicates the measurement process as assessing the extent of interference will require multiple measures.

Second, the management of driving is intelligent; simply because there is an in-vehicle demand doesn't necessarily mean it will interfere with driving. However, the addition of a task may load some drivers to their limit, with additional tasks leading to a degradation of driver performance. Thus, in some situations, measurements of driver performance may show no effects of adding tasks, even though they are present (e.g., the elimination of reserve capacity).

Third, because the tasks are managed, tasks can be delayed (resulting in longer response times) or completed, but not as well as normally (resulting in increased error rates). It is difficult to know which outcome might occur (or which measure to collect) in advance, complicating the measurement process.

Fourth, the task management strategy generally adopted results in graceful degradation in the face of overload. Hence, identifying a single point at which a human-machine system transitions from safe to unsafe behavior will not be obvious.

Fifth, while safety and ease of use are connected, there may be instances where a system could have a minor impact on safety but is difficult to use and ineffective. For example, in heavy traffic drivers might forgo the task of paying attention to a particular navigation system if its tasks demands are high.

Thus, assessing the safety and ease of use of an in-vehicle information system will be difficult because one is considering a complex, adaptive system responding to external demands that vary as a function of time. Multiple measures of driver performance will be required, with the appropriate measures varying with the external task. The particular measures that will show degradation will depend on the in-vehicle task, with degradation likely to be gradual.

The point to be made is that there is no well understood, general model of driving behavior. To effectively select methods and measures to evaluate the specific impact of Advanced Traveler Information System (ATIS) technologies on driver safety and performance, the known factors that influence the driving task should be understood and should be explainable by a general model. This document is an important step in collecting that information and providing insight in to what methods have been successful in the research arena. Unfortunately, the connections of driving behavior and performance to an explicit conceptual model of the driving task are not strong as the model is still being developed.

PREVIOUS REVIEWS SIMILAR TO THIS ONE

This report reviews evaluation methods and measures for in-vehicle information systems. Three key efforts, the DRIVE task forces, Zaidel's review, and the work of Robertson and Southall are covered in the section that follows.[22, 23, 24] In a previous report completed as part of this project methods and measures are also discussed, though that topic is not the central theme.[3] For related information on this topic also see reference 25. The purpose of the section that follows is to give a sense of the types of measures that have been used.

DRIVE SAFETY TASK FORCE

This report was written by a committee, by correspondence.[26,27] The work was coordinated by the DRIVE Central Office. Contributing were several individuals well known for their contributions to driving research from the following organizations: University of Leeds Computer Studies Department, University of Leeds Institute of Transport Studies, University of London, Saab, University of Lund, HUSAT-Loughborough, BMW, and Traffic Research Center, Groningen. This report does not contain a listing of experimental measures, experimental protocols for assessing safety or ease of use, or, as alluded to by its title, design guidelines. Rather, it provides an overview of safety and human factors issues, and of the likely consequences if those issues are ignored.

The scope of the report is apparently much broader than driver information systems as it concerns all aspects of what are referred to as Advanced Transport Telematics (ATT). While ATT is not formally defined, its domain is apparent from the examples given.

The DRIVE report considers driver performance, safety, and driver comfort/convenience. Safety and human factors matters are classified into three categories: systems safety, man-machine interaction, and traffic system. The Task Force defines systems safety as problems that "arise as a result of a design fault or system malfunction."[22] For example, a traffic signal might show green to both roads at an intersection. The emphasis appears to be on approaches (e.g., fault tree analyses) in which there are a limited number of actions and consequences, and where their probabilities are readily identified. Man-machine interaction concerns "the usability of a system." Here the emphasis is on perceiving and understanding information and immediate reactions to it. Demanding secondary tasks associated with using in-vehicle information systems can lead to overload and distraction from the primary task, control of the vehicle. Traffic safety is a more global category which includes adverse behavioral changes in drivers. For example, if the automobile is equipped with an icy-road warning system, the absence of a warning might induce drivers to drive faster than they otherwise would have, under the assumption that a warning would be present if conditions were hazardous.

Primarily, those who will benefit from reading the DRIVE report are administrators who need a very general overview of what the issues are (particularly those of systems safety), and newcomers to the field who need to know about some of the likely problems. Those actively engaged in IVHS human factors research may find the report lacking in detail.

The DRIVE report contains several thought-provoking figures and tables in the man-machine interaction section. The first (table 1) concerns the context in which data are to be collected, which should be quite familiar to those engaged in research on this topic. Noteworthy is the distinction between real road test trials at a micro and macro level. The micro level refers to tests of single vehicles, the macro level to fleets. In the United States, macro tests are often referred to as operational field trials.

Table 1. Test environments.[22]

Increasing confidence that data correspond to real phonomena	↑	· Real Road Field Trials (Macro) · Real Road Test Trials (Micro) · Test Track Studies · Dynamic Vehicle Simulations · Static Vehicle Simulations · Part Task Evaluations	↓	Increasing control of variables and replication

Table 2 shows how well suited the Task Force believes each test context is for various evaluation dimensions. The check marks are never explained, but it appears they imply a higher level of suitability than the x's. The evaluation criteria are only generally defined in the DRIVE report.

Table 2. Task force's appraisal of various test environments.[22]

	TEST ENVIRONMENT					
EVALUATION CRITERIA	Real Road Field Trials	Real Road Test Trials	Test Tracks	Dynamic Simulators	Static Simulators	Part Task Evaluations
System Performance	√	√	x	x	x	x
User Workload	√	√	√	√	√	x
User Acceptance	√	√	x	x	x	x
Adaptation	√	√	x	x	x	x
System Effects	√	x	x	x	x	x
Task Match	√	√	x	x	x	x
Ease of Learning	√	√	√	√	√	√
Ease of Use	√	√	√	√	√	√

Table 3 lists some of the possible experimental measures. Notice that the list is quite extensive. Although not stated in the report, the task level distinctions are based on work by Michon, the behavior distinctions from work of Rasmussen. Strategic driving tasks are those associated with planning a complete trip, such as determining the route from Detroit to Ann Arbor. Tactical tasks are at a lower level and have a shorter time frame, and for example, concern the rules for changing lanes on an expressway. Operational tasks are associated with moment to moment control of the vehicle, such as the feedback-control loop for keeping the vehicle in the lane. Finally, reactive tasks are responses of the driver to higher level tasks. Reactive tasks are not voluntary, as are other classes of tasks.

Readers interested in evaluation should examine the original report for further details.

Table 3. Possible experimental measures.[22]

Driving Task Level	Driver Behavior	Tools	Measurement or Measurement Category
Strategic	Knowledge Based	Interviews	Structured Unstructured
		Verbal Protocols	Post Hoc Concurrent
		Questionnaire/ Survey	
Tactical	Rule Based	Vehicle Dynamics	Speed Acceleration Deceleration
		Visual Behavior	Patterns Glance duration Glance frequency
		Observation	Conflict studies Traffic flows critical incidents
Operational	Skill Based	Control Actions	Steering wheel rotation/reversal Pedal actuation Gear selection
Reactive	Autonomic	Psychophysiological	GSR, EKG, EEG, EOG, EMG, HR, HR variability, Adrenaline secretion

DRIVE 2 HARDIE PROJECT

This project, which began in January 1992, is a 3-year program to develop recommendations for the presentation of messages to drivers based on (1) understandability, usability, and safety while driving; (2) harmonization of text and symbols; (3) compatibility of different systems; and (4) harmonization with externally presented information.[28] Partners in this effort include the Transport Research Laboratory (UK), INRETS (France), Transportokonomisk Institutt (Norway), HUSAT (UK), Universidad Politecnica de Madrid (Spain), British Aerospace (UK), and TUV-Bayern eV (Germany), organizations well know for their contributions to transportation research. Tasks include reviewing existing standards, developing a theoretical framework and experimental methods for assessing in-vehicle driver interfaces, performing evaluations of demonstrator interfaces, and developing recommendations for information presentation.

Stevens and Pauzie found that most standards concern information presentation in offices, not in motor vehicles. In terms of European standards (their focus), the authors identify constraints from UK Construction Standard 109 and the Spanish telephone regulations. The 109 standard allows the use of video only for presentation of information about the state of the vehicle and its equipment, its location, route and destination, and information to help see the road. Systems such as electronic yellow pages may fall outside these limits. Spanish telephone regulations ban the use of hand-held phones while driving.

Further details will emerge from this project as it progresses.

REVIEW OF HUMAN FACTORS ISSUES RELATING TO ADVANCED DRIVER INFORMATION SYSTEMS

This is certainly the most comprehensive review of research concerning methods for evaluating human factors and advanced driver information systems to date.[29] The purpose of the report is to identify critical, generic human factors issues relating to advanced driver information systems and to propose a research agenda. This includes reviewing the literature relevant to Advanced Driver Information System (ADIS) evaluation; developing an evaluation framework; identifying relevant road, task, and measurement variables; discussing research needs; and describing example studies. The report is exhaustive in its coverage of the issues, including considerable work from DRIVE and PROMETHEUS, and is thoughtful in interpreting the results. It is less comprehensive in proposing specific research. It is clearly a report those involved with driver information systems must read. Zaidel's review of the literature will not be reiterated here. However, his ideas concerning evaluation methods are summarized in the section that follows.

With regard to research emphasis, Zaidel states the following: "There are presently neither conceptual reasons nor methodologic solutions to justify investing the large effort needed to simulate emergency situations."...[The] "navigation task appears to be a good overall cover-task for an evaluation scenario, even if the ADIS [Advanced Driver Information System] is not directly concerned with navigation."[29]

While describing many types of evaluation procedures (using rapid prototyping, videotape-based tasks, simple driving simulators, etc.), Zaidel favors carrying out on-the-road studies, with traffic level as a variable. He views traffic negotiation as being particularly important. In particular, he proposes the idea of using the subject's own vehicle and adding to it both the ADIS system and the instrumentation. The advantage of this approach is that the subject is already accustomed to his or her own vehicle and may behave more naturally. While this is an extremely interesting idea, there are significant challenges in terms of power and cooling requirements, and of interfacing the instrumentation to the vehicle (steering wheel, brake, accelerator). If the set of measures was limited, this may be possible, though the instrumentation would certainly not be inconspicuous. He realizes this approach is ideal but not always achievable. "The engineering development stage of the interface device, personal and organizational research preferences, and practical consideration influence the choice of the evaluation method as much as theoretical considerations."[29]

Zaidel comments that measures commonly fall into three categories: those that reflect simple vehicle control and guidance aspects of driving (traditional measures of driving performance such a speed and lane position), those that reveal time-sharing of ADIS with other driving functions (such as video records and verbal reports), and those that relate to the quality of the driving and the driving safety envelope (judgments obtained from driving instructors). Table 4 shows a more complete listing of these and other measures.

Table 4. Zaidel's list of potential measures.[29]

	Information Type		
Information Source	Driver State	Task Processing	Quality of Driving
Sensors	Physiological Control Input	Eye movement Speed variance Task performance Headway	Lane keeping TLC Obstacle avoidance
Driver Report	Self-evaluation TLX SWAT	Interface design Priorities Difficulties Sequencing	Error recovery Distraction Incidents Situation awareness
Expert Observer	Stress Load Inattention	Control errors Procedural errors Task sharing Strategies	Safety envelope Error recovery Anticipation Situation awareness
Traffic Data		Gap acceptance Speed distribution	Accidents

Note: TLX is the NASA Task Loading Index. SWAT is the Subjective Workload Evaluation Technique.

Of these measures, Zaidel believes that lane exceedance, glance frequency and duration, verbal comments, SWAT or NASA TLX scores, task completion times and errors, steering wheel motions (as indicators of attention), and expert judgments of the quality of driving, are the key measures to collect.

For those unfamiliar with the literature, SWAT is a method for rating workload on three dimensions: time load, mental effort load, and psychological stress.[30,31] To assess workload, a scale is developed in which operators provide an assessment of their perception of workload at three levels on each of the scales. They then rate the workload of specific tasks on each scale, from which numerical scores are developed.

TLX is a weighted score derived from subjective ratings on six scales (mental demand, physical demand, temporal demand, effort, performance, and frustration level).[32] Each scale has 20 equal intervals and verbal anchors (e.g., low and high). Scale weights are derived from paired comparisons of scales in which subjects indicate which of each pair contributed more to workload. The overall workload score is derived from the individual scale scores and the scale weights.[33]

Zaidel proposes some interesting ideas concerning development of research methods. In particular, he identifies the need for research on (1) methods to assess safety envelopes, (2) a driving-specific TLX scale, (3) rules for grouping traffic situations, (4) measures of driving workload (both momentary and overall), (5) the measurement of the quality of driving, and (6) protocol analysis of task processing. These topics are worthy of further investigation.

CODE OF PRACTICE FOR IN-VEHICLE DRIVER INFORMATION SYSTEMS

This report was commissioned by the Department of Transport (UK) to determine the feasibility of developing a Code of Practice for the designers, manufacturers, and users of in-vehicle driver information systems.[24] This project was carried out under contract to the Institute for Consumer Ergonomics (ICE). To identify the relevant literature, the authors searched several computer data bases, then made inquiries to several research organizations. The reference list and bibliography are comprehensive and contain many reports of which U.S. researchers may not be aware.

Those organizations were also asked questions regarding a Code of Practice. While the effort was apparently broad in attempting to get a wide range of opinions and perspectives, a statistical summary of those responses was not provided. From those efforts, ICE concluded that there was currently not sufficient data to set objective safety performance standards (e.g., eyes-off-the-road time), though a code of good practice could be developed based on current knowledge. Accordingly, protocols for developing performance standards should be given priority. Among the research organizations, there was almost unanimous support for a Code, but it should have a European and possibly global application with the flexibility to consider future research. It was clear the Code should be legally enforceable. The evaluation procedure should consider simulator versus road trials, individual differences (age, sex, vision, hearing, experience with displays), and the driving context (traffic, vehicle speed, lighting, weather).

SUMMARY

Thus, of the previous reviews, the work of Zaidel is the most insightful and the most complete. He also identifies the key measures to collect: lane exceedance, glance frequency and duration, verbal comments, SWAT or NASA TLX scores, task completion times and errors, steering wheel motions (as indicators of attention), and expert judgments of the quality of driving. However, in considering that recommendation, readers should bear in mind that the measure that is appropriate may depend upon the type of task for which the IVHS device is to serve as an aid: strategic, tactical, or operational.

WHAT ARE SOME OF THE KEY STUDIES OF DRIVER INFORMATION SYSTEMS?

While the previous section gives a sense of the types of measures one might use, it is important to have a detailed understanding of the measures that have been used. Following is a description of several key experiments pertaining to driver interface evaluation. Studies are listed in chronological order. (A more extensive listing appears in the tabular summaries in the section following this one.)

This section is intended primarily as a tutorial for human factors scientists with backgrounds in aviation or computing and as a refresher to scientists engaged in automotive human factors. While more general research on attention, time-sharing, and secondary task procedures is interesting background information, it is not reviewed here.[34,35]

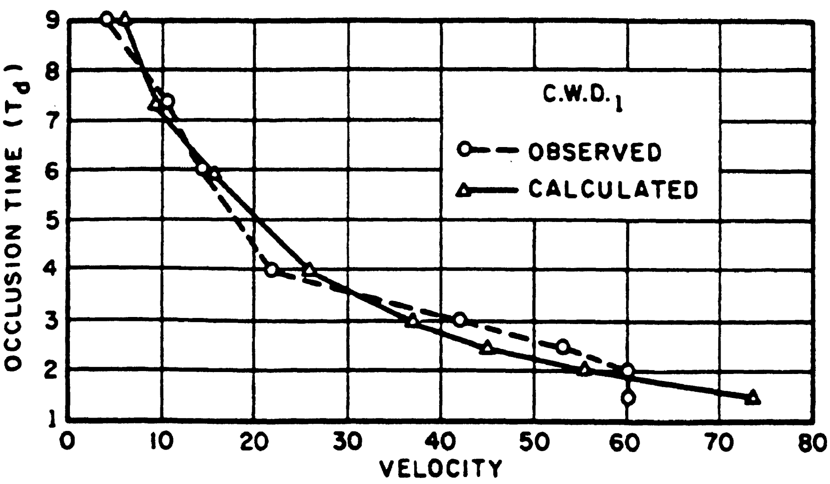
OCCLUSION STUDIES AND ATTENTIONAL DEMAND OF DRIVING

A particularly interesting method for assessing the attentional demand of driving was developed by BBN.[36] (See also reference 37 for a nearly identical technical report, and reference 38 for a more detailed, earlier report.) Their method limited the amount of time the driver could look at the road. A special helmet was devised consisting of a mechanism to lower a translucent face shield (the type used on protective helmets) that occluded the driver's view of the road. There were two experimental conditions: (1) the experimenter controlled lowering of the shield while the driver controlled his speed and (2) the driver controlled the shield while the speed was constant. The purpose of the research was to develop an empirical model relating the percentage of time the driver could look at the road to speed and other measures. (See reference 37 for details of the model.)

The research program consisted of two experiments, one conducted on an unused section of I-594 in Massachusetts with few curves, the second on a closed circuit sports car racing course. The 2.6 km (1.6 mi) track had 10 turns ranging from virtually straight to hairpin. In each case, there were two experimental conditions; either the occlusion (1.0 - 9.0 s) and viewing time (0.25, 0.5, 1.0 s) were fixed and the driver varied the speed, or the speed and viewing time were fixed and the driver varied how often the road was viewed. The test vehicle was a modified 1965 Dodge Polara sedan.

Five people served as subjects, including two authors of the occlusion report. Only one person served as a subject in all four experiments.

Figure 2 shows the relationship between occlusion time and terminal speed for two subjects. The term "calculated" in the figure refers to predictions of a model of driver behavior. (See reference 36 for details.) The data are for look times of 0.50 and 0.25 s, respectively. Both look times led to functions of the same general shape. There was little difference in the function when look times were extended to 1.0 s. The results are indeed remarkable. For example, using the data from the left panel in figure 2, one subject drove at roughly 97 km/h (60 mi/h) looking for 0.5 s every 2 s (for situations when there was no traffic).



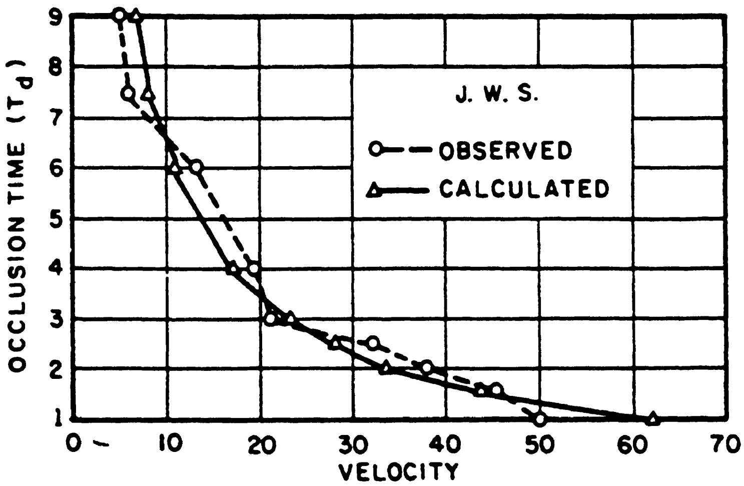


Figure 2. Relationship between occlusion time and speed for two drivers.[37]

Figure 3 shows speed vs. occlusion times for a look time of 0.5 s. Notice that individual differences were large. Further complicating matters is that occlusion times also varied with the radius of curves (larger radii had longer occlusion times) and their angular change (larger changes had shorter occlusion times).

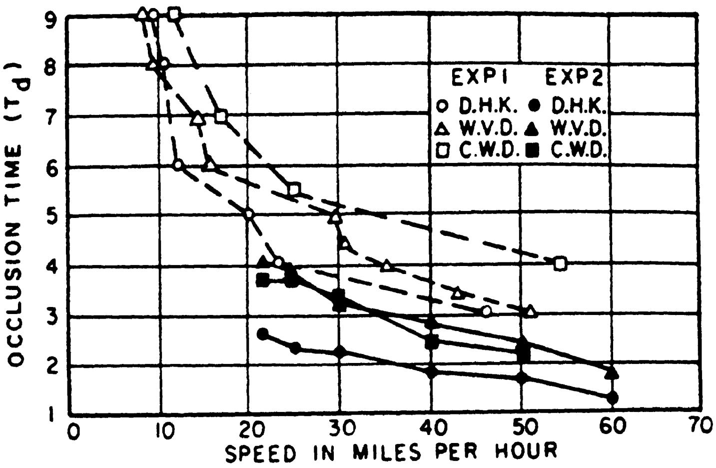


Figure 3. Speed vs. occlusion time for the race track.[37]

The Senders work makes two key contributions. First, it proposes and demonstrates the utility of the visual occlusion method as a means for examining driver workload. Useful data can be obtained by either fixing look time, occlusion time, or driving speed.

Second, this work presents a rigorous quantitative expression (not discussed here) that can be used to assess attentional demands. While other researchers have used the occlusion method, there have been few applications of the expression. Limited use of this approach is most likely due to the risk it presents to experiment participants and other motorists. Limited use the expression has occurred because it is difficult to relate to physical attributes of the road, traffic, or the driver.

COMPARISON OF WORKLOAD MEASURES WHEN DRIVING WITH CROSSWINDS

This study is one of the few to examine the merits of alternative methods of assessing the workload of driving based on the physical attributes of the road.[39] The paper also includes a reasonably comprehensive review of workload.

In this experiment, 30 college students drove the Virginia Tech Driving Simulator. It had a 6 degree-of-freedom, single channel visual display, a 4 degree-of-freedom motion base and 4-channel audio. The cab was open and the visual scene was austere. While driving at a simulated 89 km/h (55 mi/h), the effects of simulated crosswinds were imposed on the driving task. There were three levels of workload—low, medium, and high—achieved by varying the distance between the center of pressure of the crosswinds and the vehicle center of gravity. After considerable practice, data were collected for 90-s test trials. The three dependent measures were lateral deviation, yaw angle deviation, and the number of two-degree steering wheel reversals/unit time.

Five sets of secondary task measures were collected from subjects. In the digit-reading task, participants were shown a random sequence of digits by a seven-segment LED mounted in front of them. Initially, baseline data were collected to determine the presentation rate at which each participant could read 80 percent of the digits shown correctly (1.5 to 4.25 presentations/second) when the digit reading task was done alone. Subsequently, that task was performed concurrently with driving. In the heart-rate condition, an ear-mounted plythysmograph was used to measure heart rate, from which heart-rate variability was computed. In the rating scale condition, participants rated crosswind effects for each session on two 11-point scales (A = extremely harsh and troublesome, K = extremely small or imperceptible) and attentional demand (A = extremely high attention needed, K = extremely low attention needed). In the occlusion condition, each time the driver said "now," the road scene was presented for 200 ms.

The primary analysis of the data was an Analysis of Variance (ANOVA) of workload where, for each of the three dependent measures, workload was defined as:

$$100 * (1 - (\frac{\text{dual task performance}}{\text{single task performance}})) \qquad (1)$$

In terms of the primary task measures, workload variation led to significant differences in all three dependent measures. These differences were slightly more pronounced in steering wheel reversals than in yaw deviation, but more so than in lateral deviation. This suggests that steering wheel reversal is the most sensitive measure of driving workload of those examined.

In terms of the secondary tasks, only rating scale data reflected the differences in workload (crosswinds location). There were no differences between the two ratings obtained. Digit reading performance, heart-rate variability, and occlusion times were not affected by workload. As Hicks and Wierwille explain, there were many possible reasons why these measures did not show significant effects, such as lack of experience with the equipment, small sample sizes, etc. For example, when the occlusion method was used, participants were willing to traverse more of the lane than they might have for a real road.

From these results, Hicks and Wierwille recommend "primary task measures and rating scale measures (as constructed here) should be used in assessing driver workload, particularly if it is of a psychomotor nature."[39] Here the primary measures were steering wheel reversals, yaw deviation, and lateral deviation, and the rating scale measures were of attentional demand. The particular measures selected for each context require some thought. In this experiment, people were driving on basically a straight road and their course was perturbed by wind gusts. The immediate effect of such is to cause the vehicle to yaw suddenly, not move laterally, and hence yaw angle should reflect workload, as it did in the experiment. Further, the effect of crosswinds in driving are felt (as a torque on the steering wheel) before they are seen, but it is unclear what force feedback cues were provided in the simulation.

TIME-TO-LINE CROSSING AND DRIVING STRATEGY

Godthelp, Milgram, and Blaauw (1984) describe an evaluation of TLC or time-to-line-crossing, a measure of driving strategy.[40] In reviewing control models of driving, Godthelp, et al. note that most models assume drivers behave as error-correcting mechanisms who continually attend to the steering task. In contrast, Godthelp believes drivers behave as intermittent controllers, sampling the road scene, making corrections, and then not sampling for a while. Godthelp and others suggest that in making sampling decisions, the time-to-line-crossing is a critical factor. The time-to-line-crossing is how long it would take a vehicle to reach either lane edge if the steering wheel is not moved. At each moment the vehicle is assumed to have some heading error and may not be in the center of the lane. In some sense, this represents a margin of safety that drivers maintain.

Godthelp, et al. had six drivers steer an instrumented car on a 2-km (1.2-mi) section of an unused section of a 4-lane divided highway. While driving, they wore a bicycle helmet with a translucent visor that could be raised for 0.55 s looks whenever the horn button was struck. For each run the speed [20 to 120 km/h (12 to 75 mi/h)] was held fixed by a cruise-control-like device. Steering wheel angle, yaw rate, and lateral position were sampled at 4 Hz.

Godthelp, et al. found that the time-to-line-crossing (TLC) decreased as speed increased, with the 15 percent TLC being about 0.3-0.4 s greater than the occlusion times over the range of speeds examined. (See table 5.) Further, the data show that over the range of speeds examined, the time-to-line-crossing at the end (TLCe) of each occlusion period was about 1.57 times the occlusion duration. This suggests a constant relative safety margin (constant fraction of time) maintained by the drivers. Since looking away from the road has the same effect as occluding vision of the road ahead, it seems reasonable to propose that TLCe values (divided by 1.57) for roads might provide estimates of the time available to view in-vehicle displays.

Table 5. Values of occlusion and time-to-line-crossing.[40]

Speed (km/h)	Speed (mi/h)	T Occlusion (s)	15% TLC	TLCe (s)	TLCe/T Occlusion
20	12	5.32	6.7	8.88	1.67
40	25	4.23	4.5	6.33	1.49
60	37	3.45	3.9	5.32	1.54
80	50	3.15	3.5	4.77	1.51
100	62	2.67	3.1	4.35	1.63
120	75	2.38	2.9	3.74	1.57

Godthelp (1988) describes another experiment to strengthen the concept of alternating between open- and closed-loop driving.[41] The same instrumented vehicle, and probably the same highway [with 3.5-m (11.5-ft) lanes] from the previous experiment was used. The dependent measures were also the same; so, too, was the number of drivers. Only 3 speeds were examined—20, 60, and 100 km/h.

Participants were instructed to drive the car normally until a tone was presented. At that time, they were to ignore the path error and stop steering until the vehicle reached the point at which it could just be comfortable correcting the heading to avoid reaching the lane boundary.

From these data, two measures of time-to-line-crossing were obtained. TLCs, the value for when the steering wheel was turned, and TLCmin, the smallest value of TLC obtained just after the correction began. The mean values of both of these measures remained fairly constant with speed (TLCs=1.3 s, TLCmin=1.1 s) and were the same for both lane boundaries (left and right). Also remaining constant was the minimum lateral distance to the lane boundary (about 15 cm). Other dependent measures varied with forward speed: the lateral distance at which corrections were made increased, as did the mean lateral speed.

Thus, in this experiment, drivers resumed steering when there was a constant amount of time left before the vehicle reached the lane boundary, implying a temporal safety margin of just over 1 s.

OCCLUSION VS. HANDS OFF STEERING WHEEL

Zwahlen and his colleagues have carried out several studies that are relevant to the safety evaluations of in-vehicle displays. Zwahlen and Balasubramanian (1974) examined steering behavior when there was no visual input.[42] The rationale for this research is that drivers looking at an in-vehicle display are obviously not looking at (or paying attention to) the road ahead. For all practical purposes, not allowing drivers to look at the road and keeping their eyes closed should lead to identical driving performance since road-related visual input is required to steer. The purpose of Zwahlen's research was to determine acceptable eye fixation behavior, namely how long and how often drivers could look away from the road, and still adequately maintain lateral position.

Two 23-year old students drove a 1965 Volkswagen sedan and a 1971 AMC Ambassador down an airport runway. A container of liquid dye attached to the rear bumper, dripped regularly onto the pavement when the car was driven. Subjects drove at 16 to 64 km/h (10 to 40 mi/h). When they reached a starting point, they closed their eyes until they had traveled either 153 m (500 ft) or had reached the side of the runway. There were 47 runs with no visual input and 13 with no steering control. (The subjects took their hands off the wheel.) Path deviations were recorded every 4.6 m (15 ft) to the nearest 1.2 cm (1/2 in).

Originally, Zwahlen proposed that uncertainty of the lateral position of the vehicle (the standard deviation in lateral position) could be derived from the work of Senders and others, as shown below.

$$s_y = k * V^2 * T^{1.5} \qquad (2)$$

Substituting V = D/T

$$s_y = k * D^2/T^{.5} \qquad (3)$$

where: s_y = standard deviation to vehicle displacement from the centerline
k = constant
V = velocity
T = occlusion time
D = distance traveled at constant speed while occluded

Using the experimental data for 74 km/h (46 mi/h) and 64 m (210 ft), k = 0.0004. Plots of occlusion distance versus standard deviation of road position suggest that for this value of k, the standard deviation at each distance is greater at higher speeds, which is the opposite of what Zwahlen's data show.

A second analysis, based on the steering model of Wier and McRuer described in their paper, assumes the initial heading error is negligible. (See the original paper for details.) Based on that analysis Zwahlen proposes the following expressions of the standard deviation of lane position:

$$s_y = k * D * T^{0.5} \tag{4}$$

A summary of the experimental results for one person and one car appears in figure 4. From those data, Zwahlen and Balasubramanian proposed k = 0.025 (0.683 in metric units) for steering with no visual input. As a check, using the data in that figure, a value of 0.0272 was computed for D = 110 m (360 ft), close to the reported value. For that same distance for no steering control, k was calculated to be 0.0467 from the figure.

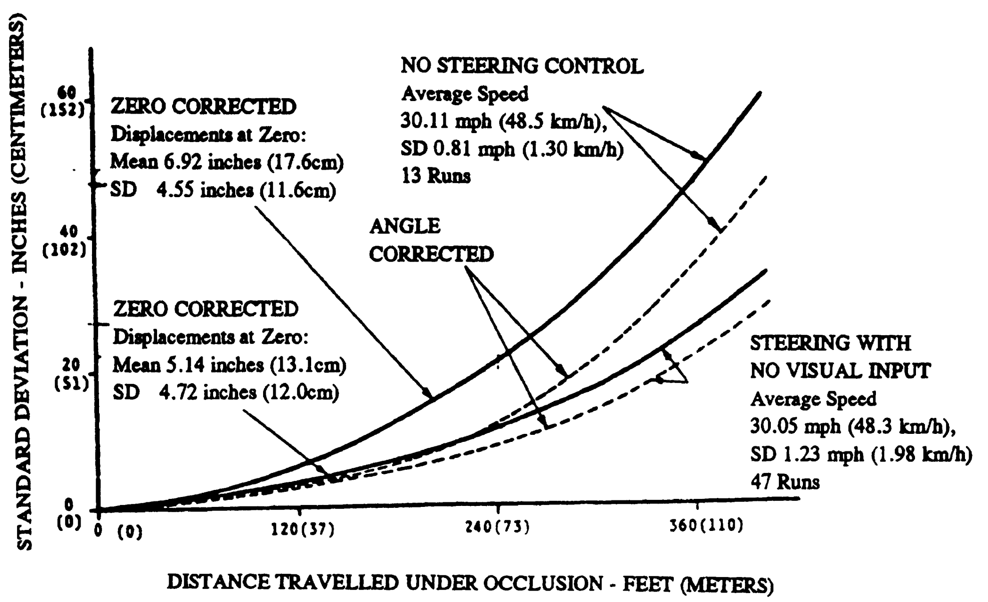


Figure 4. Standard deviations of vehicle displacements for no steering and no visual input.[42]

The actual standard deviation as a function of time is larger than computed using these equations which are for the zero-corrected case, this assumes drivers started in the middle of the road. In fact, their initial positions were 13 to 18 cm (5 to 7 in) off-center with a standard deviation of 11 cm (4.5 in). No heading error was assumed.

To apply this approach to safety assessments, additional data on driver, vehicle and road surface differences are needed to compute the appropriate values of k. With that information, safety criteria for inattention (at least for straight-road driving) could be computed.

Occlusion Vs. Text Reading

Zwahlen and DeBald had 12 people drive either a large or small car, again recording lane error using liquid dye.[43] Mounted midway in the center console were either article clippings or sections of a road map. Participants drove at 48 km/h (30 mi/h) and at the zero point either drove normally, with their eyes closed, or began reading text inside of the car. How drivers were to divide their attention in the reading task is not described, though it is assumed they were not to look at the road.

There were no significant differences in the variance of lateral deviation between cars for the first 15 m (50 ft). Interestingly, there was also no difference in the lateral variance between the eyes closed and reading conditions for distances up to 69 m (225 ft) (5 s at 30 mi/h). This suggests that for moderate distances, closing one's eyes has an identical effect on driving as looking inside the vehicle for information. Figure 5 shows the standard deviation of lateral deviation for the conditions examined.

Using the second model described previously, Zwahlen and DeBald estimate k = 0.076 (averaging across car sizes) with a mean absolute error of 27 cm (10.7 in) for driving with eyes closed. (From the graphics presented, a value of 0.073 is estimated.) This estimate of k is about triple the earlier estimate. For reading text k = 0.041 with a mean absolute error of 11 cm (4.4 in). (A check from the figures provided gives an estimate of k of 0.042—quite close.) The k for reading text is about 50 percent of that for eyes closed.

Figure 6 shows a comparison of the measured and estimated values for reading text. Notice the difference is a maximum at about 122 m (400 ft) [about 25 cm (10 in) less than the 140 cm (55 in) predicted]. For engineering estimates, this 20-percent difference is a bit large. In applying the model, it is important to consider the critical range of the lane variation when computing k. Also, the estimate for k is much lower [close to 0.03 for distances something less than 92 m (300 ft)] is the initial mean error [20 cm (7.7 in)] is included in the calculation (that is, $s_y = k * D * T^{0.5} + 7.7$).

To put these numbers in context, Zwahlen gives the example of a 1.8-m (6-ft)-wide car in a 3.6-m (12-ft)-wide lane. A lateral deviation of 1 m (3 ft) would put the driver at the edge of the lane. At 48 km/h (30 mi/h) for a reading task of 2, 4, and 6 s, respectively, the chance of leaving the lane (100 * probability) are shown in table 6. The chances computed are for two one-tailed tests, since lane departure can occur in either of two directions. However, since there is an initial bias, the departure is most likely in the direction of the bias, though probabilities must be calculated in both directions. The values reported in the paper do not agree with the author's calculations, either using the reported value of k, or a somewhat smaller value that provides a better fit to standard deviations at shorter distances. Clearly, the probability of departing from a lane (and concern about claims of risk) are very dependent upon the estimates of k.

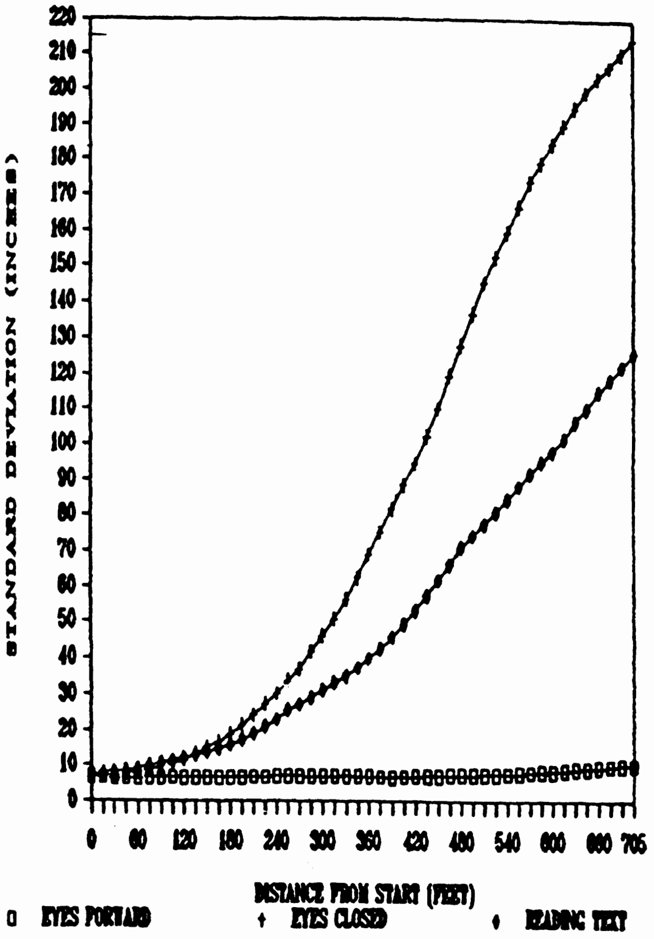


Figure 5. Standard deviations versus distance traveled.[-3]

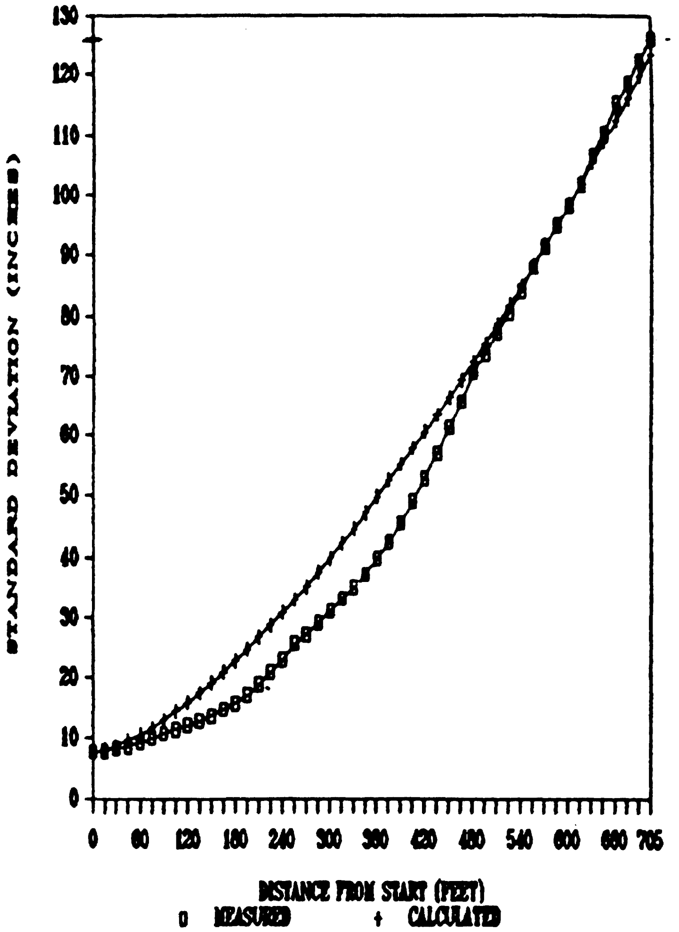


Figure 6. Comparison of measured and predicted standard deviations.[43]

Table 6. Recomputed standard deviations.[43]

Lane W (ft)	Distance D (ft)	Time T (sec)	Sy (in) k=0.041	Sy (in) k=0.030	Chance of Departure (%) (author, k=0.041)	Chance of Departure (%) (author, k=0.030)	Chance of Departure (%) (Zwahlen, p. 260)
10	88.	2	5.1	3.7	0.07	0.01	1.25
10	176.	4	14.4	10.6	15.31	6.31	6.30
10	264.	6	26.5	19.4	38.99	25.10	18.41
12	88.	2	5.1	3.7	<0.01	<0.01	0.04
12	176.	4	14.4	10.6	2.60	0.39	1.10
12	264.	6	26.5	19.4	19.41	8.44	8.69

Note: $Sy = k * D * T^{(.5)}$

bias = 19 cm (7.7 in)

Occlusion Vs. Operation of Panel Components

This is the best known of Zwahlen's recent studies.[44] He presented eight young drivers with a paper mockup of the CRT touch screen in the 1986 Buick Riviera. It was mounted in a 1981 Oldsmobile Cutlass in a position comparable to the Riviera's center console or in a lower position. The "airport runway strip chart technique" was again used to record lane position. Drivers accelerated to 64 km/h (40 mi/h) and then used the simulated touch panel to turn the radio on, adjust the volume, etc. Drivers either looked directly and continuously at the simulated panel until the task was completed, or could look at the road ahead as necessary. Drivers' visual behavior was observed by an experimenter seated next to the driver. Given the description of fixation behavior provided in the original document, readers are left with the impression that "looks" were classified as either inside or outside of the vehicle. Hence, each look could consist of multiple fixations.

Figure 7 shows the uncorrected standard deviations of lateral lane position. Some aspects of this figure make sense, while others do not. In general, the standard deviation of lane position error increases with time in the beginning of the run and then, as the in-vehicle task is completed, decreases. For all conditions drivers had completed their tasks after they had driven 244 m (800 ft). Notice that overall the standard deviation for Condition B (high, not looking) was approaching 0.6 m (2 ft) in mid-run, but, for Condition D (low, not looking) was close to 0.3 m (1 ft). From a practical perspective, a 0.3 m (1 ft) difference in deviation matters. But what is unexpected is that the data for the looking conditions (A and C) were between B and D. They should be less than B and D. Zwahlen, et al. reported that subjects usually glanced outside the vehicle after a page change (from radio to climate). It could be that because external demands were low (there was no other traffic) drivers attached a lower priority to steering and it was no longer treated as a "protected" task.

With regard to the in-vehicle task times, Zwahlen, et al. report they are normal with means and standard deviations of 5.02/0.98 and 8.39/1.63 s for the radio and climate control tasks, respectively.

Figure 8 shows the number of looks to the displays for each task. There were no significant differences due to display location (2.79 looks for high, 2.67 for low for the radio, 1.42 for high vs. 1.46 for low for the climate control). There were four instances where drivers did not look outside while operating the radio and three instances for the climate control.

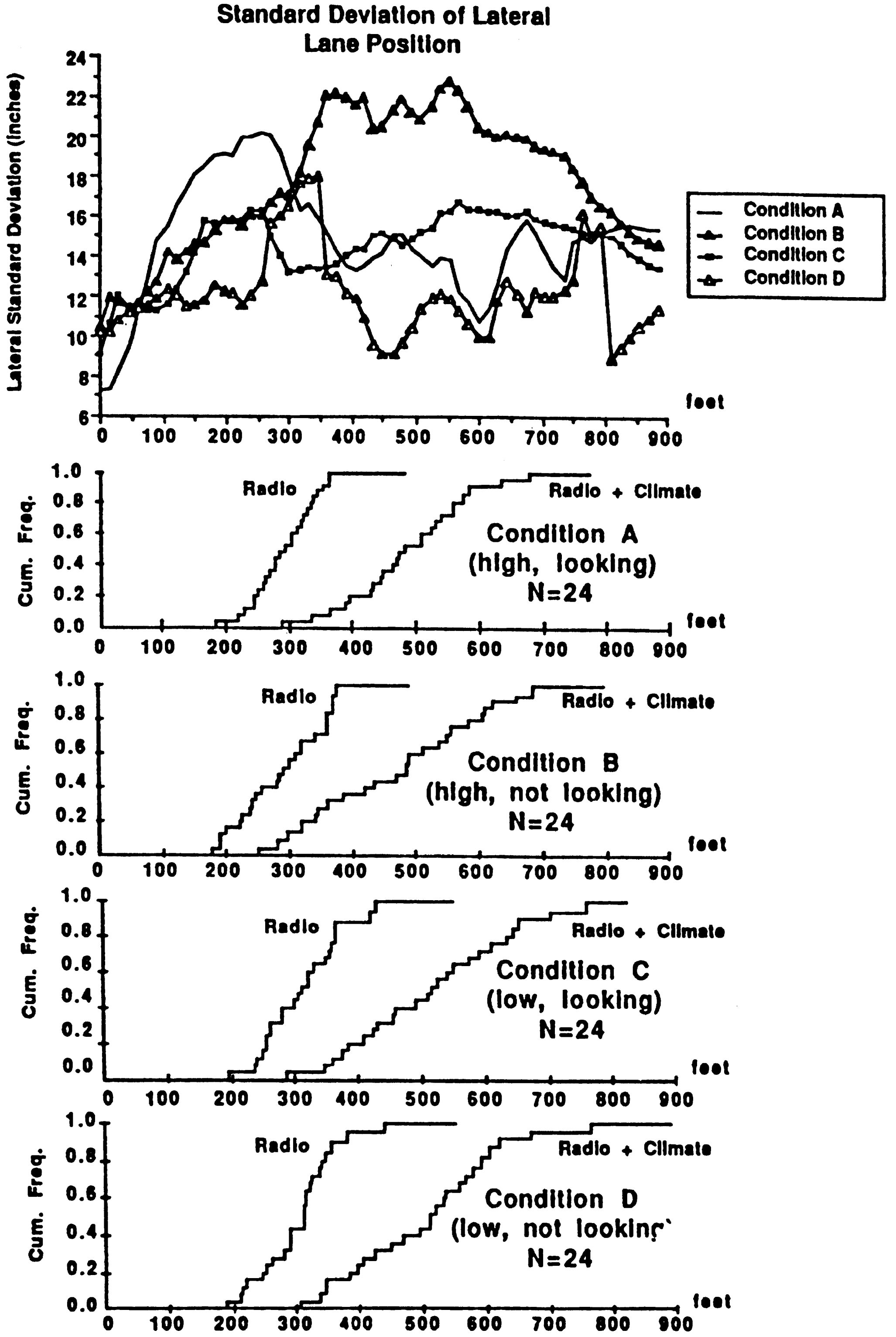


Figure 7. Uncorrected standard deviation of lateral lane position.[44]

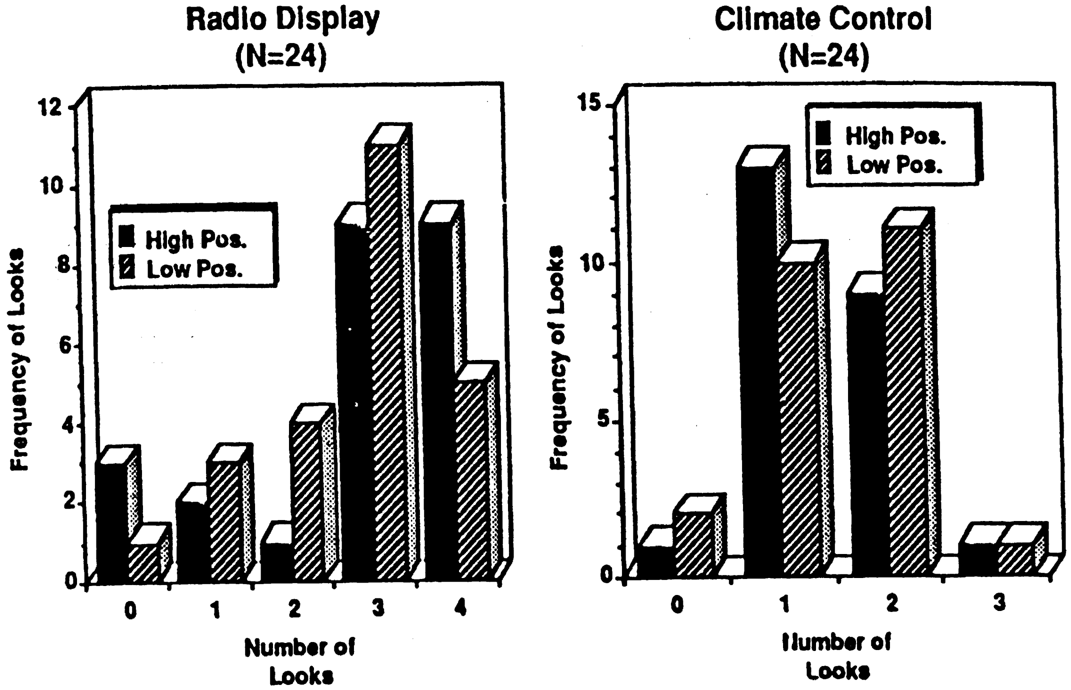


Figure 8. Direct looks made by drivers to the radio and climate control.[44]

Combining the data for all four conditions for the entire run [0 to 270 m (885 ft)], Zwahlen calculates a standard deviation of 42.2 cm (16.62 in). Using that value, he estimates that under ideal conditions (dry pavement, daytime, etc.) at 64 km/h (40 mi/h), there is a 3-percent chance a driver would deviate from a 3.6 m (12-ft) lane while operating a touch panel in a 1.8-m (6-ft) wide vehicle. (Note: On each side of the vehicle is a 1-m (3-ft or 36-in) clearance. Hence, 36/16.62 = 2.166 = z. According to the normal distribution tables, the 1-tailed p is 0.015, so for the 2-tailed case the probability is 3 percent.) For a 3-m (10-ft) lane, the probability of lane exceedence is 15 percent.

What is wrong, then, with these data? In brief, the problem is with the task demands. Because there was no traffic, it is likely steering was given a lower priority than normal. Drivers gave priority to the internal task over lane maintenance, something they would not do while driving on the highway. As described elsewhere, research from Wierwille and his associates shows that drivers alter their attention allocation based on external demands, both anticipated and unanticipated. That does not mean, however, that these data should be discarded.

From these and other data, Zwahlen et al. propose the design guide shown in figure 9. The constraint on the number of fixations comes from his work on pavement markings, and from Senders' classic study of the attentional demands of driving.[36,37] Specifically they say the following:

> The areas...are based partially on the conceptual model...which suggest[s] that the number of consecutive looks required to obtain a specific chunk of information...while

driving along a straight path needs to be limited to about three, partially based on the results of ...[his work on edge markings, Senders et al.]...which indicated out of view times (rear view mirror, speedometer, etc.) or occlusion times in the range of 0.5 to 2.0 seconds for tangent sections, 0.32 to 0.34 second for curve sections and partially upon experimental results from ...[references 42 and 43]...and this study which indicate that the lateral lane position standard deviations and lane exceedence values reach unacceptable values after 2 to 4 seconds of occlusion of visual road/traffic input or when working...inside a vehicle.[44]

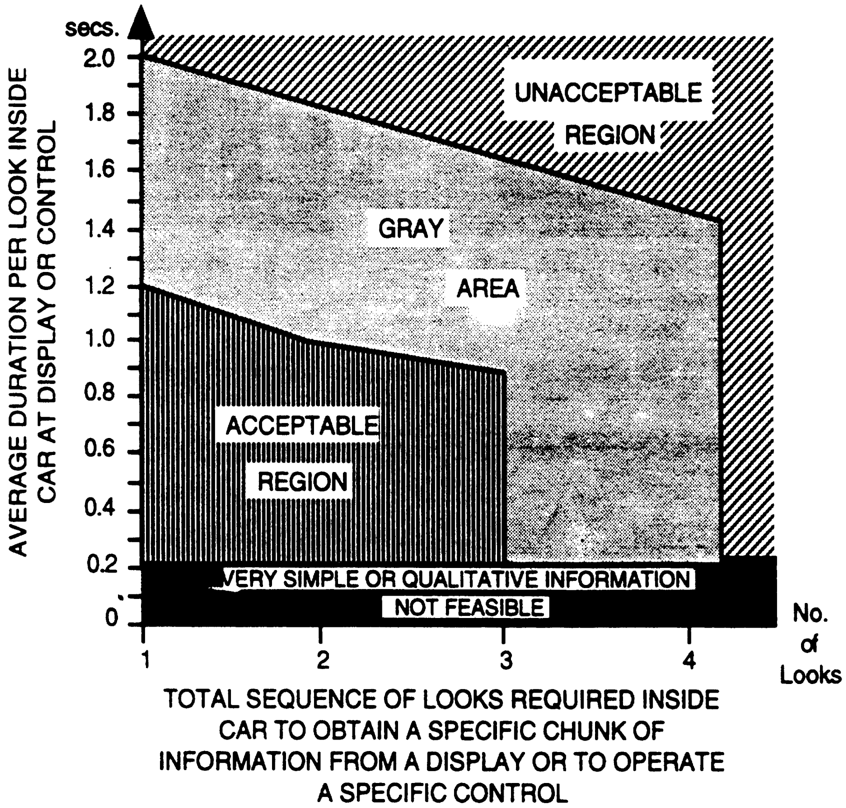


Figure 9. Proposed design guide for in-vehicle displays.[44]

Extrapolating the data from Senders' famous helmet study described earlier, to 105 km/h (65 mi/h) (the highest posted speed on U.S. roads) for straight roads, an occlusion time of 1.0 s is associated with a viewing time of 0.5 s. This suggests that, for the easiest of driving contexts (a straight road in daylight with no traffic and an experienced person), distractions from the road 1 s (with 1/2 s in between) is the most drivers will accept. (Or, this is the least the three drivers tested would accept.)

Occlusion Vs. Phone Operation

In two related experiments, a total of 20 young drivers steered a car equipped with an automatic transmission at 64 km/h (40 mi/h) down an airport runway instrumented in a manner identical to the previous studies.[45] A standard pushbutton phone (simulating a cellular phone) was installed in either a 1982 Pontiac station wagon (part 1) or a 1985 Plymouth Turismo (part 2). It is not clear what feedback each keypress provided, if the dialed number was displayed, and if there was a send/dial key to be pressed. There were four conditions (car phone mounted low/high on the dash vs. driver allowed/not allowed to look at the road while dialing). The 11-digit number to be dialed was on a piece of paper near the phone. Drivers did not correct dialing errors (as they would with a real cellular phone).

Figures 10 and 11 show standard deviations of lateral position and dialing task completion times. These lane deviation data make more sense than those in the previous study. In general, in terms of increasing lane variance, the order was look/high on the panel, look/low, no-look/high, no-look/low. Statistical tests of these differences are not presented. Initial lateral standard deviations ranged from 13 to 18 cm (5.2 to 7.1 in), with most of the deviations being close to 14 cm (5.5 in). Table 7 shows the standard deviation of lateral deviation when dialing was completed. Notice that both positioning and look/no look affect these values. For 6 of the 8 cases shown, the maximum standard deviation was close to 0.6 m (2 ft), which would put the lane exceedence probability at 30 percent, a large value.

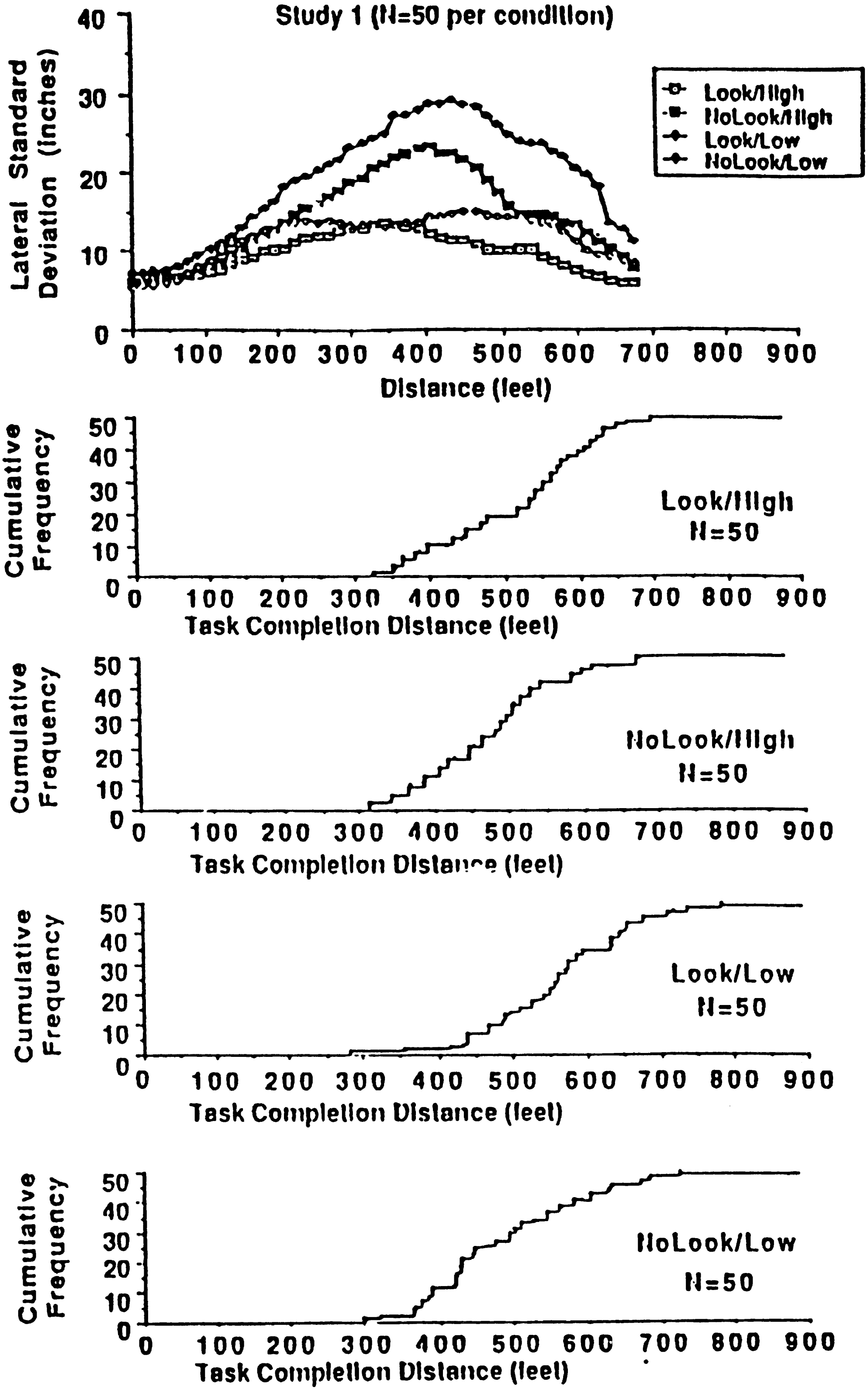


Figure 10. Standard deviation of lateral position and task completion times for the Plymouth Turismo.[45]

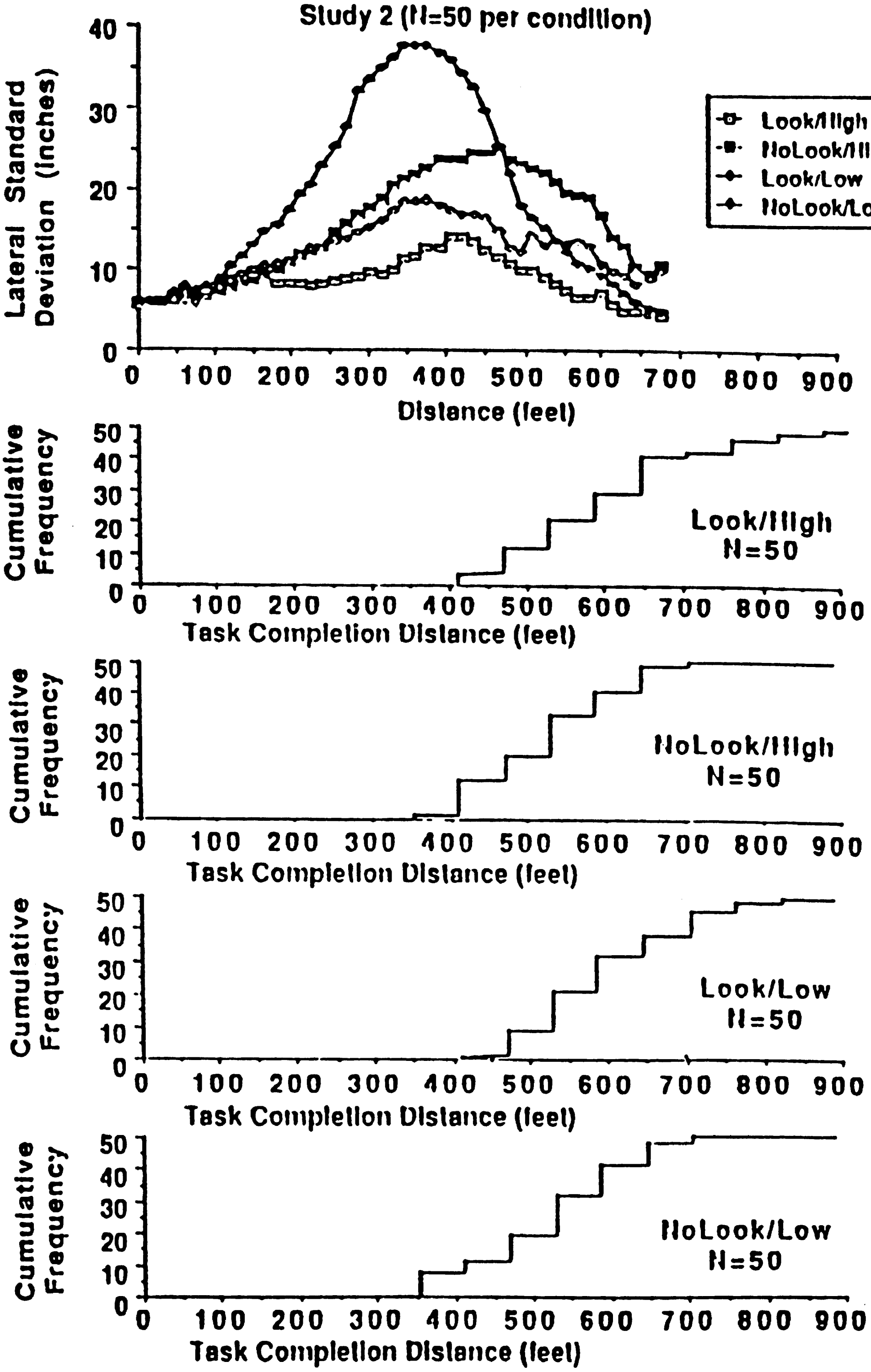


Figure 11. Standard deviation of lateral position and task completion times for the Pontiac Wagon.[45]

Table 7. Standard deviation of lateral position.[45]

	Look/High	No-Look/High	Look/Low	No-Look/Low
Dialing Completed				
wagon	8.8	20.7	11.1	25.2
small car	7.0	17.5	15.0	20.8
Maximum				
wagon	16.8	27.7	23.0	38.1
small car	16.6	27.2	25.3	40.7

With regard to the dialing task, the distributions appear normal, with an overall mean of 9.1 s (or 0.83 s/digit) and a standard deviation of 1.72 s. Looking at the road increased dialing times slightly but panel location had no effect (look/high = 9.5 s, no look/high = 8.5, look/low = 9.8, no look/low = 8.6). For the wagon, the number of outside looks was 2.9 times/dialing sequence (standard deviation of 1.1 looks) for the high location and 2.2 times with a standard deviation of 1.4 looks for the low location. Dialing errors were few and did not differentiate between conditions. There were three errors for each of three conditions and two in the other.

Averaging across both vehicles, the standard deviation for lateral position was 39 cm (15.4 in) for the 197-m (675-ft) run. For a 3.7-m (12-ft) lane, Zwahlen claims that 1.9 percent of the dialing tasks would result in a lane excursion under ideal conditions (daytime, dry pavement, straight road, etc.) at 65 km/h (40 mi/h). For a 3-m (10-ft) lane the associated quantity is 11.9 percent. Zwahlen considers these amounts to be unacceptable to driver safety and the use of design enhancements to prevent phone dialing in curves or heavy traffic, the use of voice recognition input, and other modifications. He also argues that this evidence can be used to support his design guide, though it is not clear how the results here are linked to the specific recommended values in the design guide.

As with his previous studies, the main criticism of this research is the nature of the task to be performed. Because there was no traffic or other significant external demands, steering was not a protected task and hence lane variation was much larger than it should be while driving. It does, however, represent a worst case scenario, such as a very important phone call to which the driver devotes more attention than is reasonable. Wierwille's research, described elsewhere in this section, argues that drivers behave in a manner different from this experiment when in traffic.

Zwahlen's experiments are covered in detail because they serve as the basis for his design guide, the only quantitative one in the literature at the moment. Those data have been used as a basis for arguing against including phones, navigation systems, and other driver information systems in cars. "Based upon the results of this study, the development and introduction of sophisticated in-vehicle displays and/or touch panels requiring consecutive eye fixations of several seconds inside of the automobile should be halted."[43] "The introduction of in-vehicle CRT touch panel controls which require a number of consecutive eye fixations and a fairly stringent eye-hand-finger coordination and touch accuracy for several seconds inside a moving vehicle should be reconsidered and delayed."[44] "The results of these two studies, especially the lane exceedence probabilities, appear to indicate that the use of cellular pushbutton telephones in automobiles requiring finger touch inputs of relatively long number strings while driving, should be reconsidered and may be unacceptable from a driver safety point of view."[45]

UNSAFE DRIVER ACTIONS

This experiment developed a scheme for classifying unsafe driving actions and examined its application.[46] As noted earlier, it is an approach that Zaidel finds attractive. Forty-eight young and middle-aged Australians participated. After driving the test vehicle (Mitsubishi Sigma) around the test facility grounds, participants drove a 60 km (37 mi) route that took 90 to 100 minutes to complete. The route included major and minor roads, local streets, and intersections with traffic signals and roundabouts. Subjects were guided by an experimenter in the front seat. Quimby, seated in the rear, identified unsafe driver actions as they occurred. While driving, subjects were encouraged to comment on aspects that might be dangerous either to the driver or others. Comments were recorded on audio tape. No other instrumentation for recording driving was present.

For the purpose of this research, an unsafe driving action (UDA) was considered to be, "any action or lack of action on the part of the driver that increased their risk of an accident"[46] Poor driving practices, e.g., failing to signal, were considered UDAs only if other road users or vehicles were involved. There were 28 types of UDAs (e.g., too fast for conditions, following too closely) combined into 15 categories.

Table 8 shows the number and severity levels of different types of UDA's. UDA's were classified as slight or serious, and as being conflict or nonconflict types. The associated percentages are shown in the table. Conflicts were situations "where the subject, or another driver, pedestrian or cyclist, had to take action, such as braking or steering, to avoid an accident occuring."[46]

Participants made a total of 2016 UDA's (roughly 42 per driver, 1 every 2 min, or 1 every 0.7 km (0.4 mi) of driving). The number per driver ranged from 10 to 113. Most common were following too closely, driving too fast for conditions, and positioning while turning (e.g., being too far in the intersection at the start of a turn). Notice that the relative frequency with which particular UDA's lead to serious conflicts varies between UDA's.

These data were examined in considerable detail with regard to age, sex, personality, maneuver type, etc. The intent was to compare these data with accidents at several intersections along the route. There was little correlation between the two sets of data, though the number of accidents used in the correlation was small. It was thought that part of the problem may have been that other behaviors should be recorded as well.

Thus, while this research provides an interesting insight into the development of the method for certifying safety, further work is needed before it can be applied to the evaluation of in-vehicle systems. Some of that work is being carried out as part of the DRIVE program. Validation of the method is also required.

Table 8. Number and severity of different UDA's.[46]

Type of UDA	# Committed	% Serious	% Conflicts	% Serious Conflicts
Too fast for conditions	328	10.4	4.0	0.6
Following too closely	414	27.8	8.7	2.4
Emerging into traffic	176	30.1	70.5	28.4
Emerging when stopping	16	43.8	62.5	31.3
Turning across approaching traffic	32	34.4	40.6	15.6
Late through traffic signal	18	22.2	0.0	0.0
Overtaking/passing vehicles	70	10.0	15.7	5.7
Positioning going ahead	156	5.8	8.3	0.6
Positioning while turning	351	5.4	4.8	0.0
Observation/anticipation	178	10.7	33.7	6.7
Erratic maneuver	44	9.1	27.3	6.8
Giving/taking priority	72	13.9	27.8	4.2
Rear observation	34	5.9	2.9	4.4
Signaling	68	2.9	4.4	1.5
Steering/hitting curb	59	1.7	0.0	0.0
Total	2016	14.7	16.5	5.8

TIME-TO-COLLISION

Van Der Horst and Godthelp (1989) describe two techniques for collecting data on driver behavior: an instrumented vehicle and video recording of traffic scenes.[47] The car contains potientiometers for recording steering wheel and throttle angle, switch to record shift lever position, a pushbutton unit an experimenter can use to code driver actions, sensors for speed and distance travelled, and a lane tracker to determine lateral position. In addition, attentional demands can be assessed using a device to occlude the drivers vision momentarily, and physiological responses (heart rate, respiration rate, galvanic skin response, etc.) can be measured. This vehicle is used for evaluating experimental visual and auditory driver interfaces.

To record traffic scenes, the scientists at the Institute for Perception (TNO) mounted a video camera about 4 m (13 ft) above the road surface, generally on a lamppost, balcony, or building close to a location of interest. In some situations there may have been more than one camera. Video frames were recorded on a VCR and time coded.

To analyze the data, tapes are played back frame by frame and the position of key objects is digitized manually. This is believed to be a very labor-intensive and time-consuming task. The on-screen coordinates are then transformed into road coordinates via computer software. This information is then used to compute predictions of vehicle velocities and time-to-collision (TTC). TTC is defined as the time for two vehicles to collide if they were to continue at their present speed and on the same path. The operational rationale behind this measure is to encourage drivers to attempt to maintain a protected space around their vehicles so that if an unexpected event should occur (for example, a lead driver suddenly braking), they will generally have adequate time to react and avoid a collision.

Van Der Horst and Godthelp, 1989, p. 79 note, "In general, only interactions with a minimum of 1.5 seconds are considered critical.[47] Trained observers are able to consistently apply this threshold value." However, Van Der Horst and Godthelp did not explore the ability of trained observers to identify a range of values, something that would clearly be difficult to do on a continuous basis. Hence, digitized data is still required. While the evidence on TTC is far from complete, Van Der Horst believes that drivers use this parameter in making maneuvering decisions.

IN-VEHICLE STUDIES OF NAVIGATION

Wierwille and students working with him have carried out several experiments that provide quantitative data on the usability of navigation systems and that suggest a method for evaluating them. (See reference 3 for a more complete description.) All experiments involved an instrumented 1985 Cadillac Sedan DeVille fitted with an ETAK navigator (with the smaller screen). The car was equipped with a camera to record the forward scene and a second aimed at the driver to determine the direction of gaze. Steering wheel movements, speed, and foot control use were recorded by a computer. Lane excursions were recorded manually. Experiments typically involved 20 to 30 people varying in age. Test routes involved streets, two-lane State roads, and expressways, taking about 20 min to drive. Prior to testing, drivers were given extensive training in the use of the navigation system.

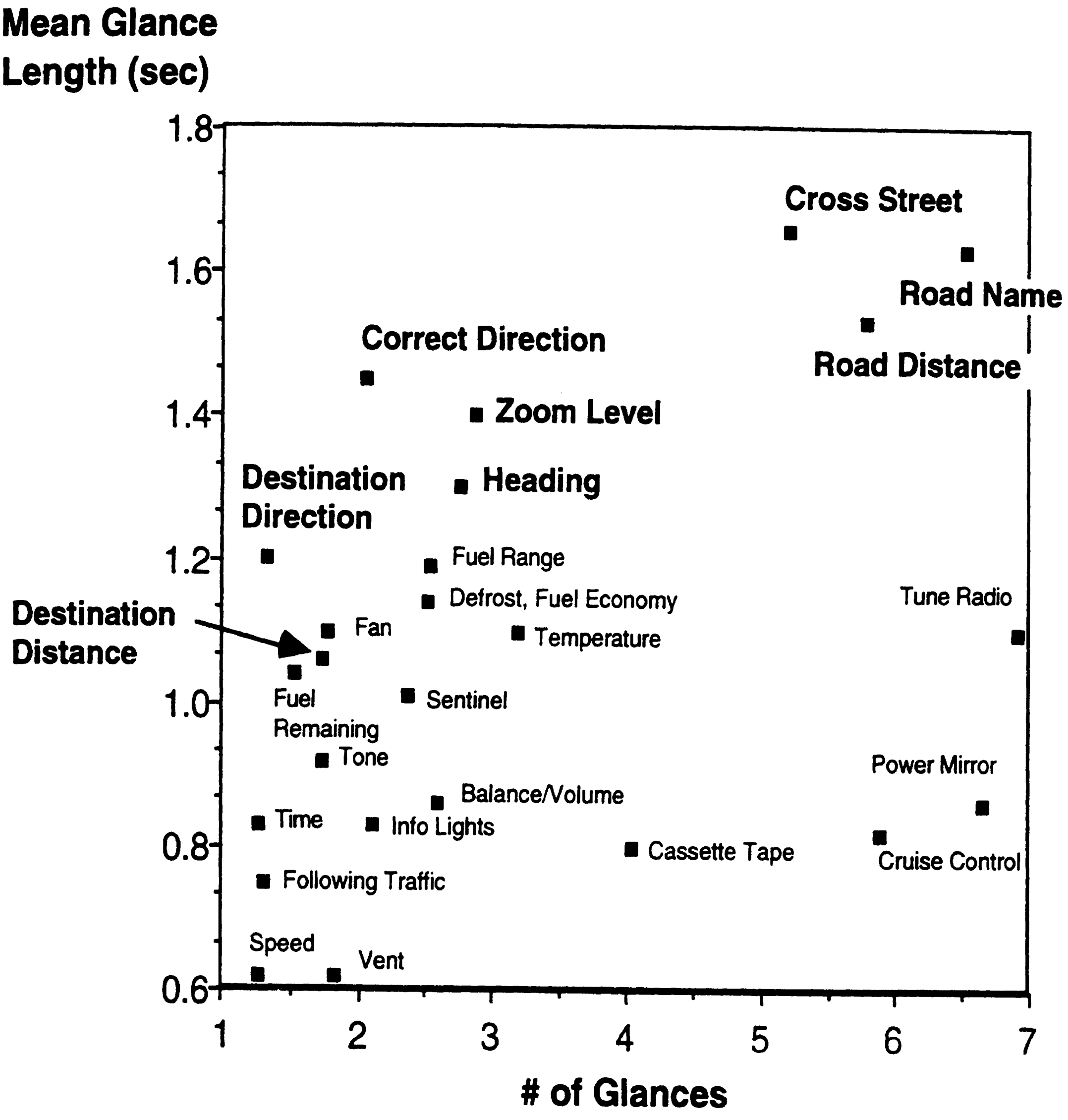
The first experiment, Tom Dingus' dissertation, concerned the attentional demand of the ETAK Navigator (a moving map display).[48] (See also references 49, 50, and 51). While driving, participants were verbally cued to perform certain tasks, such as reading the speedometer, reading the time, adjusting the fanand reading the name of the next cross street. Table 9 shows the total glance times, which equal the sum of the individual glance times. Notice the long times associated with tasks associated with the navigation system. Clearly, looking away from the road scene for an extended period of time increases the risk of driving.

Those data have been replotted in figure 12. It is reasonable to argue that controls and displays should not be added to the vehicle that are worse than those that are provided now. Using that rationale, the mean glance durations should be less than 1.2 s. Further, the author believes that drivers find the number of glances required to manually tune a radio, use of the power mirror, the cruise control, and possibly handling the cassette tape to be unacceptable.

(There is no quantitative data to support this, only his expertise.) If that judgment is accepted, then the number of glances required should be four or less.

Table 9. Total display glance time for each task.[48]

Task	Mean Time (s)	Standard Deviation (s)
under 1.0 s		
Speed	0.78	0.65
Following Traffic	0.98	0.60
1.0 to 2.5 s		
Time	1.04	0.56
Vent	1.13	0.99
Destination Direction	1.57	0.94
Remaining Fuel	1.58	0.95
Tone Controls	1.59	1.03
Info. Lights	1.75	0.93
Destination Distance	1.83	1.09
Fan	1.95	1.29
Balance Volume	2.23	1.50
Sentinel	2.38	1.71
Defrost	2.86	1.59
Fuel Economy	2.87	1.09
Correct Direction	2.96	1.86
2.5 to 4.0 s		
Fuel Range	3.00	1.43
Temperature	3.50	1.73
Cassette Tape	3.23	1.55
Heading	3.58	2.23
4.0 to 8.0 s		
Zoom Level	4.00	2.17
Cruise Control	4.82	3.80
Power Mirror	5.71	2.78
Tune Radio	7.60	3.41
over 8 s		
Cross Street	8.63	4.86
Roadway Distance	8.84	5.20
Roadway Name	10.63	5.80



Note: The labels in bold are navigation system functions.

Figure 12. Glance time data.[48]

Another perspective is that using controls and displays should not cause drivers to drift from the lane in which they are driving. Table 10 shows the lane excursion data from that experiment, a somewhat insensitive measure of safe driving behavior. Drivers should never leave the lane, but it is not clear whether that is an achievable criterion in a practical context.

Table 10. Number of lane exceedences and mean duration.[48]

Task	Total Number of Lane Exceedences (All Drivers)	Mean Duration (s)
Following traffic	0	
Time	0	
Speed	0	
Vent	0	
Destination distance	0	
Destination direction	0	
Turn signal	0	
Fan	1	0.46
Remaining fuel	1	0.95
Tone controls	1	0.97
Correct direction	1	1.00
Sentinel	2	0.28
Balance	2	0.55
Defrost	3	0.67
Heading	3	0.62
Info. Lights	3	0.83
Fuel economy	3	2.25
Zoom level	4	0.94
Fuel range	5	0.84
Temperature	8	0.65
Cross street	8	0.93
Roadway name	8	1.38
Roadway distance	9	1.17
Tune radio	10	1.86
Cassette tape	13	0.99
Power mirror	21	1.10

Also important was Hulse's thesis, which examined how anticipated increases for attentional demand related to the steering task affected use of a navigation display.[52] (See also references 53, 54, 55, 56, and 57). To examine demand, an expression was developed to estimate workload, also referred to as attentional demand (Q) from roadway geometry. Q has a range of 0 to 100. The calculation is shown on the following page. Details concerning the derivations of the workload calculations are not given in the open literature. (For example, why the sight distance factor is a function of the log of the inverse of the sight distance?)

$$Q = 0.4A + 0.3B + 0.2C + 0.1D \tag{5}$$

where:

$$A = 20 \log2(500/Sd) \qquad \text{(Sight Distance Factor)} \tag{6}$$

where Sd = sight distance (m)

if $Sd > 500$, then $A = 0$
if $Sd < 15.6$, then $A=100$

$$B = (100*Rmax) / R \qquad \text{(Curvature Factor)} \tag{7}$$

where R = radius of curvature
$Rmax$ = maximum value of the radius of curvature
(set to 18.52 m (60.7 ft)
the turn radius for a city street)

note: $R = 360X / (2\_a)$

X = arc length along the curve (m)
a = change in direction (degrees)

$$C = -40So + 100 \qquad \text{(Lane Restriction Factor)} \tag{8}$$

where: So = distance of closest obstruction to road (m)
(phone pole, fence, ditch, etc.)
if $So > 2.5$, then $C=0$

$$D = -36.5W + 267 \qquad \text{(Road Width Factor)} \tag{9}$$

where: W = road width for 2 lanes (m)
if $W > 7.3$ (24 ft, 12 ft lanes), then $D = 0$
if $W < 4.57$ (15 ft, 7.5 ft lanes), then $D = 100$

To calibrate these expressions, five graduate students studying human factors engineering drove several road sections twice and then rated them. In the ratings, 1 corresponded to being able to look away from the road for long periods (4 s or more), 5 was for being able to look away for periods of 1 to 1.5 s, and 9 corresponded to not being able to look away at all. Traffic on the test roads varied from light to heavy.

The correlation between the subjective ratings of workload and the calculations of Q was 0.72, with sight distance being the best predictor of the subjective rating. Correlations of those ratings with the percentage of time fixating objects related to driving and the percentage of time fixating the navigation display were low, typically 0.10 to 0.20. This may be because of the detailed analysis (short road segments).

An interesting result of this series of studies is where drivers looked as a function of traffic load and when incidents occurred. ("Incidents" included certain kinds of traffic at intersections, vehicles ahead changing lanes, etc.) In response to increasing traffic, drivers were more likely to look at the road center and for longer periods of time (and less likely to drivers were more likely to look at the road center and for longer periods of time (and less likely to look at the navigation display). The same behavior was observed for incidents. Thus, drivers adapted to the imposed demands placed upon them in a sensible manner.

BOS, GREEN, AND KERST (1988) - LABORATORY ASSESSMENT OF DISPLAYS

In contrast to these field studies are laboratory studies conducted by UMTRI to assess display legibility.[58] (See reference 59 for a review of the literature.) After conducting several pilot experiments to fine tune the test methods, eight young drivers participated in a response time experiment.

The basic task involved showing drivers (seated in a vehicle mockup) slides of instrument clusters in the location they would normally appear. Speedometers varied in their size, contrast, illumination level, and location on the instrument panel. The drivers' task was to find the numeric speedometer and indicate, by pressing a button, if the speed was over the limit [89 km/h (55 mi/h)]. Drivers either responded to speedometers alone or in concert with one of two other tasks (responding to arrows, driving a simulator). In the arrows condition, drivers fixated at a screen well in front of the vehicle on which an arrow might appear. If it did, they pressed one of two keys (left or right). For trials on which an arrow slide was not presented, drivers looked from the far screen to the instrument panel and responded to the panel slide. In the third condition, subjects drove a simulator and, at random times, responded to slides that appeared on the instrument panel. The pattern of results from the driving and arrows conditions were quite similar. Response times when there was no additional task tended to significantly underpredict response time in other task combinations when the viewing conditions were poor (e.g., low contrast). Further, the variability in the response times to speedometer slides within designs in the arrows condition was less than in the driving condition, so the more sensitive arrows condition was used for subsequent studies of instrumentation.[60]

Thus, what emerged from this research was a laboratory method for testing display legibility. This research highlighted the importance of driver accommodation to the road scene as a factor that influenced the time to read instrumentation. Thus, if instrument panel display legibility is to be assessed, an external task must be provided. In addition, visual search for the relevant display was an important factor. All of the speedometers tested met minimum legibility standards. However, increasing the size of speedometer digits to several times the minimum led to large reductions in response time, since these increases made it easy for drivers to find the speedometer.

NOY'S (1987) REVIEW OF SECONDARY TASK METHODS

This report is an exhaustive review of secondary-task methods as they apply to evaluating intelligent automotive display systems.[61] The reasoning behind the secondary (or subsidiary) task method is that carrying out a task or collection of tasks requires some fraction of one's attentional capacity. To assess how much is required, one is asked to complete another task (of less importance) in concert with the existing or primary task. Ideally, the secondary task should not affect performance of the primary task. The score on the secondary task is indicative of the amount of processing resources available to carry out additional work.

Shown in table 11 is an abridged version of Noy's tabular summary of the literature. In creating this table, the original sources cited by Noy, which have additional details, were not retrieved. Studies were primarily concerned with driver workload and driver fatigue. Some of the experiments did not employ tracking or steering-like activity as a primary task. These data do not suggest a single ideal secondary task. Popular tasks included visual detection, response time (RT) to tones, response time to lights on the instrument panel (IP), various short-term memory tasks (such as digit shadowing), cognitive tasks involving mental arithmetic, and, in one case, antonym naming.

Table 11. Secondary task list.[61]

Study	Secondary Task	Issue	Effect on Driving	Effect on Secondary Task
Boadle, 1976	detect light on IP	vigilance	interference	unknown
Brown & Poulton, 1961	auditory digit shadowing	general methodology	unaffected	sensitive to traffic density
	mental addition		unaffected	sensitive to traffic density
Brown, 1965	auditory digit monitoring (odd-even sequences)	nature of secondary task	intrusive	more sensitive than memory task
	detect repeated letter in series		intrusive	
Brown, et al., 1967	voice interval production task	compare 2 tasks	slight intrusion	not stated
	visual detection		not stated	performance improved with time
Brown, et al., 1969	auditory grammatical transforms	interference with telephone use	little effect on control skills, affected gap judgment	interference with perception, not motor skills
Dobbins, 1963	visual detection	vigilance	no effect	no effect
Drory, 1985	visual choice RT	vigilance and rest periods	unclear	unclear
	read speedometer		improved performance	unclear
Fagerstron & Lisper, 1977	RT to tone	vigilance	not stated	affected by car radio listening, experience, etc.
Harms, 1986	mental subtraction of auditory digits	general methodology	RT inversely correlated with speed	RT increased in high accident areas

Table 11. Secondary task list (continued).

Study	Secondary Task	Issue	Effect on Driving	Effect on Secondary Task
Harms, 1986	mental subtraction of auditory digits	general methodology	RT inversely correlated with speed	RT increased in high accident areas
Heimstra, 1970	visual detection RT	effect of stress from electric shocks	not stated	not stated
	detect meter deflections		not stated	not stated
	detect brightness change		not stated	not stated
Hicks and Wierwille, 1979	read random numbers	comparison of workload techniques	no intrusion	not sensitive
Hoffman & Joubert, 1966	auditory digit shadowing	effects of vehicle dynamics	not stated	correlated with number of cones hit
Johnston and Cole, 1976	visual detection	effect of distractions (advertising)	not examined	affected by signs
Laurell & Lisper, 1978	auditory RT	general methodology	not stated	RT correlated with obstacle detection
Lisper, et al., 1986	auditory RT	fatigue	not stated	RT increased over time
Lisper, et al., 1979	auditory RT	diurnal variation	not stated	slight change with time of day
Lisper, et al., 1973	auditory RT	physiological indicators	not stated	RT better than physiological measures
McDonald & Ellis, 1975	visual digit shadowing	attentional demands-curves & speed	not stated	unclear
Quenalt, 1977	mental addition (auditory presentation)	careless vs. normal drivers	not examined	road conditions affected performance
	antonym naming			road conditions affected performance
	count tones			road conditions affected performance

Table 11. Secondary task list (continued).

Study	Secondary Task	Issue	Effect on Driving	Effect on Secondary Task
Riemersma, et al., 1977	report changes every 20 km			sensitive to fatigue
	detect colored light change	fatigue	not stated	sensitive to fatigue
Sanders & Noble, 1975	count targets	general methodology	interference	affected by # of targets
Snyder & Monty, 1986	operate radio, trip computer, climate controls	effect of programmable displays	degraded lane keeping and speed control	not stated
Stephens and Michaels, 1964	identify target word on sign	interference between search and tracking	degraded by secondary task	not stated
Wetherell, 1981	mental addition	compare 6 secondary tasks	secondary tasks intruded for women only	no single task outstanding
	grammatical transforms			
	auditory detection of digit in stream			
	recall of route instructions			
	generate random digits			
	Sternberg auditory memory task			
Wierwille et al., 1977	read random digits	vehicle handling parameters	not stated	sensitive to some steering parameters
Wierwille & Guttman, 1978	read random digits	general methodology	intrusion	less sensitive than primary task performance
Zeitlin & Finkelman, 1975	random digit generation	general methodology	no interference	sensitive to task difficulty
	digit shadowing		no interference	not sensitive to task difficulty
Zwahlen, 1986	reading text	reading text and occlusion	interference	reading text and occlusion degraded performance equally

ALTERNATIVE METHODS FOR ASSESSING WORKLOAD WHEN FOLLOWING A VEHICLE

Noy's recent work on methods for assessing driving workload was summarized in a paper and a technical report.[62, 63] This very comprehensive effort builds upon the literature review just described. Experiments were carried out using a moving-base DC-8 flight simulator at the University of Toronto. The rear of the cab was configured to represent a compact car. Near the center console was a 30-cm (12-in) monitor for presenting an auxiliary task. Drivers wore head tracking transducers and Electro-oculogram (EOG) electrodes to record head movements and eye fixations.

Drivers looked at a 30 by 40 degree scene with low detail (300 polygons) that updated at 20 to 30 Hz. A two-lane winding road with a dashed centerline was shown. Their task was to maintain a constant headway from a lead truck and minimize lateral drift. Periodically the truck decelerated rapidly, in response to which the driver stopped. At various times short vertical lines (6.5 mm) appeared on the in-vehicle, auxiliary display in the presence of 8 mm vertical distracters. The task difficulty was varied by altering the number of distracters (1, 4, 8, or 12). The driver's task was to identify whether a shorter line was present. At other times drivers were presented with a Sternberg memory task for set sizes of 2, 3, 4, and 5 letters with 3 probes. For a set of four, the following would occur. "W, T, F, R", ... followed by a delay... "Yes or no, were any of the following letters presented, A, F, E?" For both auxiliary tasks, response time was the dependent measure and the yes/no probabilities were equal.

In addition, there were six dependent measures related to driving performance (standard deviation of lane position, lane exceedence ratio, time-to-line crossing, headway, speed, and standard deviation of speed), three related to attentional demand (dwell time, look frequency, auxiliary display to road scene viewing ratio), and seven (time load index, mental demand, physical demand, temporal demand, effort, performance, and frustration) associated with TLX, the NASA Task Load Index, a subjective measure of workload.

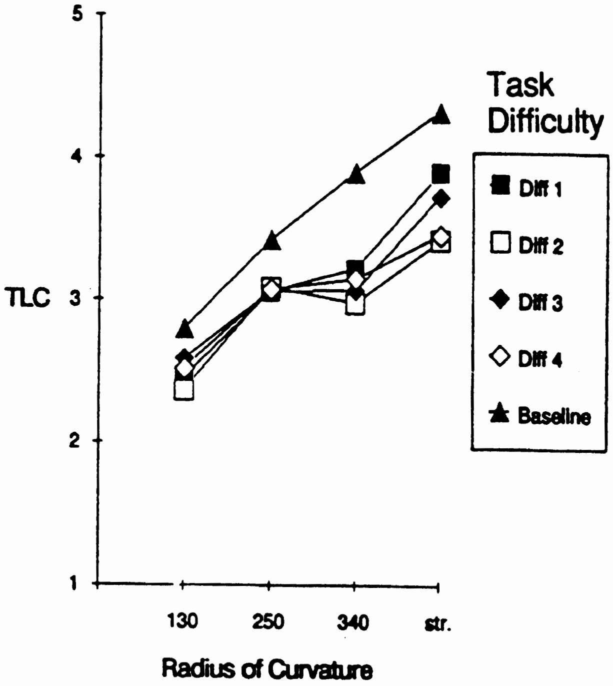
Twenty students served as subjects. They were given 4 h of training in single and dual task conditions prior to the experiment. The main experiment consisted of two 2-h sessions separated by a 30-min break.

Analysis of the data showed there were no significant differences in driving performance between the perception (line discrimination) and the memory tasks, so further analysis focused on the perception task. The presence of the auxiliary task significantly affected time-to-line crossing, standard deviation of lane position, headway, and standard deviation of speed. Table 12 shows task decrements ranked from left to right. Except for headway, the level of task loading had no effect on driving performance.

Table 12. Auxiliary task decrements in dual task conditions.[51]

Condition	Std. Dev. of Lane Position (m)	Std. Dev. of Speed (m/s)	Time-to-Line Crossing (s)	Headway (m)	Lane Exceedence (%)
Driving	0.2	0.79	3.47	53.5	0.0
Dual Task	0.25	0.94	2.90	56.7	0.02
% Change	+26.	+19.	-16.	+6.	+2.

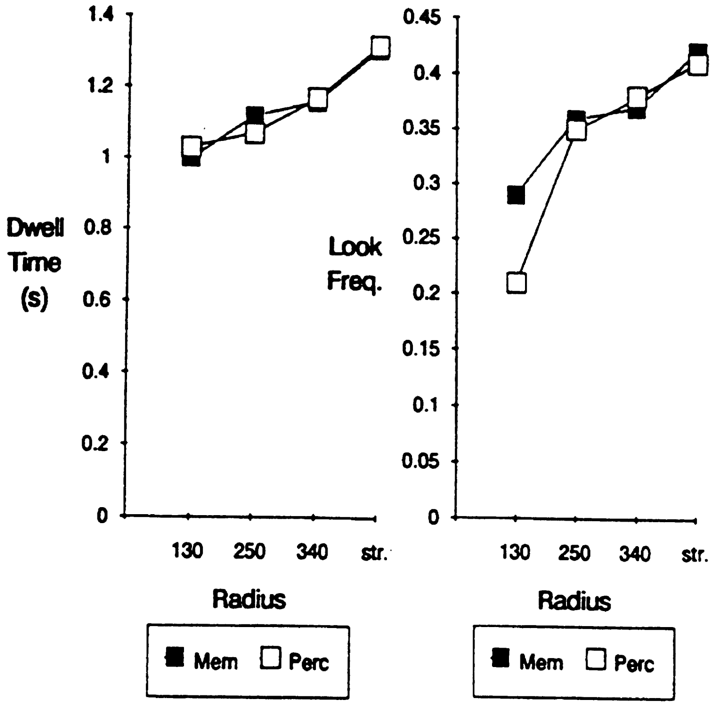
Of the driving performance measures, only time-to-line crossing and standard deviation of lane position were significantly affected by the difficulty of the driving task. Figure 13 shows the effect of road curvature on time-to-line crossing.



Note: Difficulty refers to the secondary perception task present.
Baseline refers to the condition where a secondary task was not present.
Str. represents the straight road condition (no curvature).

Figure 13. Time-to-line crossing (s) as a function of road curvature radius (m).[51]

With regard to the visual/attentional measures, both dwell time and look frequency were affected by the curve radius. (See figure 14.) In response to greater external demands, drivers looked at the auxiliary display less often with the viewing ratio display (the percentage of run time spent looking at the in-vehicle display) changing from 50 percent, in the low driving load conditions, to 20 percent, in the high load conditions.



Note: Mem and Perc refer to the memory and perceptual secondary tasks.

Figure 14. Visual/attentional performance as a function of road curvature.[51]

One of the more interesting results concerns partitioning the driving performance measures based on where the driver was looking at any moment. In fact, standard deviation of lane position, standard deviation of speed, and exceedence ratio were all larger, and time-to-line crossing was smaller, when the driver was looking inside rather than outside. That is, drivers tended to look at the inside display only when it was relatively safe to do so.

Results concerning TLX appear only in the technical report. Only two of the rating scales (mental demand and physical demand), and to a lesser degree, the weighted composite TLX rating, were significantly linked with driving task difficulty.

Thus, these data show that driving task difficulty (here curve radius) had a larger effect on performance and attention than auxiliary task difficulty. According to Noy, the visual attentional variables were more sensitive to experimental manipulations than primary driving variables. Noy believes that drivers "were able to maintain primary task performance within the desired bounds by judiciously modulating their scanning behavior in response to changing task requirements."[53] Of the driving performance measures, standard deviation of lane position, standard deviation of speed, and time-to-line crossing were most noticeably affected by the addition of a secondary task and should be examined in future driver performance studies. In some cases, time-to-line crossing may be difficult to obtain.

With regard to the auxiliary tasks, the presence of any of them significantly degraded driving task performance and, to the extent they mimic future in-vehicle displays, that is a matter of concern. Potentially of greater concern is not their effect on the mean performance level, but their effect on increased performance variance. Also important, however, is the extent to which drivers adapt to the addition of secodary tasks, and the strategy they choose to execute these tasks (for example, only when driving performance is relatively good).

INSTRUMENT PANEL EVALUATION WITH SIMULTANEOUS TASKS

This paper describes an approach that General Motors has used to evaluate instrument panel controls.[64] Detailed performance data are not provided. Individual drivers seated in a mockup carried out three tasks concurrently. In the speed regulation task, drivers monitor an analog speedometer perturbed by a random signal. Drivers maintain a constant speed using the accelerator.

In the pedestrian detection task, drivers are shown a video scene of a single lane road lined with pedestrians. (See figure 15.) At random times, a pedestrian enters the road for 50 ms. The drivers task is to indicate if that occurs from the left or right.

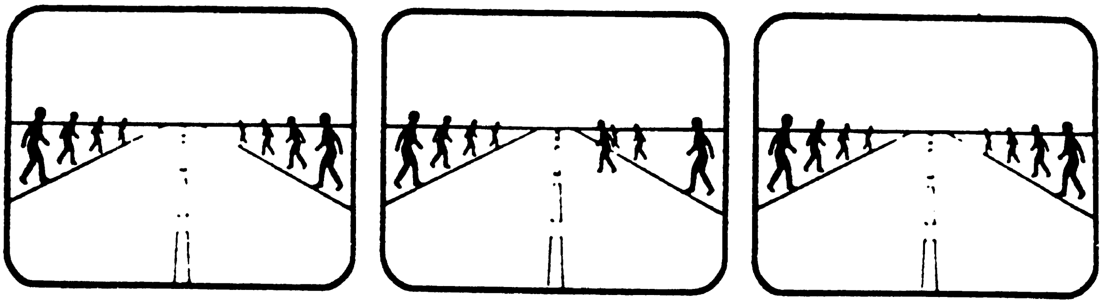


Figure 15. Three-frame sequence of the pedestrian detection task.[64]

The third task performed involves operation of instrument panel controls. Auditory commands to operate controls (wiper, lights, radio, etc.) are given, and driver behavior is videotaped.

In a typical experiment drivers are given practice in the speed regulation and pedestrian detection tasks both independently and together. Following is practice in the instrument panel task until a preset criterion is reached. The test conditions involve several 5-min blocks in which all tasks are performed concurrently. Dependent measures include the mean speed, the number of pedestrians detected, the task completion times for each control, hands-off-the-wheel time, and the frequency and types of errors associated with each control.

Hardee et al. describe two methods for analyzing the data. In the video method, tapes of the experiment are played back frame-by-frame for analysis, so times are accurate to the nearest

1/30 of a second. Experiments of this type tend to be easy to set up and counterbalance. Analysis, however, can be very time consuming. In the automated method the subject wears a conductive wrist strap interfaced to a computer, so the time his hand leaves the steering wheel can be determined. In addition, all of the controls must be interfaced to a computer, which takes time to accomplish. Also, in this method, a quad splitter is used to show a time-synchronized view of the pedestrian detection scene, the subject's eye movements, and his interaction with the controls. Based on discussions with GM personnel, both of these approaches are regularly used to analyze control designs.

CONGESTION AVOIDANCE USING CRT DISPLAYS

This method simulates driver behavior while trying to avoid nonrecurring congestion by diverting with the aid of an in-vehicle traffic information/navigation system.[65] Drivers were shown slides of road scenes (Golden State Freeway in Orange County, California), which included an instrument panel image showing speed and time. At the same time, they were shown a simulated traffic information display on a CRT. Computer generated auditory feedback of engine sounds, wind, and road noise was also provided. Across scenarios, the speed shown and the level of traffic congestion varied. The manner in which the slides were produced was quite clever. Scenes of nearly vacant freeways were shot. Separately, pictures of vehicles were shot and then superimposed in the scene in the proper location, so the scenes could be photographed again. This resulted in a series of scenes that were identical except for the number of vehicles shown. Producing these slides was painstaking work.[66]

Four different in-vehicle units were examined. They included a static map system similar to ETAK (no congestion or guidance information), a static map with congestion level information, a dynamic map with a highlighted alternative route and auditory messages on traffic, and an arrow-based route guidance system. Further details on the design of the interfaces is contained in reference 66.

Each participant utilized only one of the interface designs, but for three different delay levels. All subjects were given training in the operation of the navigation system. In test trials, slides were shown at an unknown rate. When subjects wanted to divert, in most cases they keyed in the numbers of the nodes (shown on a special map) through which their vehicles would pass.

There were significant differences between display types in terms of how early in a trip drivers would divert. Drivers with route guidance interfaces tended to divert first, followed by those with the dynamic map, followed by others.

What is interesting about this method is that it provided useful behavioral information that could serve as input to a traffic simulation model. It did so in a way that provided control over the traffic congestion level and had considerable fidelity with real driver decisions. Developing the task scenarios was nontrivial, especially the photographic work.

FACTORS CONTRIBUTING TO DRIVER WORKLOAD

This experiment examined the factors that contribute to driving workload.[67] The results could be applied to design a dialogue controller for the driver-vehicle interface that avoids overload.

The driving task involved a 3.5-km (2.2-mi) four-lane motorway and 1.4 km (0.9 mi) of rural road in the Netherlands. Participants drove at normal speeds. For the 10-min drive, six segments were analyzed. (See table 13.) Segments were 120 to 800 m long. Trips began at specific times in the morning and afternoon, corresponding to times of low, moderate, and high density traffic.

Table 13. Tasks examined.[67]

Road	Task
4-lane motorway (100 km/h)	merge and exit
4-lane motorway (100 km/h)	drive in right lane of straight section
2-lane road (100 km/h)	drive around traffic circle
rural road (80 km/h)	turn right a cross street
rural road (80 km/h)	turn left just before a curve
rural road (80 km/h)	drive on straight section

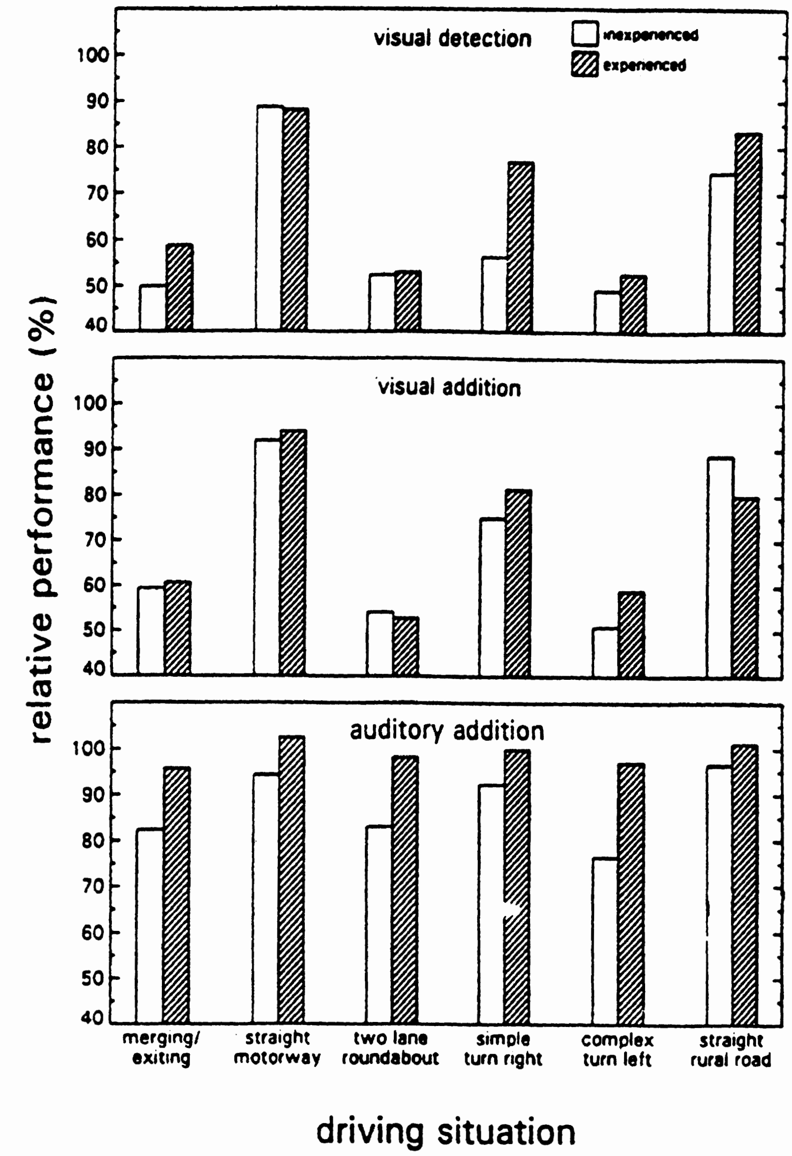
Driving took place under four test conditions: 1 no-task (control) condition and three secondary task conditions. Those secondary tasks consisted of a visual detection task, a visual addition task, and an auditory addition task. In the visual detection task, subjects said the Dutch equivalent of yes when a display mounted high on the center console was illuminated. This occurred for 0.75 s every 2 to 4 s. In the visual addition task, subjects added 12 to the number shown on the console display and spoke the answer. In the auditory addition task, the number was presented auditorally (for 1 to 1.5 s) instead of visually. Baseline secondary task data were obtained for two 2-min sessions with the car stopped.

The test vehicle was a Volvo 240 station wagon with a manual transmission. A roof-mounted camera recorded the forward scene. A dashboard-mounted camera recorded driver eye movements. Foot control and steering wheel inputs, speed, and lane position were sampled at 4 Hz.

Serving as subjects were 24 drivers familiar with the road course. Half had been licensed for less than a year; the others were experienced drivers.

Measures examined in this experiment included reduction of secondary task scores due to driving (the difference between the baseline and driving conditions), eye glance measures (glance frequencies and durations to four locations), and nine driving performance measures. Of the nine driving measures, three were collected for all segments of the test route: speed, standard deviation of speed, and steering wheel action rate (SAR). SAR is the number of wheel movements per s. According to Verwey the movement threshold was 5 degrees/s. The other six measures were specific to maneuvers such as merging or right turns: time to merge, distance required, speed after merging, and time, distance and speed of braking before intersection. Eye fixations were reduced manually in real time by an experimenter pressing a button to indicate the object fixated (right, left, interior mirror, display).

Approximately 2600 responses were obtained for each of the three secondary tasks. Error rates were 34 percent (visual detection), 40 percent (visual addition), and 22 percent (auditory addition). Figure 16 shows the relative performance on each of the three secondary tasks for various road segments for both inexperienced and experienced drivers. There were significant differences between task types, with performance for each task type depending upon the driving situation, but not the traffic density.



Note: While not fully explained in the paper, it appears that relative performance refers to the error rate relative to that of auditory addition secondary task performed while driving on a straight motorway.

Figure 16. Secondary task performance for various tasks.[67]

The primary differences in driving demands seem to be related to visual input, with both the visual detection and visual addition secondary tasks reflecting large differences between driving situations. Further, statistical analysis revealed a significant interaction of secondary tasks with driver experience, with inexperienced drivers showing much greater degradations for some task-driving situation combinations.

These results indicate that for experienced drivers, the primary limitation is visual load, not cognitive load. The visual tasks vary most widely with the driving situation whereas the auditory task did not vary very much. However, for the inexperienced drivers, cognitive loading is a somewhat of a problem, as well, since all tasks (including auditory addition) vary with the driving situation. Verwey suggests that the absence of interference with auditory addition for experienced drivers may be because the basic driving skills are so highly automated.

With regard to eye movements, there were differences due to driving situations causing drivers to glance in different locations. For example, drivers looked more to the left for left turns and more to the right for right turns. Quite noteworthy, were differences in the number of glances to the mirror and display as a function of the driving situation. (See figures 17 and 18.) When mirror usage was not central to driving (e.g., tasks other than merging), the number of mirror glances was sensitive to workload, with more mirror glances occurring for easier driving situations. Verwey suggests that mirror usage is considered as less important than other visual tasks and is one of the first tasks shed when workload increases.

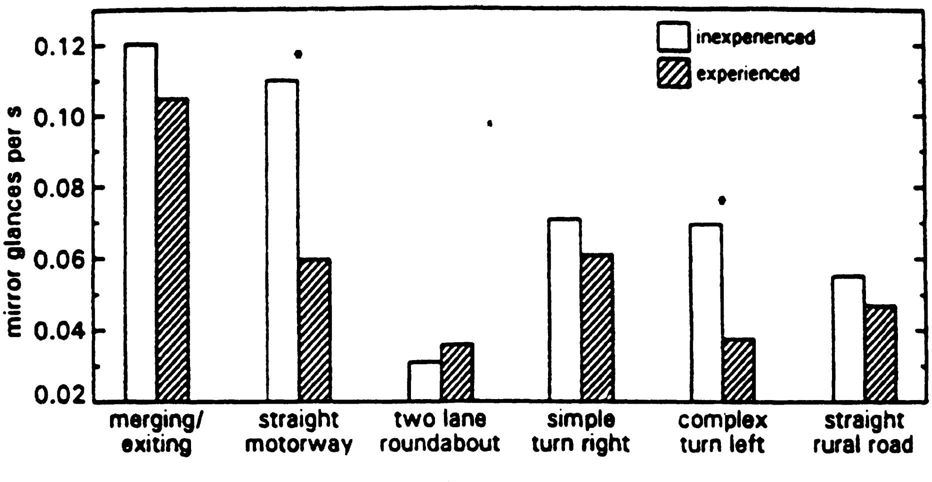


Figure 17. Number of mirror glances for various tasks.[67]

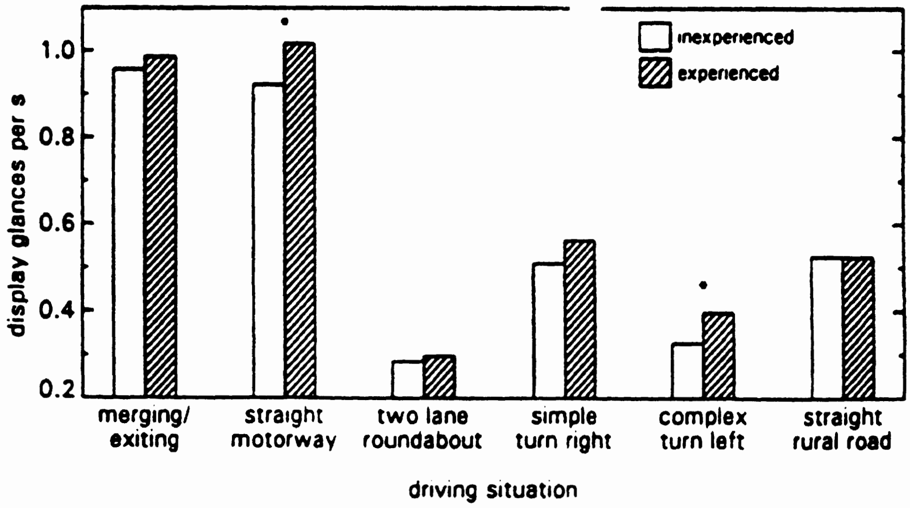


Figure 18. Number of display glances for various tasks.[67]

Patterns for glance durations were similar to those for glance frequency, though they were not examined in the same detail (in part because there were fewer significant differences). In brief, the easier the driving task, the longer the average fixation.

Many of the driving performance measures were affected by the driving task (the frequency of accelerator, brake and clutch depressions, driving speed, the standard deviation of speed, etc.). None of these performance variables was affected by the secondary task, though brake pedal frequency increased and mean speed decreased with traffic density. Verwey notes that the lack of secondary task effects supports a multiple resource theory of driving (with separate resources for visual, motor, and cognitive processing) for experienced drivers. Verwey also claims that single resource theory best supports the inexperienced drivers' data (because differences in driving situations influenced performance in all secondary tasks). According to Levison, an alternative explanation is that "multiple resource theory applies to both classes of drivers, but that the cognitive resource must be shared among all tasks—in this case, driving and the auditory side task.[68] (This assumption is made by the Integrated Driver Model.[19]) Assume that the driving task imposes a lower demand on cognitive resources for experienced drivers than inexperienced drivers because of the increase in information processing that accompanies learning. Because the experienced drivers have more "spare capacity" than the inexperienced drivers, the experienced drivers can devote the necessary cognitive resources to the audition task, whereas the inexperienced drivers come up short. Thus, one might claim that differences between driver classes are due to the efficiency of information processing associated with the driving task, and not to differences in the number of independent resources." Hence, the multiple resource theory of driving could explain the performance of both inexperienced and experienced drivers.

Steering action rate was affected by several factors and was apparently sensitive to workload. Experienced drivers moved the steering wheel less often than inexperienced drivers (0.42 vs. 0.46 times/s). The rate also varied with the secondary task, especially for tasks with visual demands. (See figure 19.)

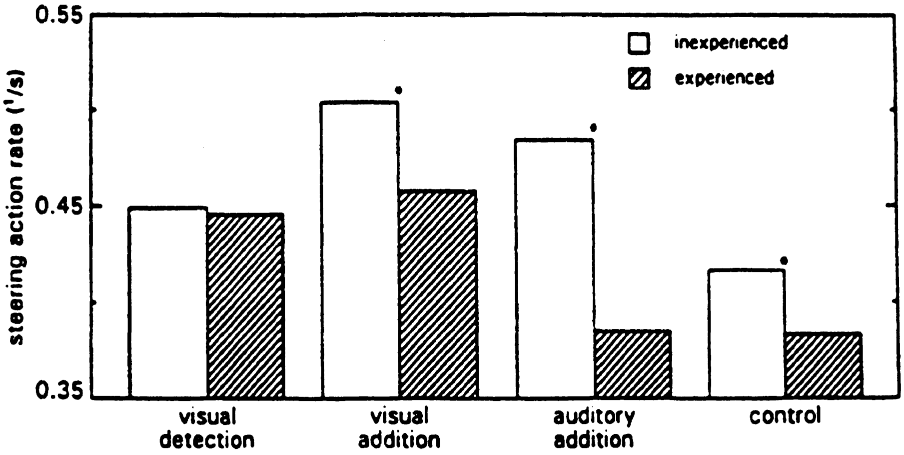


Figure 19. SAR and secondary task.[67]

The following conclusions emerged from this research.

1. Verwey claims the behavior of experienced drivers may be best explained by multiple resource theory, while inexperienced drivers fit single channel theory. Levison claims multiple resource theory explains the performance of both classes of drivers.

2. Driving workload was primarily determined by the driving scenario, not the traffic density.

3. The primary limitation of driving is visual, and hence, tasks that contain a purely visual load (e.g., peripheral detection) are most likely to be sensitive to driving demands. That is, limitations are more perceptual than cognitive in nature.

4. Steering action rate may be a useful online measure of driving workload.

5. Glance frequency appears to be more sensitive to workload than glance duration.

6. When mirror use is discretionary, it seems to be sensitive to driving workload.

TRAVTEK PREDRIVE TASK PERFORMANCE

This paper concerned the evaluation of predrive functions associated with the TravTek interface, a turn arrow and voice-based navigation system recently examined in an operational field test in Orlando, Florida.[69,70,71] During the development of the interface, subjects were given a wide variety of tasks to complete. The TravTek interface was shown on a touch screen CRT connected to a Macintosh computer. The graphics were generated in SuperCard. In addition, a video camera was placed behind the subject to record the general nature of the interaction.

A total of 72 people participated in this experiment, equally distributed among three age groups and both sexes. There were seven basic tasks—select an unfamiliar address, select a previously stored destination, determine street names where congestion is present, store a destination and route for future reference, use the yellow pages to select a business, set voice messaging options, and summon emergency service. Tasks were reasonably complex, involving as many as 7 separate screens and 5 to 20 button presses. The experimenter read a description of the task while subjects saw the main menu. The computer recorded each touch screen button press. At the end of the experiment, subjects were given a survey and rated the difficulty of each task.

Table 14 shows the task completion times. In general, these times were correlated with the number of errors, but a correlation coefficient is not given. Most of the task completion times were associated with recovery from errors, not with choosing an inefficient access method.

Table 14. Navigation task completion times.[69]

Task	Time (s)
select an unfamiliar address	130
select a previously stored destination	50
determine area street names where traffic congestion is present	240
store a destination and route for future reference	160
use a yellow pages feature to select a business	95
set voice messaging options to a desired destination	40
summon emergency service	40

The importance of this study is that it demonstrates a laboratory method for assessing the usability of driver interfaces, showing the utility of both performance and subjective data. This research was made possible by recent advances in rapid prototyping.[72]

TRAVTEK OPERATIONAL TEST

This research is quite different from those mentioned previously. Rather than evaluating a system or alternative systems by a few drivers in the laboratory or on the road, this project involved approximately 100 cars, potentially thousands of drivers, and a real, functioning traffic information system. The only other operational test in the United States, Pathfinder, conducted in Los Angeles, was far less ambitious than the TravTek project.[73] A detailed overview of the TravTek test protocol is given in in references 74 and 75. For an overview of the evaluation of other related projects, see references 76 and 77.

In brief, 10 experimental approaches were examined—(1) a field study with rental users, (2) a field study with local users, (3) a yoked driving study, (4) the Orlando test network study, (5) the camera car study, (6) a survey of rental and local users, (7) debriefing and interviews in several studies, (8) traffic probe studies, (9) modeling and analyses, and (10) a global evaluation. The first five approaches emi/hasize human performance and behavior.[75]

The rental car users study involved people who rent TravTek cars from Avis and a matched control group. Within the TravTek group, there were three subgroups (all functions, navigation and service function but no real-time data, service functions only). All drivers participated in the questionnaire study, and a subset were interviewed. Data concerning vehicle location, heading, speed, and stops were automatically time stamped and logged, as were all keypresses associated with the TravTek interface.

The local users study was similar to the rental car study, except that people familiar with Orlando had the test vehicles for several months. The dependent variables were the same.

In the yoked driving study, hired drivers were assigned to the three versions of the TravTek interface. Pairs of drivers (one without a Travtek interface, one with a TravTek interface) went from specified origins to specified destinations at the same time, so weather, congestion, etc. were matched. The focus of this approach was on the time and distance savings associated with real-time traffic information.

The Orlando Test Network Study was similar to the yoked experiment, except that there were a network of routes between origin-destination pairs. Hired subjects drove vehicles with either a hard copy map, a route guidance system, a moving map with a route guidance system, a moving information map, or voice guidance. Trip times and experimenter ratings were the primary dependent measures.

The Camera Car Study provided a detailed analysis of driver performance. Video cameras were focused on the road scene, the driver, and the outside lane line. Dependent variables included eye glance measures, speed variance, and lane excursions. Supplemental information was provided from an accompanying experimenter's log. Drivers performed predrive functions, baseline tasks (e.g., use cellular phone) and drove the Orlando test network. Each driver participated in four conditions — hard copy map, moving map with route overlay, route guidance, and voice guidance.

This operational test was extremely thorough and yielded valuable data concerning the relationship between human performance measures (e.g., glance times), behavior measures (e.g., route choices), and safety (e.g., accidents).

CLOSING COMMENT ON MEASURES AND METHODS

As a whole, these studies show that the research community is far from reaching a consensus either on the research protocol to be used or on the appropriate measures. Particularly lacking are any attempts to replicate research results. In other areas of science, "truth" is established when many independent investigators examine a question using similar methods and reach the same conclusion. Such work has not been carried out in driving science, in part because resources are so scarce that replication has been avoided.

TABULAR SUMMARIES OF METHODS AND MEASURES

Now that the reader has a sense of the methods and techniques used, it is appropriate to look for trends. In the following tables are all the studies, to the author's knowledge, that examine the use of navigation systems, as well as more general studies of timesharing. While relevant, studies of the use of cellular phones have not been included. Those studies appear in other reports funded by this project .[11] The reports listed in the following tables have been partitioned into three categories: methodological experiments (table 15), general studies seeking information regarding interface design (table 16), and specific interface experiments (table 17). Many of the experiments examined fit into several categories; however, each experiment was placed into only one category, for simplicity.

Table 15 concerns methodological studies. Included here are some basic studies concerning timesharing, experiments that concern the sensitivity of various human performance characteristics, and related issues. To keep the scope reasonable, only studies that included driving-like tasks as one of the timesharing activities have been included. Had the scope been expanded to include general tracking studies, the list would have been enormous. Just as rare were operational tests, in part because of the millions of dollars they often cost to execute.

Of the 23 items listed in table 15, 10 were conducted on the road, 5 were conducted in a driving simulator, 2 were conducted in both contexts, 4 were conducted in part task simulators, and only 1 was a true laboratory experiment. In some sense, the value for the on-road category can be misleading as 4 of those studies (conducted by Zwahlen) took place on an abandoned airport runway. None of these experiments were conducted on test tracks, most likely because of the cost of access. Somewhat unexpected to the author, was the small number of laboratory studies. This may have been due to the selection criteria, not the absence of useful material in the literature.

There was no consistent pattern for the choice of independent or dependent measures, or in the results. However, it was clear that for the simulator and on-road experiment, the number of dependent measures collected was large with a half dozen being typical. Dependent measures included response times and error rates for in-vehicle tasks, heart rate variability, eye fixations frequencies and durations, tracking-time-off-target, steering wheel angle statistics, time to line crossing, mean speed, lateral deviation, ratings of attentional demand, and the number and severity of unsafe driving actions.

Table 15. Methodological studies.

Experiment	Domain	Dependant Variable	Independent Variable	Task	Results/Comments
Bos, Green, and Kerst (1988)[58]	lab	resonse time, error rate	digit size, task condition	either look at IP and press 1 or 2 buttons (speeding/not speeding), look a screen for left/right arrow, if not arrow press button, or RT + driving simulator	RT + arrows gave best data
Bouis, Voss, Geiser, and Haller (1979)[78]	part task driving simulator	response time for secondary task, RT to unexpected signal	display format-visual, auditory, combined	track moving target and hold varying speed while pressing buttons in reponse to lights	shorter RT's for combined displays, shorter RT's for lights and tones that are continuous (vs. intermittent)
Brouwer, Ickenroth, Van Wolffelaar, and Ponds (1990)[79]	in part task simulator	similar to Ponds, Brouwer, and van Wolffelaar (1988)	driver age	don't have yet	don't have yet
Daimon (1992)[80]	on road	heart rate variability, thinking aloud comments, eye fixation durations and frequencies	map or nav system	drive route using nav system or map	lower peak power (in heart rate variability frequency spectrum) with map, fewer eye fixations to map
Daimon (1992)[80]	simulator	tone identification time (secondary task-0, 1, or 2 tones lagged), time off target in tracking, eye fixation duration and frequency	secondary task	drive and respond to secondary tasks and navigation display	longer fixation times but fewer fixations with paper map, lag 2 response time was most likely to show map-nav system difference

Table 15. Methodological studies (continued).

Experiment	Domain	Dependant Variable	Independent Variable	Task	Results/Comments
Godthelp (1988)[41]	on road	Steering wheel angle, yaw rate, and lateral position, TLC	speed	when tone presented, stop steering then resume at latest moment	(TLCs=1.3 s, TLCmin=1.1 s, constant was the minimum lateral distance to the lane boundary [about 15 cm (5.91 in)]
Godthelp, Milgram, & Blaauw (1984)[40]	on unused highway	Steering wheel angle, yaw rate, and lateral position, TLC	speed	drive and press button for 0.55 s look (occlusion)	time-to-line-crossing at the end (TLCe) of each occlusion period was about 1.57 times the occlusion duration, TLC was measure of safety
Grant and Wierwille (1992)[81]	on road and in lab	task duration	timing method (on-road, real-time, slo-mo), task duration (short, med, long)	watch real driver carry out reach for controls and time with stopwatch, watch tape in real time in lab, watch tape at 1/6 speed	slow motion led to larger time estimates, real time led to smaller time estimates, on-road led to no biases
Hardee, Johnston, Kuiper, and Thomas (1990)[64]	part task simulator	mean speed, the number of pedestrians detected, the task completion times for each control, and the frequency and types of errors associated with each control	control of interest	maintain speed, detect pedestrians on road, use control	none given

Table 15. Methodological studies (continued).

Experiment	Domain	Dependant Variable	Independent Variable	Task	Results/Comments
Hicks and Wierwille (1979)[39]	VPI driving simulator	lateral deviation, yaw angle deviation, and two-degree steering wheel reversals/unit time, rated attentional demand (extremely low to extremely high)	vary crosswinds (workload)	driver either while reading random digits (secondary task), with occlusion (200 ms looks), or while heart rate was recorded	lateral dev, yaw and reversals all affected by workload, reversals was most affected, ratings affected by workload, occlusion, secondary task performance, heart rate not affected
Korukawa and Wierwille (1990)[82]	in simulator and on road	task completion time, hand-off wheel time, duration and # of glances to road and IP	task (17-press button on radio, turn knob, etc.)	steer car or simulator and reach for control on command	simulator times were close to in-car times, crosswind reduced # glances to IP and increased to road
MacAdam (1992)[83]	driving simulator	lateral deviation, heading angle, standard deviation of wheel angle, yaw rate, lateral acceleration, mean time on side tasks	side task, choice of dependent measure	steer on straight road and wind side gusts while performing second task (RT to single letter, 2-choice RT with letters, 2 digit addition)	standard deviation of lateral position and heading (yaw) angle were most sensitive to side tasks, other measures were insensitive
Noy (1989, 1990)[63]	simulator	standard deviation of lane position, lane exceedence ratio, time to line crossing, headway, speed, and standard deviation of speed, + secondary task (various RT's), + TLX, dwell time and look frequency	secondary task present, driving task difficulty	while driving either detection of lines (perceptual task), Sternberg memory task, or driving alone	perceptual and memory task had similar effects on driving, TLC and lane variance affected by driving task difficulty, dwell time and look frequency affected by driving load, TLX linked to task difficulty, drivers executed secondary tasks when driving was good

Table 15. Methodological studies (continued).

Experiment	Domain	Dependant Variable	Independent Variable	Task	Results/Comments
Ponds, Brouwer, and van Wolffelaar (1988)[84]	in part task simulator	tracking time-on-target, % correct in dot counting	driver age (21 to 37, 40 to 58, 61 to 80)	steer on straight road to counteract side winds, count # dots shown on screen	most differences were between old and other drivers, elderly drivers had decreased ability to divide attention
Quimby (1988)[46]	on road	number and severity of unsafe driving actions	age, sex, maneuver type, etc.	just drive	little correlation with accidents on route
Senders, Kristofferson, Levison, Dietrich, and Ward (1967)[36,37]	on road (unused Interstate or test track)	viewing time, speed (varied with experiment)	speed, radius of curve, road type (varied with experiment)	occlusion-either press button to raise face shield or select driving speed	either varying speed or look time lead to same results, typical look/no look of 0.5/2.0 s at 97 km/h (60 mi/h)
Verwey (1991)[67]	on road	glance frequencies & durations to various places, speed, standard deviation of speed, steering wheel action rate; 6 measures were specific to merging or right turns-time to merge, distance required, speed after merging, and time, distance and speed of braking before intersection	secondary task (visual detection, visual addition, auditory addition), driving alone	driving or driving + visual detection, visual addition, auditory addition	visual tasks interfere most, fewer mirror glances with high task demands; glance frequency more sensitive than duration; driving performance measures affected by driving task (freq of accelerator, brake & clutch depressions, driving speed, std. dev. of speed. Brake pedal frequency increased and mean speed decreased w/ traffic density; SAR sensitive to workload, other driving performance variables unaffected

Table 15. Methodological studies (continued).

Experiment	Domain	Dependant Variable	Independent Variable	Task	Results/Comments
Verwey (1989)[85]	part task simulator	% of time off track, RMS tracking error, RT to nav information	stimulus modality (text vs. arrows), familiar vs. unfamiliar intersections	countersteer random movements of projected slide; shown guidance info and then road scene, indicated which way to go by pressing button	least tracking error with auditory nav, more with arrows and most with text, RT shows same pattern favoring auditory guidance
Walker, Alicandri, Sedney, and Roberts (1990)[86] (see also Walker, Alicandri, Sedney, and Roberts 1991, 1992)[87,88]	in simulator	heart rate, reaction time to gauge changes, speed (minimum, mean, variance, skew), lateral deviation	auditory vs. visual navigation system, complexity of system (3 levels), nature of loading (perceptual, cognitive, psychomotor)	drive route following advice of navigation system	heart rate was not sensitive to loading of nav device differences, some differences in RT due to navigation devices (longer for complex types), significant differences in speed (slower for more complex designs, especially complex visual), lateral position measures (average and variance of deviation) were unaffected by the navigation device
Zwahlen and Balasubramanian (1974)[42]	unused airport runway	path error	steer/no steer	either close eyes and drive or do that and take hands off wheel	$sy=k^{*}D^{*}T^{0.5}$ k=0.025 (0.683 in metric units) for steering with no visual input
Zwahlen and DeBald (1986)[43]	unused airport runway	path error	eyes closed or reading	drive with eyes closed or read article or road map	eyes closed and reading gave similar results, k=0.076

Table 15. Methodological studies (continued).

Experiment	Domain	Dependant Variable	Independent Variable	Task	Results/Comments
Zwahlen, Adams and Schwartz (1988)[44]	unused airport runway	path error	panel location, look/no look to road	use simulated car phone	both factors affected lane exceedence probability
Zwahlen, Adams, and DeBald (1988)[43]	unused airport runway	path error		use simulated touch panel to turn radio on, etc., looked steadily at panel until task completed or road as needed	task times of 5.02 and 8.39 s for radio and climate control, 3% excursions estimated for 3.7-m (12-ft) lane, developed eye fixation design guide

Table 16 shows studies examining interface design; for example, the nature of landmarks that drivers might find valuable. These studies are quite different from those in the previous category in that surveys and verbal protocols can be used to determine driver information needs. While these studies are less concerned with evaluation than some of the methodological work, they are nonetheless important in that they reflect how driver information systems should be designed.

Of the 15 studies listed in table 16, 8 were conducted on the road, 4 using surveys, 1 was conducted in the laboratory, 1 was conducted using a driving simulator, and 1 is unknown. Again, on-the-road experiments are most common.

Dependent measures explored included ratings of information quality, how pleased participants were with the interface, ratings of workload, response times and error rates for in-vehicle tasks, mean speed, speed variance, eye fixation frequencies and durations, recall percentages, and navigation errors. Most noteworthy about the dependent measures chosen is their variety.

Table 16. Information gathering studies.

Experiment	Domain	Dependant Variable	Independent Variable	Task	Results/Comments
Alm (1990) - exp 2[89]	survey	directions to a destination, map	3 origin/ destination combos	give directions to a destination and draw a map	landmarks, paths, and nodes all mentioned often, districts (except with map) and edges rarely used, common landmarks include traffic lights, highway signs, shops, bridges, gas stations
Alm (undated) - exp 1[90]	survey	directions to a destination	3 origin/ destination combos	describe how to get there	landmarks, paths, and nodes all mentioned often, districts and edges rarely used, common landmarks were traffic lights and signs, buildings, and parking lots, references were egocentric (not local or global)
Alm and Berlin (undated)[91]	on road	ratings of information quality (1 through 7), preferences for more information, ease of remembering info, ease of following instructions	amount of info given (1, 2, or 3 choice points)	drive to destination while guided verbally	give info about 2 choice points when the time between choices is less than 10 s, otherwise 1; need some what to repeat message
Alm, Nilsson, Jarmark, Savelid, and Hennings (undated)[92]	on road (easy route)	how pleased they were with the system, how easy it was to use, how distracting the information was, etc., workload (TLX)	landmarks (present/absent)	drive to destination while guided verbally and visually	landmarks made the system easier to use and were preferred

Table 16. Information gathering studies (continued)

Experiment	Domain	Dependant Variable	Independent Variable	Task	Results/Comments
Davis (1989b)[93]- (see also Davis, 1989a)[94]	on road	problems in navigation	intersection type	drive route using navigation aid	directions were modified (problems with closely space turns, timing, etc.)
Eberhard (1968)[95]	survey	% responding	-	show film and give people survey afterwards	94% said good idea but only 43% would buy, wanted HUD and both arrows and words for lane changes
Green and Williams (1992)[8]	in lab	response time, error rate	nav display location (HUD vs IP), display format (plan, persepctive, aerial), road graphic design (solid vs. outline)	look at road scene slide and press button if navigation system shows same or different intersection geometry	HUD better than IP, aerial and plan view better than perspective, solid slightly better than outline.

Table 16. Information gathering studies (continued).

Experiment	Domain	Dependant Variable	Independent Variable	Task	Results/Comments
Hook (1991)[96] (see also Hook and Karlgren (1991)[97]; Brown, Gustavsson, Hook, Lindevall, and Waern (1991)[98], Waern (1992)[99]	interviews	directions to a destination	resident vs. tourist	describe how to get there	landmarks were important, descriptions use different road hierarchies, descriptions from two groups differed in detail and route recommended
Kuiken, Miltenburg, and Winsum (1992)[100]	simulator	speed, speed variance, headway, gap acceptance when passing, occurrence of incidents, trip time, SWAT mental load, # of accidents, # of navigation errors	driving without assistance, driving with nonintegrated applications, driving with integrated applications	drive route as guided by navigation system, place calls while driving	in only 1 of 7 scenarios was level of support significant, 1 accident in no support condition, 3 in nonintegrated support condition, no difference in navigation errors between conditions

Table 16. Information gathering studies (continued).

Experiment	Domain	Dependant Variable	Independent Variable	Task	Results/Comments
Labiale (1989)[101]	on road	eye fixation, navigation errors, steering wheel movement, recall %, vehicle	map format (only road on itinerary and cross streets, network, maps + text directions, maps + auditory guidance; all combined with labels for some or all roads	drive route, either when moving or stopped shown map with route, recall 30 s later	no difference between and stopped, written guidance had highest recall, bare maps worst.; relative # of errors for road names than turn direction (left/right) varied with design, driver preferred map with auditory guidance, mean map glance time=1.28 s, 10.6 glances/30 s trial, driver decrease speed by 15 km/h (9 mi/h) when reading maps or maps with written guidance; use auditory when driving, text directions when stopped
Labiale (1990)[102];- exp 1 (see also Labiale, Mamberti, Baez, Conus, and Aupetit, (1988)[103]	on road	% of info units recalled, preferences, eye fixation data, steering wheel movements, speed	# of units of traffic info, format (visual text, single auditory, repeated auditory)	whle driving traffic info message is presented, recall it 30 s later	no differences due to format inverse relationship between # of items and recall, auditory information was preferred, fixation durations increase with # of info units (1.18 to 1.35 s), visual displays affected course control more than auditory, more likely to reduce speed with visual message than auditory one

Table 16. Information gathering studies (continued).

Experiment	Domain	Dependant Variable	Independent Variable	Task	Results/Comments
Labiale, (1990)[102] - exp 2 (see also Labiale, Mamberti, Baez, Conus, and Aupeti, 1988)[103]	on road	% of info units recalled, preferences, eye fixation data	map plus auditory or visual guidance, 1 or 3 turns	while driving route guidance message is presented, recall it 30 s later	significant advantage for visual 3 turn case only, auditory guidance was preferred, auditory format had fewer (8.6 vs 10.9) and shorter fixations (1.25 vs. 1.5 s)
Labiale (undated)[104]	on road	keyboard use time, screen viewing time, route selection time	number of route nodes	enter several routes and select best one while driving	time to enter route and evaluate it was 86 s for 1st, 55 s for alternative, therefore use when vehicle is stationary
Schraagen (1990)[105]	on road	navigation errors, # of times landmarks, etc. are mentioned	navigation ability, enlarged street names at turns	study map then think aloud as 4 routes are driven	enlarged road names at turn points led to fewer navigation errors, poor navigators memorized fewer turns and spent more time on street names, street names attended to most (about 1/2 of time) followed by road signs, topological knowledge, landmarks and road signs
Sperandio and Dessaigne (1988)[106]	% (in French)	reading speed, recall	visual or auditory with or without repetition	?	auditory messages more convenient, maps or graphics improve efficiency of visual messages

Table 17 shows the specific interface evaluation experiments, 19 studies in all. These are closest to the focus of this section. Of the 19 studies listed, 12 were conducted on the road, 3 using true laboratory methods, 2 using a part task simulator, 1 using a field survey, and only 1 using a driving simulator. Hence, in contrast to the methodological studies listed in table 15, most of the application-oriented experiments were conducted on the road. If the past is a predictor of the future, interface evaluations will generally be conducted using instrumented vehicles.

There was not any consistency across studies of the dependent measures examined, ranging the gamut from various measures of steering wheel movements, lane excursions, in-vehicle task completion times and errors, workload estimates, frequency and duration of glances to various locations, violations of traffic laws, etc.

Contained in these 3 tables (tables 15, 16, and 17) are the studies known by the author when this report was drafted, that related to in-vehicle information systems. With a knowledge of methods and techniques used for studies, as discussed in the first part of this report, these summaries of methods (road, lab, simulator, etc.) for each of the research types (methodological studies, interface design, interface evaluation) may reveal important trends. Looking at tables 15 through 17 as a group, it is clear that on-road studies of driving predominate with 30 of the 57 being conducted on the road. The relative fraction increases and the focus of the research moves from basic research to application. Perhaps this trend supports the concept that as research moves from basic to application, the design of the study should include more on-road data collection. The method selected for research done for in-vehicle information systems should then consider the operational impact of the technology being examined. If the IVHS technology is to be used in conjunction with the operational driving task, then the research method should be one which allows data collection in the operational driving environment. An isolated lab test of a car phone, which is to be used while driving, would not provide the desired data. An isolated lab test of a pre-trip planning tool, which would be used on the roadside, or in the home or office prior to driving, would provide useful data.

Along with selection of research methods, research measures need to be considered. It is noteworthy that in the research reported in tables 15 through 17 there is a lack of any consistent pattern in the selection of dependent measures. Across all three tables, the same dependent measures are mentioned, however.

An examination of these measures appears the discussion section that follows and in a subsequent report.[18]

Table 17. Interface evaluation studies.

Experiment	Domain	Dependent variable	Independent variable	Task	Results/comments
Allen, Stein, Rosenthal, Ziedman, Torres, and Halati (1991)[65]	part task simulator (slides of road scene & nav display on CRT)	% diverting at each exit	display type - ETAK-like static map, static map w/ congestion info, dynamic map w/ alt route & auditory traffic messages, arrow-based route guidance	drive simulated trip and decide when to divert	route guidance led to earliest diversion
Antin (1987)[107]	on road	Steering wheel movements, speed, foot control use (computer recorded), lane excursions (manual), direction of gaze from camera, time to read ETAK	memorized route (the control condition), vs. map vs. ETAK	drive route from memory, using map, or using ETAK	no difference between map and ETAK, eye fixation frequencies and durations used as key measures (see also Antin, Dingus, Hulse, and Wierwille, 1986), Antin, Dingus, Hulse, and Wierwille, 1990, Dingus, Antin, Hulse, and Wierwille, 1986)

Table 17. Interface evaluation studies (continued).

Experiment	Domain	Dependent variable	Independent variable	Task	Results/comments
Burgette (1991)[74]/ Fleischman (1991)[75] - TravTek evaluation	on road	trip time, route errors, times for various inputs, etc. (varies with experiment)	visual or auditory format, route type & length, etc.	(1) field study w/ rental users, (2) field study w/ local users, (3) yoked driving study, (4) Orlando test network study, (5) camera car study, (6) survey, (7) debriefing & interviews	in progress
Dingus (1988)[48]	on road	Steering wheel movements, speed, foot control use (computer recorded), lane excursions (manual), direction of gaze from camera, time to read ETAK or use existing control on display on command	control (tone), display (speedo), or function (next cross street) to select	read ETAK or use existing control on display on command while driving	means and standard deviations for each dependent variable for each item to use, few lane excursions, workload measure, driver adapted in response to external demands (see also Dingus, Antin, Hulse, and Wierwille (1986), Wierwille, Antin, Dingus, and Hulse (1988), Dingus, Antin, Hulse, and Wierwille (1989)
Dingus, Hulse, Krage, Szczublewski, and Berry (1991)[69]	Mac in lab	task completion times	task to complete	select address or destination, determine st name, store dest., use yellow pages, set voice options, summon emer service	task times of 40 to 240 s

Table 17. Interface evaluation studies (continued).

Experiment	Domain	Dependent variable	Independent variable	Task	Results/comments
Hulse (1988)[52]	on road	Steering wheel movements, speed, foot control use (computer recorded) Lane excursions (manual), direction of gaze from camera, time to read ETAK, workload estimates	attentional demand (anticipated, unanticipated)	drive route and use ETAK to get there	use of navigator was responsive to anticipated external demands (to minimize overload), correlation of workload estimates with fixation % on road and on display was low, good correlation between objective and subjective workload estimates (see also Wierwille, Hulse, Fischer, and Dingus (1987); Wierwille, Hulse, Fischer, and Dingus (1988); Hulse, Dingus, Fischer, and Wierwille (1989); Wierwille, Hulse, Fischer, and Dingus (1991)

Table 17. Interface evaluation studies (continued).

Experiment	Domain	Dependent variable	Independent variable	Task	Results/comments
Labiale (1992)[108]	on road	% correct recall	map format (only road on itinerary and cross street, network, maps + text directions, maps + auditory guidance; all combined with labels for some or all roads), driver age	driver route; and recall map 30 s after being shown	recall was best with map+written instructions, followed by simplified map, map+aural, map alone, age differences but no large interactions
McKnight and McKnight (1992)[109]	laboratory -part task simulation	time looking at navigation display, % of missed turns, % of hazards missed, steering wheel and brake position	driver age, navigation display (static area map, strip map, strip map with position, guidance arrows, strip map with position & arrows)	watch 25 min videotape of route, respond to hazards by braking, accelerating, or turning, operate turn signal to signal when turn street is next street	tone prior to turn helped, % of missed turns for guidance display was half of others (including guidance + position), % of time spent looking for guidance was 1/3 of others, no effect on failure to respond of display type, drivers perferred position+guidance, guidance alone received a low rank

Table 17. Interface evaluation studies (continued).

Experiment	Domain	Dependent variable	Independent variable	Task	Results/comments
Morita and Ogawa (1992)[110]	on road	brake and accelerator applications associated with use, # glances to display, general impression	visual vs. auditory guidance	drive route using guidance system	fewer fixations when auditory guidance added, foot control data not obtained satisfactorily-not usable, timing of messages had big affect on system usability
Pauzie and Marin-Lamellet (1989)[111]	on road	eye fixations, navigation errors	intersection type, paper vs. map	drive route as directed by map or arrow display	screen watched more than rear view mirror, nav display required more time for older drivers (gives mean times and frequencies for mirrors, landmarks, road, etc.)
Popp and Farber (1991)[112]	Mercedes driving simulator	frequency and duration of glances to displays, std deviation of lane position, speed variance, mental load (heart rate)	location of display-cluster vs. center console, amount of map detail	drive two routes	peripheral display required more glances and they were longer, lane variance was greater with central display, but no differences in speed, heart rate for 2 locations differed
Rothery, Thompson, and von Buseck (1968)[113]	in lab	RT	symbol vs. text	show road scene, operate turn signal if called for by message (keep left, exit left, etc.)	symbols better for exiting, text better for lane positioning
Rothery, Thompson, and von Buseck (1968)[113]	in lab	RT	upper vs. lower case, addition of arrows (Keep right, etc.)	move turn signal lever in correct direction	no significant differences

Table 17. Interface evaluation studies (continued).

Experiment	Domain	Dependent variable	Independent variable	Task	Results/comments
Rothery, Thompson, and von Buseck (1968)[113]	on road	response time, preference ratings	micromap vs. text	approach traffic circle, view display (controlling duration), then drive route	words took less time and were preferred
Staal (1987)[114]	field	response to survey questions	-	survey completed after returning rental car with nav system	most thought ETAK was easier to use than a map, easy to learn, and would be useful in major cities
Streeter, Vitello, and Wonsiewicz (1985)[115]	on road	navigation errors	customized route maps, voice guidance or both	drive route using navigation aid	drivers who listened to directions drove fewer miles, took less time, and made 70% fewer errors than map users, performance with both was between map alone and voice alone
Verwey, and Janssen (1988)[116]	on road	violations of traffic laws, driving time, # of nav errors, mental load (SWAT)	map vs. arrows vs. auditory, route complexity	drive three routes using a navigation system	no differences in traffic rule violations, as traffic became heavier the advantage of electronic systems increased, for complex routes driving time was 30% less with electronic systems, they made most errors with maps and fewest with auditory guidance, subjects felt more load with maps
Voss and Haller (1982)[117]	on road	rating of workload, route decision errors, glance frequencies and durations	graphic shown	driver route following instructions on visual display (ALI system)	1-2 glances of 0.8 - 0.9 s required to read display, not overload

Table 17. Interface evaluation studies (continued).

Experiment	Domain	Dependent variable	Independent variable	Task	Results/comments
West, Kemp, and Hack (1989)[118]	on road	% correct of symbols recognized, % of retrieval/entry tasks completed, % completing test drives, % of drivers reporting difficulty (command late, symbol unclear, etc.)	manual vs. verbal instructions, day vs. night, professional vs. domestic drivers	instructed in use of Autoguide, say what symbols meant, enter and retrieve destinations, driver four routes	found Autoguide helpful, safe, and easy to use, just as usable at night, no problems with entering destinations as grid coordinates, problems with instruction timing could lead to hazards

WHICH MEASURES HAVE BEEN USED TO ASSESS DRIVER INFORMATION SYSTEMS?

A summary of the measures that have been considered in the research examined for this report follows. Measures can be divided into two types: measures of input, or what the driver does to the car, and measures of output, or what the car does as a result of the driver (performance).

Input control measures are relatively easy to obtain. They are summarized in table 18. Input control measures are divided into three categories, primary (related to real-time control of the car), secondary (not related to real-time control), and overall measures of driver input. Primary measures include a variety of parameters related to the steering wheel and accelerator use, including both means and standard deviations of positions, and statistical measures of movement. Secondary tasks include using a car phone or adjusting the radio, as well as task added to assess spare information handling capacity. Most would agree that the safety of the driving task should not be compromised by the addition of the secondary task. With this logic, the output measures of the combined tasks would be merely looking at the driving performance, the same measures as for the primary task alone. Of the measures listed, considering driver vision as an input measure is not an ideal fit to this scheme. The rationale used was that measures of vision are indicators of input to the driver (as opposed to the vehicle) and hence is more appropriate to consider as an input rather than an output.

Table 18. Input measures of driving behavior and performance.

Category	Subcategory	Measure
Primary task	control input-lateral	number of steering wheel movements per unit time, number of steering wheel reversals, Steering wheel Reversal Rate (SRR), Steering wheel Action Rate (SAR), mean steering wheel angular change, variance of steering wheel angular changes
	control input-longitudinal	mean throttle position, throttle variance, number of brake applications, mean brake pressure/brake application force, braking pressure variance, number of clutch depressions
Secondary task	In-vehicle system use	response time, error rate or percentage
	detection performance	response time (to brake lights), error rates
Overall	driver vision	frequency and duration of glances to road, mirrors, in-vehicle display

If the IVHS technology is used in conjunction with the driving task, input measures for the driving task itself must not be neglected. Table 18 should be helpful in ensuring that the various input measured used by researchers to date have been considered. Again, these are distilled from an examination of all the studies listed in tables 15 through 17, and are felt to represent the major categories of input that should be considered to measure effects of advanced IVHS technologies on driver performance.

Measures of control performance (output) that have been examined are summarized in table 19. For the primary driving task they include absolute measures of lane position and yaw, their variance, and the first and s derivatives of yaw angle (rate and acceleration). Generally, lateral rate and lateral acceleration are not examined, though lateral acceleration (g) is believed to be an important determinant of driver comfort in lane change and turning maneuvers. For secondary tasks such as those considered in IVHS technologies, output measures are the consequences of slow responses or missed warning signals.

Table 19. Output measures of driving behavior and performance.

Category	Subcategory	Measure
Primary task	vehicle response (output)-lateral	mean and mean absolute lateral deviation (path error), number of lane exceedences and percent of time outside of lane, lateral deviation variance (lane variance), lateral acceleration, yaw angle, yaw rate, yaw acceleration
	vehicle response (output) - longitudinal	mean speed, speed variance, mean acceleration/deceleration, number of decelerations exceeding a specified g level, mean headway, headway variance (range rate)
Secondary task	In-vehicle system use	navigation errors, ratings of ease of use
	detection performance	number of pedestrians or lead vehicles struck, etc.
Overall	crashes	accidents, near misses
	quality of driving	QOD, unsafe driver actions, TTC, TLC
	workload	TLX, SWAT
	travel operations	trip time, distance traveled, average speed, number of stops
	physiological	heart rate variability, GSR

WHICH MEASURES SHOULD BE CONSIDERED FOR FUTURE ASSESSMENT PROTOCOLS?

SELECTION CRITERIA

To select measures for assessment, a set of criteria is needed to guide the selection process. There is a considerable parallel between the selection of the test protocol (discussed in the followon report) and the selection of particular dependent measures described here.[18] The criteria for the selection of dependent measures are as follows:

1. Indicativeness—The dependent measure reflects the underlying notion or hypothesis which the study is to address. It is important to clearly state the research question prior to selection of measures. As an example, safety and ease of use can be two quite different research parameters, and may have different measures.

2. Sensitivity to design differences—This is an extension of indicativeness. Changes in the product or service that have real impact should be measureable. This is necessary for engineering analyses.

3. Risk to drivers and experimenters—Driving can be dangerous. Measures that add unnecessary risk to the driving task should be avoided. Also needing consideration are minimizing pain and even embarrassment to subjects.

4. Ease of Measurement—In deciding when to collect data, the cost of the collection effort must be weighed against the benefits of the data. Easy to measure implies minimal equipment and minimal software.

5. Analyzable—Some measures are either difficult to reduce because of the physical format of the data collected or the need for special statistical tools.

6. Repeatable—Replicability is a cornerstone of scientific methods and in establishing truth. For measures to be repeatable, it is important to know which factors affect repeatability and to be able to control those factors. For example, the radius of curvature affects the difficulty of driving a road and the associated workload. Hence, comparable (or preferrably identical) roads should be driven for comparisons of in-vehicle displays.

7. Acceptance by the scientific and engineering community—The results of research are to be applied by both designers and researchers. If the likely users of the experimental results do not accept the measurement protocol, they are unlikely to be convinced by the results. Part of this involves understanding of the measurement itself. In the human factors domain, this has been a problem with the application of spatial frequency-based measures of vision to the assessment of image quality. Users often do not understand those measures.

8. Fits into an available experimental context—Driving measures tend to be context specific. For example, the measurements collected in a survey are usually quite different from those collected while driving a vehicle on a test track. Hence, if for some reason a test track protocol has been selected, consideration of survey-related measures in most cases should be dropped. In other cases, the data collection capability may not exist. For example, on-the-road measurement of driver eye fixations can be very informative, but collection of eye fixations requires a vehicle outfitted with very special recording equipment, equipment that is not widely available.

The selection of measures, both of driver input and output (performance), should be done with the above considerations. The choice of measures will to a great extent determine the worth of the research. As stated, to examine the indicativeness of a measure, a clear description of the purpose of the measurement is required. As noted in the preface to this report, some of the goals of IVHS are to improve traffic operations, reduce accidents, and reduce air pollution from vehicles. Activities in this project includes (1) quantifying the influence of safety, (2) quantifying the effectiveness of information transfer, and (3) assessing driver comfort, convenience, and confidence. Hence, the qualities of interest, both to this project and to IVHS in general are:

- Safety—Reducing crashes.
- Operational—Being more efficient, saving time and energy, providing increased capacity and increased functionality.
- Enhancing the experience of driving—Making driving more enjoyable, even fun. This is the personal aspect of driving.

It is in the context of these various sets of goals and the selection criteria given previously that the measures of interest will be considered.

Even considering the above in the design of the study, it should be understood that selecting good measures of driving performance is not simple. It is clear that there is a need for multiple criteria, as suggested by the DRIVE Task Force and others. However, identifying exact levels of those criteria that represent safe (and acceptable) driving is premature given the current state of knowledge.[24] Therefore, interpreting the meaning of a change in driver performance is somewhat arbitrary. Short of creating a danger to other drivers on the road, secondary tasks which require some amount of driver attention are commonly acceptable. The amount of attention that should be reserved for safe driving is left up to the driver's personal judgment. With this qualification in mind, research can be done examining the relative impact of secondary tasks, but conclusions on the significance of these results will be subject to nonscientific interpretation. Looking to existing research for insights on the complicated problem of interpreting measures of performance, Wierwille has shown that people adapt to task demands. Those demands can either overload drivers' overall capacity, or overload particular channels. The literature suggests the most common overload is of the visual channel.[67]

Having discussed matters that should be considered when selecting research measures, the following section considers the various categories of output measures and input measures likely to be used in IVHS research.

OUTPUT MEASURES - LATERAL CONTROL

In spite of these limitations, the current state of knowledge provides considerable insight into the selection of measures. One approach is to first consider measures that have the most direct impact on consequences (the output measures of table 19). Clearly, primary task output measures are indicative of safety. If a vehicle is wandering in the lane, crashes are more likely. Variations in lateral position will also have an operational penalty by disrupting traffic flow, and make driving more difficult, making the driving experience less pleasant.

Of the measurements in this category, lane exceedences would seem to be an obvious choice, since they represent collision opportunities. As Zwahlen has shown, giving the driver an attention-demanding task can cause the driver to deviate from straight ahead. When that reaches extremes, drivers actually wander outside of their lanes (and potentially collide with another vehicle or roadside object). However, lane exceedences occur infrequently, so that measure tends to be insensitive to differences in the attentional demands of various in-vehicle displays. As Wierwille has shown, lane exceedences are not well correlated with other driving performance measures.

As suggested by MacAdam, the standard deviation of lateral position is a more sensitive measure than mean deviation from the center.[83] Noy also found that the standard deviation was affected by task difficulty.[62,63] In brief, when drivers pay less attention to the control/steering task (due to fatigue or the attentional demands of in-vehicle displays), they make fewer path corrections, but the corrections they make are larger. This behavior is most directly reflected in the increase in lane variance. To a lesser extent, this is also reflected in an increase in mean yaw angle.

Hence, measurements in this category are indicative of safety and operational problems, and at least at a surface level, this notion is accepted within the scientific and engineering communities. Exactly how these measures reflect the personal experience of driving is unknown though clearly less lateral control is not desired. It certainly will make drivers uncomfortable. The sensitivity of measures based on vehicle lateral position to design differences varies widely with the particular measurement in this category, with further research needed to determine the exact relationship. In itself, collecting these measurements poses no risk to drivers and experimenters. However since the basic research to determine the association may require exploration of risky situations, basic research may pose some risk to drivers.

As indicated previously, there are significant technological hurdles to be overcome in the measurement of vehicle lateral position. Few researchers have collected measures of this type, and data on repeatability are limited. Further, collection of this class of measures typically requires an instrumented vehicle with a lane tracker.[119] At this time there are probably fewer than 10 vehicles in the world with that capability. In most lane trackers, a video image is scanned for lane markers, and after geometric transformations, the lateral position is determined. Only a few lane trackers can determine yaw angle.

If only lane exceedences are desired, they can be obtained by periodically looking out the window and manually recording position, or by post-test review of a forward scene videotape (or from a camera attached to the side of the car and aimed downward). In a driving simulator, lateral position (as well as yaw angle) is one of the results of vehicle dynamics calculations and can readily be saved to a file.

OUTPUT MEASURES - SPEED CONTROL

If a vehicle is driving at a variable speed or too fast, crashes are more likely. While high speeds are associated with greater throughput and a more pleasant driving experience, variability in speed reduces road throughput, and makes driving more difficult, diminishing the experience of driving. Thus, measures of speed control are indicative of safety, operational, and personal aspects of driving, though in a complex manner.

Both mean speed and speed variance may be affected by the use of in-vehicle displays and have been shown to be affected by external demands, though additional research to address the sensitivity of speed control measures is desired.[67] In general, when people are given in-vehicle tasks with heavy attention demands, they tend to slow down to provide themselves with a greater safety margin. This is sometimes an unconscious behavior. Also, because they attend to speed less, their speed may be more variable, even likely to increase because mean speed and speed variance tend to be correlated. Obviously, in braking situations, rates and accelerations could be affected by task demands associated with in-vehicle displays; however, such measures concern transient events, which, again, are more difficult to assess.

A consequence of choosing a particular speed while driving is the headway between the subject's vehicle and a lead vehicle. Headway and headway variance are linked to the frequency of rear-end collisions, and are therefore worth considering.

Measurements of speed control do not usually pose any special risk to drivers or experimenters. In contemporary vehicles recording speed is quite easy with the speed signal being an output of the electronic engine controller. Some filtering of the signal may be required before it can be processed by a computer. The measurement of acceleration requires somewhat more complex and expensive sensors, but the effort is only somewhat greater than that for speed measures. Headway measurement is complex, requiring either custom-made radar-, laser-, or sonar-based sensors. Headway measurement is particularly difficult on curves. In the future, when vehicles are outfitted with intelligent cruise control systems or collision avoidance systems, recording headway measures may be as straightforward as current methods for recording speed. All of these measures of speed control are analyzable and accepted; though due to current sensor limitations, there are limits to the repeatability of headway measurements. These limitations are not present for laboratory or simulator experiments.

OUTPUT MEASURES - SECONDARY TASKS

Secondary tasks include real in-vehicle tasks that may add to the driver's workload and artificial tasks used to assess the processing capacity remaining. Real tasks include dialing phones and using navigation systems, with their respective performance measures being the number of calls successfully completed and navigation errors. These measures are directly indicative of operational performance (in this case, the effectiveness of information transfer).

Measures of detection performance, intended to assess spare capacity, have not been used very often. In brief, the concept is that driving involves not only maintaining a path, but searching for objects of concern, and are indicative of safety margins. This includes pedestrians that might dart into the path of a moving car, responding to the brake lamps of a lead vehicle, looking for vehicles to cross one's path at intersections, etc.[64] Tasks can also be somewhat artificial such as pressing buttons when instrument panel gauges go out of tolerance, response time to single letters, two-choice response time with letters, two-digit addition, dot counting, etc.[83,84,86] There does not seem to be a simple pattern to explain the results. Sometimes the secondary task is affected by the presence of an in-vehicle task and sometimes it is not. The clearest perspective comes from the work of Noy, which emphasizes the importance of within-modality interference.

There is no standard method for collecting or analyzing secondary task data. Each researcher chooses a method compatible with the equipment and resources available to them.

These measures should be sensitive to in-vehicle attentional demands because many of the situations can be precursors to accidents, however their sensitivity to design variations has not been established. The weaknesses of these measures is that they are discrete. To assess an in-vehicle display, the timing of the event relative to in-vehicle system use (and the extent to which the event unfolds over time) is important. Further, while such events are relatively easy to schedule in a driving simulator, many of them (e.g., pedestrians crossing the vehicle's path) are difficult to safely execute on the road, posing a risk to the driver, experimenter, and other road users. With some creativity (such as using foam core outlines of pedestrians), risk can reduced, but the development of test facilities using such approaches may be costly. Also, because these events are unique, once they have occurred, their surprise value is gone and their repeated presentation diminishes their utility. Repeatability within individuals is therefore difficult to assess. The initial outcome, however, can be telling in terms of the safety of a system.

The extent to which secondary task measures relate to drivers' comfort, convenience, and confidence in information systems is unknown. This topic has not been examined in the literature.

OUTPUT MEASURES - OVERALL

In the past, unsafe designs have been identified by counting how often those designs were associated with accidents. However, many IVHS technologies have yet to be implemented. Their association with crashes has yet to be established. Even after they are implemented, most crash data bases do not provide a means for identifying if an IVHS device was present or being used prior to a crash, so establishing a connection between devices and crashes will be difficult. The extent to which crash data reflect operational benefits (e.g., ease of information transfer) or the quality of the experience of driving are unknown.

Looking at the selection of crash measures as a source of output information is another possibility for researchers. Traditionally, such information has been contained in data bases assembled by the Federal and State governments. Each data base has its own structure, though fairly routine statistical methods are used to identify relationships between variables. There are minimal drawbacks with analysis or acceptance of the results. For IVHS applications, these analyses are not yet possible as the presence of IVHS devices is not coded, so there is no data to analyze.

In searching for surrogate measures, subjective assessment of the quality of driving could be indicative of when crashes might occur. Subjective quality of driving has not been used to examine operational or individual performance on the road, and its sensitivity to interface design differences has not been examined. Quality of driving is not a prime candidate for assessing ease of operation, driver comfort with in-vehicle systems, or related matters. As another potential measure, it is commonly believed that driving experts, such as driver trainers, can identify dangerous acts that drivers perform. Those acts can be precursors to accidents. Quimby's work suggests that the correlation is not very good; nonetheless, common belief in the linkage persists.[46] The weakness of this method is the reliance on trained observers and the difficulty of calibrating those observers to achieve repeatable results. Quantification of unsafe driving behaviors and their validation using simulation is needed. Thus, while the equipment needs for quality of driving assessments are minimal, there are many unresolved questions about the data obtained from such evaluations.

Direct subjective assessment of driving workload is also a possibility (e.g., TLX, SWAT). SWAT and TLX were described in the initial section of this report in conjunction with the discussion of Zaidel's 1991 report. Work by Wierwille and others suggests that workload ratings can be indicative of primary task (safety-related) demands, but it remains unclear what should be emphasized—average workload or peak workload. Workload ratings are sensitive to operational differences of in-vehicle devices but not as sensitive as direct measures of driving performance. It is not known if they reflect differences related to the experience of driving. The workload literature is voluminous, clearly establishing that workload measures are analyzable, repeatable, and well accepted. Workload assessments can be conducted in a wide variety of contexts.

Summary measures of driving show promise of being useful for practical assessment of the safety of in-vehicle systems. Researchers at TNO have expressed interest in both TLC and

TTC.[40,41] (For a description of TLC and TTC, see earlier sections of this report describing Godthelp's research.) The driver's goal on a moment-to-moment basis is to minimize the opportunity for collisions; hence, TTC should be a measure of how safely one is driving. The difficulty with TTC is that computing it requires a human analysis of each video frame, computation of the trajectories of everything in the scene, and then predictions about potential conflicts with each object. These calculations require such a considerable effort that few studies have examined these measures. The development of equipment to compute TTC on an ongoing basis should be a priority item.

TLC is somewhat easier to determine, in that it requires information only on a vehicle's lateral position, yaw angle and rate, and forward velocity and acceleration, as well as data on road curvature. This information can be obtained from a lane tracker and from vehicle speed and acceleration sensors. In some cases it may be possible to obtain all the needed data from an advanced video lane tracker. This computational capability is not often available, which is one reason why TLC is often not used. Development of hardware to determine TLC automatically is appropriate for technical development. TTC and TLC show considerable promise, but matters pertaining to analyzability, repeatability, and data collection hardware need to be addressed.

It seems likely that these summary measures could reflect ease of operation. If the driver is distracted by the in-vehicle system, lane position will be more variable, resulting in decreases in TTC and TLC. Driver comfort with the system should also decrease. However, there is no data to address the operational and personal connections with these summary measures.

Travel operations measures (trip times, number of turns, etc.) are accepted measures of the operational performance of an interface. In a secondary way, they are connected to safety in that greater exposure to the road (more time on the road, more turns) provides more opportunity for crashes. Travel operations measures were widely used in the TravTek project. They are straightforward to collect and analyze. Distance data may come from manual reading of odometers or counting of wheel pulses from a speed sensor in a instrumented vehicle. In a simulator these data are directly available from the vehicle dynamics calculations. Data on turns may be manually recorded in real time or postprocessed in a manual review of videotapes of test sessions.

Potential physiological indicators reported in driving studies include heart rate, the variance of heart rate (arrhythmia), respiration rate, and galvanic skin response (GSR). Heart rate is generally not sensitive to measures of attentional demand, but rate variability may be.[86,80] Physiological measures tend to be more common in studies conducted in Japan, than in studies conducted in the United States and Europe. In general, physiological measures are most sensitive to the experience of driving and less sensitive to operational differences and safety. These measures require considerable experience to collect. Special instrumentation is also required to amplify and filter signals. There is some debate as to how best to analyze this type of data.

INPUT MEASURES - LATERAL CONTROL

Driving task execution measures concern the actions the driver carries out to sense and maneuver the car on a moment-to-moment basis. Measures of interest include mean and variance of steering wheel angle, steering wheel reversals, and the spectrum of steering wheel input. Of the steering wheel measures, the number of reversals over time and the spectrum of input appear to be the most sensitive to changes in driving behavior. Spectral qualities of steering wheel position are more difficult to analyze than other measures of steering behavior. Of these measures, steering wheel action rates seem to be most indicative of task loading, though further research on the topic of lateral control is desired. These measures should be sensitive to the operational demands of in-vehicle devices since time spent operating the device is not spent steering. They should also be indicative of safety since not attending to the primary task of steering may lead to an accident. In fact, it could be that steering input is a better measure of safety than various measures of lateral position because the inertia of the vehicle "filters out" some of the input differences.

Recording of steering wheel position is usually accomplished using a string potentiometer connected to a computer. In future drive by wire vehicles, the steering wheel angle may be directly accessible from a steering motor controller.

INPUT MEASURES - SPEED CONTROL

Speed control measures of interest include mean throttle (accelerator) position, throttle variance, the number of brake pedal actuations, and the number of throttle actuations. Just as with lateral control measures, throttle position measurements may be better indicators of driver performance that the vehicle output measures (speed, lane position) because the output is not smoothed by the vehicle dynamics. Throttle opening, a measure directly related to throttle position, can be obtained from the electronic engine controller and recorded by a computer.

It is suspected that speed control measures may be indicative of both the safety and operational performance of in-vehicle systems, though the strength of those relationships is unknown. The connection of measures of speed control with the experience of driving is also unknown.

INPUT MEASURES - SECONDARY TASKS

Response times and response errors are the most commonly used measures of secondary task performance, with the measure depending upon the task selected. The data collection protocol is task specific. The reservations expressed concerning output measures of secondary tasks also hold for input measures as well, since the reservations are related more to the task than the measure. Most of these drawbacks are not present in fairly simple secondary tasks, such as pressing buttons on the steering wheel when lights mounted on the hood of a test vehicle are detected. It is uncertain, however, how strong the connection is between the secondary task dependent measure (light detection time, percentage of lights not detected) and safety-related variables such as crashes. The connection with operational and personal characteristics is even more remote, and acceptance of them by the engineering community is less than for other measures.

In contrast to abstract tasks, measurement of performance in the completion of real in-vehicle tasks (such as the time to dial a phone) is well accepted as a measure of the operational performance of the device used. Such measures are indicative of design differences.[11] Those measures should be related to the enjoyment of using a phone or other in-vehicle device. For warning systems (such as IVSAWS), performance measures such as detection time and errors are viewed as operational measures of such systems. Task performance measures tend to be easy to collect and analyze, and are repeatable.

As with many of these measures, the extent to which secondary task input measures reflect the experience of driving is unknown. There is no reason to expect a direct linkage.

INPUT MEASURES - DRIVER VISION

Eye fixation data can be extremely informative, but are very difficult to collect and analyze.[120] They can be indicative of safety, operational, and personal aspects of interfaces, and are sensitive to design differences. Repeatability within individuals has not been given much attention. Eye fixation data are widely accepted by scientists and engineers. Clearly, the likelihood of an accident increases with the number and length of eye fixations away from the road scene. The literature suggests that when presented with in-vehicle visual demands, the first task to drop out is mirror scanning.

Eye fixations can either be collected by aiming a camera at a driver and recording where he or she looks or by using special recording devices. While the direct recording method seems straightforward, the frame-by-frame reduction of the data can take 30 to 40 times the recording time, a costly process. Analysis beyond fixation durations and frequencies (to examine patterns) is very time consuming.

Systems that automatically record fixation coordinates cost from $25,000 up to $100,000, and are beyond what most research organizations can afford. Further, many systems restrict the field of view, making driving more difficult. Nevertheless, visual scanning behavior can be an important index of potential safety problems. Where eye fixations can be economically recorded, they should be.

Some sense of the attentional demands of driving can be obtained indirectly using either helmets or goggles that temporarily block the view of the driver.[36,37] To date, this approach has been used primarily to determine the demands of the primary task, not the loading of in-vehicle tasks. One potential manipulation would be to reduce input and make the primary task so difficult that use of an added in-vehicle display would sharply degrade the primary task. This degradation is likely to occur since the primary source of overload is visual, as mentioned earlier. Use of such a method for routine assessment of in-vehicle systems seems excessively complex, though it can be useful for theoretical analyses. There are also significant risks to the driver.

DATA ANALYSIS

Most of these measures described in this section are collected in real time by instrumented vehicles or simulators, with sampling typically occurring several times per s. To reduce the data, the data are first filtered to identify and correct faulty data. This process involves examining histograms of data to identify outlier and short-term measures of variability. This is often done manually for each test session for each driver subject. Faulty data can occur as the result of electrical malfunctions, environmental interference with the lane tracker, typing errors in identifying file names, and for a variety of other reasons. Because the environment is less harsh, there are generally fewer problems with simulator data than data collected on the road. Anomalies may require the manual review of session videotapes.

The next step involves computing summary statistics for each driver by task and road segment. Typically this is done using computer software for one driver session at a time as a further check of the data. The results of those analyses are then entered into standard statistics packages for computation of ANOVA and regression statistics, as well as correlation statistics.

SUMMARY ON MEASURES

It should be apparent that there is no single best measure or limited set of measurements that are appropriate for assessing the safety, operational, and personal aspects of driver information systems. Considerations pertaining to the selection of measures was provided. For many of the input and output measurements discussed in this section, data on repeatability is lacking. Several of them have significant instrumentation requirements; others have significant data analysis requirements. Given these limitations, the next section provides general recommendations as to which measures to consider collecting in studies of driver interfaces. Readers interested in further discussion of these measures should see reference 18.

CONCLUSIONS

Again, the selection of measures by the researcher should reflect the use of the equipment or system being examined. In the foregoing discussion, the primary emphasis has been on safety. Ease of use is also important, but the measures of usability tend to be very system-specific. In the case of route guidance systems, the measurements of interest are the time to learn the system, the number of wrong turns made per trip, and the time to reach a destination. For vehicle monitoring and IVSAWS systems, the appropriate dependent measures are the time to read a display or hear a message, and the probability that a correct response ensues. For car phones, candidate measures include the time to dial a phone number and the frequency of errors.

Thus, this report suggests that the standard deviation of lane position, mean speed, and speed variance are likely to reflect safety, and, to some extent, ease-of-use problems with in-vehicle displays. Speed-related measures are easy to collect. Lane position, especially on the road, is more difficult. However, because maintenance of speed and lane position are protected tasks, they are unlikely to be perturbed when the risk to the driver is moderate.

More sensitive to attentional demands are eye fixation data. Ordinarily, as demands increase, the fraction of time spent looking at mirrors decreases. Displays that are difficult (and potentially less safe) to use have longer fixation times and require more glances. The drawback of eye fixation data is the considerable difficulty in collecting and analyzing it. The standard deviation of lane position, speed and speed variance, and eye fixation distributions are the preferred measures for the assessment of in-vehicle displays.

Direct performance measures, such as response times and error rates, certainly reflect the ease of use of in-vehicle systems. However, data linking specific response times and error rates to specific numbers of accidents do not exist. Response time and error data are most useful for comparing alternative interface design and, using simple experiments, deciding which design is best. Hence, they may be difficult to relate to safety. Nonetheless, to assess operational performance, it is essential that time and error measures be collected.

Also of interest are TTC and TLC, measures suspected to be tightly linked with accidents. While estimates for them can be readily obtained in simulators, obtaining these measures in test vehicles is problematic. The development of equipment to measure TTC and TLC is needed.

Less useful are measures of secondary task performance. While they can be indicative of specific overloads (especially visual), task performance is difficult to relate to levels of driving safety (as measured by the number of crashes).

Finally, some researchers favor the use of physiological measures as indicators of driving workload. While the connection of some with driving pleasure is clear, the link with performance is not. However, given this interest in exploring driver comfort, convenience, and comfort, these measures need further attention.

Overall, direct measures of driving performance (standard deviation of lane position, speed and speed variance) are preferred as indicators of the safety and ease of use of driver information systems. For visually based systems, eye fixations should also be examined. If ease-of-use requirements are to be taken seriously, then system-specific performance measures (e.g., number of wrong turns for a navigation system) should also be collected. Physiological measures also need further attention, but at the level of basic research rather than product evaluation. Again, for a further discussion of these measures, see the followon report.[18]

For many aspects of automotive engineering—development, design, and production—there are tradeoffs. That is true in safety engineering evaluations as well. Objectives vary, as do the funds, equipment, and schedule to achieve them. While the selection of measures of effectiveness in the foregoing discussion considers what is scientifically reasonable to do, not everyone has the resources necessary and it may not be practical to collect these measures. In evaluating a test protocol, this must be kept in mind.

Hopefully, this report has provided a summary of research methods and measures employed in key studies pertaining to driver interface evaluation; provided insight into the selection process; and offered useful suggestions for the selection of measures. This information has been provided with the intent of guiding assessment protocols to facilitate the evaluation of IVHS technologies.

REFERENCES

1 Green, P. (1993). *Human Factors of In-Vehicle Driver Information Systems An Executive Summary* (Technical Report UMTRI-93-18), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

2 Green, P., Williams, M., Serafin, C., and Paelke, G. (1991). Human Factors Research on Future Automotive Instrumentation: A Progress Report, *Proceedings of the 35th Annual Meeting of the Human Factors Society*, pp. 1120-1124.

3 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.

4 Brand, J.E. (1990). *Attitudes Toward Advanced Automotive Display Systems: Feedback from Driver Focus Group Discussions* (Technical Report UMTRI-90-22), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

5 Green, P. and Brand, J. (1992). *Future In-Car Information Systems: Input from Focus Groups* (SAE paper 920614), Warrendale, PA: Society of Automotive Engineers.

6 Green, P., Serafin, C., Williams, M., and Paelke, G. (1991). "What Functions and Features Should Be in Driver Information Systems of the Year 2000?" (SAE paper 912792), *Vehicle Navigation and Information Systems Conference* (VNIS'91), Warrendale, PA: Society of Automotive Engineers, pp. 483-498.

7 Serafin, C., Williams, M., Paelke, G., and Green, P. (1991). *Funcations and Features of Future Driver Information Systems* (Technical Report UMTRI-91-16), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

8 Green, P., and Williams, M. (1992). "Perspective in Orientation/Navigation Displays: A Human Factors Test," *Conference Record of Papers, the Third International Conference on Vehicle Navigation and Information Systems (VNIS'92)*, (IEEE Catalog #92CH3198-9), Piscataway, NJ: Institute of Electrical and Electronics Engineers, pp. 221-226.

9 Williams, M. and Green, P. (1992). *Development and Testing of Driver Interfaces for Navigation Displays* (Technical Report UMTRI-92-21), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

10 Paelke, G., Green, P., and Wen (1993). *Development and Testing of a Traffic Information System Driver Interface* (Technical Report UMTRI-93-20), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

11 Serafin, C., Wen, C., Paelke, G., and Green, P. (1993). *Development and Human Factors Tests of Car Telephones* (Technical Report UMTRI-93-17), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

12 Hoekstra, E., Williams, M., and Green, P. (1993). *Development and Driver Understanding of Hazard Warning and Location Symbols for IVSAWS* (Technical Report UMTRI-93-16), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

13 Williams, M., Hoekstra, E., and Green, P. (1993). *Development and Evaluation of a Vehicle Monitor Driver Interface* (Technical Report UMTRI-93-22), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

14 Green, P., Hoekstra, W., Williams, M, Wen, C., and George, K. (1993). *Examination of a Videotape-Based Method to Evaluate the Usability of Route Guidance and Traffic Information Systems* (Technical Report UMTRI-93-31), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

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

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

17 Green, P. (1993). *Measures and Methods Used to Assess the Safety and Usability of Driver Information Systems* (Technical Report UMTRI-93-12), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

18 Green, P. (1993). *Suggested Procedures and Acceptance Limits for Assessing the Safety and Ease of Use of Driver Information Systems* (Technical Report UMTRI-93-13), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

19 Levison and Cramer. (1993). *Description of the Integrated Driver Model* (BBN Report 7840), Cambridge, MA: Bolt, Beranek and Newman.

20 Levison, W.H. (1993). *A Simulation Model for the Driver's Use of In-Vehicle Information Systems* (TRB Paper 930935), paper presented at the Transportation Research Board Annual Meeting, Washington, D.C., January 10-14, 1993.

21 Wickens, C.D. (1989). "Resource Management and Time-sharing," in Elkind, J.I., Card, S.K., Hochberg, J. and Huey, B. (eds.), *Human Performance Models for Computer-Aided Engineering*, Washington, D.C.: National Academy Press.

22 DRIVE Safety Task Force (1991). *Guidelines on Traffic Safety, Man-Machine Interaction and System Safety*, June, Brussels, Belgium.

23 Zaidel, D.M. (1991). *Specification of a Methodology for Investigating the Human Factors of Advanced Driver Information Systems* (Technical Report TP 11199 (E)), Ottawa, Canada: Road Safety and Motor Vehicle Regulation, Transport Canada.

24 Robertson, J. and Southall, D. (1992). *Summary Report on the Feasibility of Developing a Code of Practice for Driver Information Systems*, Loughborough, Leicestershire, United Kingdom: ICE Ergonomics Ltd.

25 Labiale, G. (1984). *Approche Psycho-Ergonomique des Systemes D'Information et de Guidage Routier des Conducteurs Automobiles (Psycho-Ergonomic Approach to Driver Information and Route Guidance System)* (INRETS Technical Report IRT, NNE 62), Institut de Recherche des Transports Centre D'Evaluation et de Recherche des Nuisances et de L'Energie, October.

26 Carsten, O. (1991). personal communication.

27 Franzen, S. (1991). personal communication.

28 Stevens, A. and Pauzie, A. (1992). "Safety Standards for Driver Information Systems," *The 3rd International Conference on Vehicle Navigation and Information Systems (VNIS'92)* (IEEE Catalog # 92CH3198-9), Piscataway, NJ: Institute of Electrical and Electronics Engineers, pp. 26-31.

29 Zaidel, D.M. (1991). *Specification of a Methodology for Investigating the Human Factors of Advanced Driver Information Systems* (Technical Report TP 11199 (E)), Ottawa, Canada: Road Safety and Motor Vehicle Regulation, Transport Canada.

30 Reid, G.B. and Nygren, T.E. (1988). "The Subjective Workload Assessment Technique: A Scaling Procedure for Measuring Mental Workload," pp. 185-218 in P.S. Hancock and N. Meshkati (eds.), *Human Mental Workload*, Amsterdam, The Netherlands: Elsevier Science.

31 Lysaght, R.J., Hill, S.G., Dick, A.O., Plamondon, B.D., Linton, P.M., Wierwille, W.W., Zaklad, A.L., Bittner, A.C., and Wherry, R.J. (1989). *Operator Workload: Comprehensive Review and Evaluation of Operator Workload Methodologies* (Technical Report 851). Alexandria, VA: U.S. Army Research Institute.

32 Hart, S.G. and Staveland, L.E. (1988). "Development of a NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in P.S. Hancock and N. Meshkati (eds), *Human Mental Workload*, Amsterdam, The Netherlands: Elsevier Science.

33 Hendy, K.C. (1990). "Measurement and Prediction of Operator Information Processing Load," pp. 19-25 in R.M. Heron (ed.), *Proceedings of the Second Seminar in Transportation Ergonomics* (Publication TP 8855), Montreal, Quebec, Canada: Transport Canada, Transportation Development Center.

34 Knowles, W.B. (1963). "Operator Loading Tasks," *Human Factors*, 5, 155-163.

35 Michon, J.A. (1966). "Tapping Regularity as a Measure of Perceptual Motor Load," *Ergonomics*, 9, 401-412.

36 Senders, J.W., Kristofferson, A.B., Levison, W.H., Dietrich, C.W., and Ward, J.L. (1967b). *The Attentional Demand of Automobile Driving* (BBN Technical Report 1482), Cambridge, MA: Bolt, Beranek and Newman, Inc.

37 Senders, J.W., Kristofferson, A.B., Levison, W.H., Dietrich, C.W., and Ward, J.L. (1967a). "The Attentional Demand of Automobile Driving," *Highway Research Record #195*, Washington, D.C.: Highway Research Board, 15-32.

38 Senders, J.W., Kristofferson, A.B., Levison, W.H., Dietrich, C.W., and Ward, J.L. (1966). *An Investigation of Automobile Driver Information Processing* (BBN Technical Report 1335), Cambridge, MA: Bolt, Beranek and Newman, Inc. (NTIS PB 170 879).

39 Hicks, T.G. and Wierwille, W.W. (1979). "Comparison of Five Mental Workload Assessment Procedures in a Moving Base Driving Simulation," *Human Factors*, April, 21(2), 129-144.

40 Godthelp, H., Milgram, P., and Blaauw, G.J. (1984). "The Development of a Time-Related Measure to Describe Driving Strategy," *Human Factors*, June, 26,(3), 257-268.

41 Godthelp, H. (1988). "The Limits of Path Error-Neglecting in Straight Line Driving," *Ergonomics*, 31(4), 609-619.

42 Zwahlen, H.T. and Balasubramanian, K.N. (1974). "A Theoretical and Experimental Investigation of Automobile Path Deviations When Driver Steers with No Visual Input," *Transportation Research Record #520*, 25-37.

43 Zwahlen, H.T. and DeBald, D.P. (1986). "Safety Aspects of Sophisticated In-Vehicle Information Displays and Controls," *Proceedings of the Human Factors Society-30th Annual Meeting-1986*, Santa Monica, CA: The Human Factors Society, pp. 256-260.

44 Zwahlen, H.T., Adams, C.C., Jr., and DeBald, D.P. (1988). "Safety Aspects of CRT Touch Panel Controls in Automobiles," in *Vision in Vehicles II* [A.G. Gale, et al. (eds.)], The Netherlands: North Holland, pp. 335-344.

45 Zwahlen, H.T., Adams, C.C., Jr., and Schwartz, P.J. (1988). "Safety Aspects of Cellular Telephones in Automobiles," *18th International Symposium on Automotive Technology and Automation* (Paper 88058).

46 Quimby, A.R. (1988). *In-Car Observation of Unsafe Driving Actions* (Research Report ARR 153), Victoria, Australia: Australian Road Research Board, April.

47 Van Der Horst, R. and Godthelp, H. (1989). "Measuring Road User Behavior with an Instrumented Car and an Outside-the-Vehicle Video Observation Technique," *Transportation Research Record #1213*, Washington, D.C.: Transportation Research Board, 72-81.

48 Dingus, T.A. (1988). *Attentional Demand for an Automobile Moving-Map Navigation System* (unpublished Ph.D. dissertation), Blacksburg, VA: Virginia Polytechnic Institute and State University, Department of Industrial Engineering and Operations Research.

49 Wierwille, W.W., Antin, J.F., Dingus, T.A., and Hulse, M.C. (1988). "Visual Attentional Demand of an In-Car Navigation Display," in Gale, A.G. et al. (eds.), *Vision in Vehicles II*, Amsterdam, The Netherlands: Elsevier Science, pp. 307-316.

50 Dingus, T.A., Antin, J.F., Hulse, M.C., and Wierwille, W.W. (1986). *Human Factors Test and Evaluation of an Automobile Moving-Map Navigation System, Part I Attentional Demand Requirements*, (IEOR Department Report 86-03), Blacksburg, VA: Virginia Polytechnic Institute and State University, Department of Industrial Engineering and Operations Research, September.

51 Dingus, T.A., Antin, J.F., Hulse, M.C., and Wierwille, W.W. (1989). "Attentional Demand Requirements of an Automobile Moving-Map Navigation System," Transportation Research, 23A(4), 301-315.

52 Hulse, M.C. (1988). *Effects of Variations in Driving Task Attentional Demand on In-Car Navigation System Usage* (unpublished master's thesis), Blacksburg, VA: Virginia Polytechnic Institute and State University, March.

53 Wierwille, W.W., Hulse, M.C., Fischer, T.J., and Dingus, T.A. (1987). *Effects of Variations in Driving Task Attentional Demand on In-Car Navigation System Usage* (IEOR Department Report 87-02, General Motors Research Laboratories Report CR-88/02/OS), Blacksburg, VA: Virginia Polytechnic Institute and State University, September.

54 Wierwille, W.W., Hulse, M.C., Fischer, T.J. and Dingus, T.A. (1988). *Strategic Use of Visual Resources by the Driver While Navigating with an In-Car Navigation Display System*, (SAE Paper 885180), Warrendale, PA: Society of Automotive Engineers.

55 Hulse, M.C., Dingus, T.A., Fischer, T.A., and Wierwille, W.W. (1989). "The Influence of Roadway Parameters on Driver Perception of Attentional Demands," pp. 451-456 in Mital, A. (ed.), *Advances in Industrial Ergonomics and Safety*, 1, New York, NY: Taylor and Francis.

56 Wierwille, W.W., Hulse, M.C., Fischer, T. J., and Dingus, T.A. (1987). *Effects of Variations in Driving Task Attentional Demand on In-Car Navigation System Usage* (IEOR Department Report 87-02, General Motors Research Laboratories Report CR-88/02/OS), Blacksburg, VA: Virginia Polytechnic Institute and State University, September.

57 Wierwille, W.W., Hulse, M.C., Fischer, T.J., and Dingus, T.A. (1988). *Strategic Use of Visual Resources by the Driver While Navigating with an In-Car Navigation Display System*, (SAE Paper 885180), Warrendale, PA: Society of Automotive Engineers.

58 Bos, T., Green, P., and Kerst, J. (1988). *How Should Instrument Panel Display Legibility By Tested?* (Technical Report UMTRI-88-35), Ann Arbor, MI: The University of Michigan Transportation Research Institute, (NTIS No. PB 90 141375/AS).

59 Green, P., Goldstein, S., Zeltner, K., and Adams, S. (1988). *Legibility of Text on Instrument Panels: A Literature Review* (Technical Report UMTRI-88-34), Ann Arbor, MI: The University of Michigan Transportation Research Institute. (NTIS No. PB 90 141342/AS).

60 Boreczky, J., Green, P., Bos, T., and Kerst, J. (1988). *Effects of Size, Location, Contrast, Illumination, and Color on the Legibility of Numeric Speedometers* (Technical Report UMTRI-88-36), Ann Arbor, MI: The University of Michigan Transportation Research Institute. (NTIS No. PB 90 150533/AS).

61 Noy, Y.I. (1987). "Theoretical Review of the Secondary Task Methodology for Evaluating Intelligent Automobile Displays," *Proceedings of the Human Factors Society-31st Annual Meeting*, Santa Monica, CA: The Human Factors Society, pp. 205-209.

62 Noy, Y.I. (1989). "Intelligent Route Guidance: Will the New Horse Be as Good as the Old?" *Vehicle Navigation and Information Systems Conference (VNIS'89)*, Warrendale, PA: Society of Automotive Engineers, pp. 49-55.

63 Noy, Y.I. (1990). *Attention and Performance While Driving with Auxiliary In-Vehicle Displays (Transport Canada Publication TP 10727 (E)), Ottawa, Ontario, Canada: Transport Canada, Traffic Safety Standards and Research, Ergonomics Division.*

64 Hardee, H.L., Johnston, C.M., Kuiper, J.W., and Thomas, W.E. (1990). "Toward a Methodology for Evaluating Instrument-Panel Controls," *Proceedings of the Human Factors Society 34th Annual Meeting*, Santa Monica, CA: Human Factors Society, pp.613-617.

65 Allen, R.W., Stein, A.C., Rosenthal, T.J., Ziedman, D., Torres, J.F., and Halati, A. (1991). "A Human Factors Simulation Investigation of Driver Route Diversion and Alternate Route Selection Using In-Vehicle Navigation Systems," *Vehicle Navigation and Information Systems Conference Proceedings (VNIS'91)*, Warrendale, PA: Society of Automotive Engineers, pp. 9-26.

66 Allen, R.W. (1991). personal communication.

67 Verwey, W.B. (1991). *Towards Guidelines for In-Car Information Management: Driver Workload in Specific Driving Situations* (Technical Report IZF 1991 C-13), Soesterberg, The Netherlands: TNO Institute for Perception.

68 Levison, W. (1993). personel communication

69 Dingus, T.A., Hulse, M.C., Krage, M.K., Szczublewski, F.E., and Berry, P. (1991). "A Usability Evaluation of Navigation and Information System "Pre-Drive" Functions," *Vehicle Navigation and Information Systems Conference (VNIS'91)*, (SAE Paper 912794), Warrendale, PA: Society of Automotive Engineers, pp. 527-536.

70 Krage, M.K. (1991). "The TravTek Driver Information System," *The 3rd International Conference on Vehicle Navigation and Information Systems (VNIS'92)* (IEEE Catalog #92CH3198-9), Piscataway, NJ: Institute of Electrical and Electronics Enigneers, pp. 739-748.

71 Rillings, J.H. and Lewis, J.W. (1991). "TravTek," *The 3rd International Conference on Vehicle Navigation and Information Systems (VNIS'92)* (IEEE Catalog #92CH3198-9), Piscataway, NJ: Institute of Electrical and Electronics Engineers, pp. 729-737.

72 Green, P., Boreczky, J., and Kim, Seung-Yun (Sylvia) (1990). "Applications of Rapid Prototyping to Control and Display Design." (SAE Paper #900470, Special Publication SP-809), Warrendale, PA: Society of Automotive Engineers.

73 Mammano, F. (1993). personal communication

74 Burgett, A.L. (1991). "Safety Evaluation of TravTek," *Vehicle Navigation and Information Systems Conference Proceedings (VNIS'91)*, Warrendale, PA: Society of Automotive Engineers, pp. 819-825.

75 Fleischman, R.N. (1991). "Research and Evaluation Plans for the TravTek IVHS Operational Field Test," *Vehicle Navigation and Information Systems Conference Proceedings (VNIS'91)*, Warrendale, PA Society of Automotive Engineers, pp. 827-837.

76 de Groot, M.T. (1992). "Rhine-Corridor: An RDS-TMC Pilot for Radio Traffic Information," *The 3rd International Conference on Vehicle Navigation and Information Systems (VNIS'92)* (IEEE Catalog #92CH3198-9), Piscataway, NJ: Institue of Electrical and Electronics Enginers, pp. 8-13.

77 Underwood, S. (1992). "Concept for Evaluation of the FAST-TRAC Program," *The 3rd International Conference of Vehicle Navigation and Information Systems (VNIS'92)*, (IEEE Catalog #92CH3198-9), Piscataway, NJ: Institute of Electrical and Electronics Engineers, pp. 626-631.

78 Bouis, D., Voss, M., Geiser, G., and Haller, R. (1979). "Visual vs. Auditory Displays for Different Tasks of a Car Driver," *Proceedings of the Human Factors Society-23rd Annual Meeting*, 35-39, Santa Monica, CA: Human Factors Society.

79 Brouwer, W.H., Ickenroth, J., Van Wolffelaar, P.C., and Ponds, R.W.H.M. (1990). "Divided Attention in Old Age: Difficulty in Integrating Skills," in R.J. Takens, et al. (eds.), *European Perspectives in Psychology*, New York, NY: Wiley.

80 Diamon, T. (1992). "Driver's Characteristics and Performances When Using In-Vehicle Navigation Systems [sic]," *The 3rd International Conference on Vehicle Navigation and Information Systems (VNIS'92)* (IEEE Catalog #92CH3198-9), Piscataway, NJ: Institute of Electrical and Electronics Engineers, 251-260.

81 Grant, B.S. and Wierwille, W.W. (1992). "An Accuracy Analysis of Techniques for Measuring the Durations of In-Car Manual Tasks," *Proceedings of the Human Factors Society 36th Annual Meeting*, Santa Monica, CA: Human Factors Society, pp. 1190-1194.

82 Korukawa, K. and Wierwille, W.W. (1990). "Validation of a Driving Simulation Facility for Instrument Panel Task Performance," *Proceedings of the Human Factors Society 34th Annual Meeting*, Santa Monica, CA: Human Factors Society, pp. 1299-1303.

83 MacAdam, C.C. (1992). *Truck Driver Informational Overload* (Technical Report UMTRI-92-36), Ann Arbor, MI: The University of Michigan Transportation Research Institute.

84 Ponds, R.W.H.M., Brouwer, W.H., and van Wolffelaar, P.C. (1988). "Age Differences in Divided Attention in Simulated Driving Task," *Journal of Gerontology*, 43(6), 151-156.

85 Verwey, W.B. (1989). "Simple In-Car Route Guidance Information from Another Perspective: Modality Versus Coding," *First Vehicle Navigation and Information Systems Conference (VNIS'89)*, Piscataway, NJ: Institute of Electrical and Electronics Engineers, pp. 56-60.

86 Walker, J., Alicandri, E., Sedney, C., and Roberts, K. (1990). *In-Vehicle Navigational Devices: Effects on Safety and Driver Performance* (Technical Report FHWA-RD-90-053), Washington, D.C.: U.S. Department of Transportation, Federal Highway Administration.

87 Walker, J., Alicandri, E., Sedney, C., and Roberts, K. (1991). In-Vehicle Navigational Devices: Effects on Safety and Driver Performance, *Vehicle Navigation and Information Systems Conference Proceedings (VNIS'91)*, Warrendale, PA: Society of Automotive Engineers, 499-526.

88 Walker, J., Alicandri, E., Sedney, C., and Roberts, K. (1992). "In-Vehicle Navigational Devices: Effects on Safety and Driver Performance," *Public Roads*, 56(1), 9-22.

89 Alm, H. (1990). "Drivers' Cognitive Models of Routes," in Van Winsum, W., Alm, H., Schraagen, J.M., and Rothengatter, T., *Laboratory and Field Studies on Route Representation and Driver's Cognitive Models of Routes* (Deliverable GIDS/NAV2, DRIVE Project V1041), Haren, The Netherlands: Traffic Research Centre, University of Groningen.

90 Alm (undated). Drivers' Cognitive Models of Routes (unpublished manuscript), Linkoeping, Sweden: Swedish Traffic Institute.

91 Alm, H. and Berlin, M. (undated) What is the Optimal Amount of Information from a Verbally-Based Navigation System? (unpublished manuscript), Linkoeping, Sweden: Swedish Traffic Institute.

92 Alm, H., Nilsson, L., Jarmark, S., Savelid, J., and Hennings, U. (1992). *The Effects of Landmark Presentation on Driver Performance and Uncertainty in a Navigation Task—A Field Study*, (Report No. 692A), Linkoeping, Sweden: Statens Vaeg-och Trafikinstitut.

93 Davis, J.R. (1989). *Back Seat Driver: Voice Assisted Automobile Navigation* (unpublished Ph.D. dissertation), Cambridge, MA: Massachusetts Institute of Technology.

94 Davis, J.R. (1989). "The Back Seat Driver: Real Time Spoken Driving Directions," *First Vehicle Navigation and Information Systems Conference (VNIS'89)*, Piscataway, NJ: Institute of Electrical and Electronics Engineers, pp. 146-150.

95 Eberhard, J.W. (1968). Driver Information Requirements and Acceptance Criteria, H*ighway Research Record* #265, Washington, D.C.: National Research Council, Highway Research Board, 19-30.

96 Hook, K. (1991). *An Approach to Route Guidance* (Licentiate thesis), (Technical Report 91-019-DSV), Stockholm, Sweden: Stockholm University, Department of Computer and Systems Sciences.

97 Hook, K. and Karlgren, J. (1991). *Some Principles for Route Descriptions Derived from Human Advisors* (SICS Technical Report R91:06), Stockholm, Sweden: Swedish Institute of Computer Science.

98 Brown, C., Gustavsson, R., Hook, K., Lindevall, P., and Waern, A. (1991). *Final Report on Interactive Route Guidance* (SICS Technical Report T91:18), Kista, Sweden: Swedish Institute of Computer Science, Knowledge Based Systems Laboratory.

99 Waern, A. (1992). *Planning and Reasoning for Human-Machine Interaction in Route Guidance* (licentiate thesis) (Technical Report 92-0006-DSV), Stockholm, Sweden: Stockholm University, Department of Computer and Systems Sciences.

100 Kuiken, M.J., Miltenburg, G.M., and van Winsum, W. (1992). "Drivers' Reactions to an Intelligent Driver Support System (GIDS) Implemented in a Driving Simulator," *The 3rd International Conference on Vehicle Navigation and Information Systems (VNIS'92)* (IEEE Catalog #92CH3198-9), Piscataway, NJ: Institute of Electrical and Electronics Engineers, pp. 176-181.

101 Labiale, G. (1989). Influence of In-Car Navigation Map Displays on Drivers Performances (SAE Paper 891683), Warrendale, PA: Society of Automotive Engineers.

102 Labiale, G. (1990). "In-Car Road Information: Comparisons of Auditory and Visual Presentations," *Proceedings of the Human Factors Society 34th Annual Meeting*, Santa Monica, CA: Human Factors Society, pp. 623-627.

103 Labiale, G., Mamberti, M.L., Baez, D., Conus, Y., and Aupetit, J. (1988). *Comprehension et Memorisation des Messages Visuels et Auditifs D'Information Routiere (Comprehension and Retention of Visual and Auditory Road Safety Messages)* (INRETS Report 66), Bron, France: Laboratoire Energie Nuisances, Institut National de Recherche sur Ies Transports et leur Securite.

104 Labiale, G. (undated). Ergonomics of In-Car Road Information System: RDS Inf-Flux, unknown source, 1489-1491.

105 Schraagen, J.M.C. (1990). "Use of Different Types of Map Information for Route Following in Unfamiliar Cities," in van Winsum, W., Alm, H., Schraagen, J.M. and Rothengarter, T., *Laboratory and Field Studies on Route Representation and Drivers' Cognitive Models of Routes* (Deliverable GIDS/NAV2), Haren, The Netherlands: University of Groningen, Traffic Research Center.

106 Sperandio, J.C. and Dessaigne, M.F. (1988). "Une Comparaison Experimentale Entre Model de Presentation Visuels ou Auditifs de Messages D'Informations Routieres a des Conducteurs Automobiles (An Experimental Comparison of Visual and Auditory Messages Displayed to Car Drivers)," *Le Travail Humain*, 51(3), 257-269.

107 Antin, J.F. (1987). *Automotive Navigation Methods: Effectiveness, Efficiency, and Strategy*, (unpublished doctoral dissertation), Blacksburg, VA: Virginia Polytechnic Institute and State University.

108 Labiale, G. (1992). "Driver Characteristics and In-Car Map Display Memory Recall Performance," *The 3rd International Conference on Vehicle Navigation and Information Systems (VNIS'92)* (IEEE Catalog #92CH3198-9), Piscataway, NJ: Institute of Electrical and Electronics Engineers, pp. 227-232.

109 McKnight, A.J. and McKnight A.S. (1992). *Effect of In-Vehicle Navigation Information Systems Upon Driver Attention* (Technical Report), Washington, D.C.: AAA Foundation for Traffic Safety.

110 Morita, H. and Ogawa, M. (1992). "Experimental Studies for Toyota Voice Route Guidance Using Beacon," *Surface Transportation and the Information Age: Proceedings of the IVHS America 1992 Annual Meeting*, Washington, D.C.: IVHS America, 614-622.

111 Pauzie, A. and Marin-Lamellet, C. (1989). "Analysis of Aging Drivers Behaviors Navigating with In-Vehicle Visual Display Systems," *Vehicle Navigation and Information Systems (VNIS'89)*, Piscataway, NJ: Institute of Electrical and Electronics Engineers, pp. 61-67.

112 Popp, M.M. and Farber, B. (1991). "Advanced Display Technologies, Route Guidance Systems, and the Position of Displays in Cars," *Vision in Vehicles III* (in Gale, A.G., et al.), Amsterdam, The Netherlands: Elsevier.

113 Rothery, R.W., Thompson, R.R., and von Buseck, C.R. (1968). *A Design for an Experimental Route Guidance System, Volume III, Driver Display Experimental Evaluation* (Technical Report GMR-815-III), Warren, MI: General Motors Research Laboratories.

114 Staal, D. (1987). ETAK Navigational System Evaluation in Los Angeles and San Francisco - Final Results (letter to J. Sullivan of National Car Rental).

115 Streeter, L.A., Vitello, D., and Wonsiewicz, S.A. (1985). "How to Tell People Where to Go: Comparing Navigational Aids," *International Journal of Man-Machine Studies*, 22, 549-562.

116 Verwey, W.B. and Janssen, W.H. (1988). "Driving Behavior with Electronic In-Car Navigation Aids," *Proceedings of Road Safety in Europe* (VTI Report 342A), Linkoping, Sweden: Swedish Road and Traffic Safety Research Institute, pp. 81-97.

117 Voss, M. and Haller, R. (1982). "Processing of On-Board Route Guidance Information in Real Traffic Situations," *Analysis, Design, and Evaluation of Man-Machine Systems*, Oxford, United Kingdom: Pergammon, 371-375.

118 West, R., Kemp, R., and Hack, S. (1989). *Autoguide System Proving and Usability Trials* (Contractor Report 181), Crowthorne, Berkshire: United Kingdom: Transport and Road Research Laboratory, Department of Transport.

119 Sweet, R.E. and Green, P. (1993). "UMTRI's instrumented car," *UMTRI Research Review*, January-February, 1-11.

120 Green, P. (1992). *Review of Eye Fixation Recording Methods and Equipment* (Technical Report UMTRI-92-28), Ann Arbor, MI: University of Michigan Transportation Research Institute.