U.S. Department
of Transportation
National Highway
Traffic Safety
Administration

NHTSA
www.nhtsa.gov

DOT HS 811 547A

May 2013

Distraction Detection and Mitigation Through Driver Feedback

DISCLAIMER

Suggested APA Format Reference:

Lee, J. D., Moeckli, J., Brown, T. L., Roberts, S. C., Schwarz, C., Yekhshatyan, L., Nadler, E., Liang, Y., Victor, T., Marshall, D., & Davis, C. (2013, May). *Distraction Detection and Mitigation Through Driver Feedback.* (Report No. DOT HS 811 547A). Washington, DC: National Highway Traffic Safety Administration.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT HS 811 547A		
4. Title and Subtitle Distraction Detection and Mitigation Through Driver Feedback		5. Report Date May 2013 6. Performing Organization Code
7. Author(s) John D. Lee, Jane Moeckli, Timothy L. Brown, Shannon C. Roberts, Chris Schwarz, Lora Yekhshatyan, Eric Nadler, Yulan Liang, Trent Victor, Dawn Marshall, Claire Davis		8. Performing Organization Report No. N11-002
9. Performing Organization Name and Address National Advanced Driving Simulator, The University of Iowa 2401 Oakdale Blvd Iowa City, IA 52242		10. Work Unit No. (TRAIS) 11. Contract or Grant No. DTNH22-06-D-00043
12. Sponsoring Agency Name and Address National Highway Traffic Safety Administration 1200 New Jersey Avenue, SE, West Building Washington, DC 20590		13. Type of Report and Period Covered Final Report, September 5, 2008 – April 4, 2011 14. Sponsoring Agency Code
15. Supplementary Notes Eric Traube and David Band (Task Order Managers); Stephanie C. Binder (COTR)		

16. Abstract

Despite government efforts to regulate distracted driving, distraction-related fatalities and injuries continue to increase. Manufacturers are introducing real-time driver monitoring systems that detect risk from distracted driving and warn drivers; however, little is known about these systems. This report identifies evaluation techniques to characterize and assess these emerging technologies, presents results of their application, develops a framework for estimating systems' safety benefits, and provides safety relevant information to guide technology development. A standardized language for describing and differentiating systems was created, and its application revealed key trends in the design landscape. A novel approach to detection that provides prospective indications of safety-critical vehicle state changes is described. Two evaluation protocols were developed and to provide empirical assessments of (1) detection algorithm performance and (2) the effect of mitigations on driver performance and acceptance. The protocol included driving on different types of roadways and performing secondary tasks in the high-fidelity NADS-1 driving simulator. Four progressively complex distraction detection algorithms were compared to evaluate the ability of vehicle-based systems to distinguish between distracted and non-distracted drivers. Algorithm performance varied across road types and distraction tasks. A safety benefits framework appropriate for distraction mitigation systems is proposed that expands on past benefit analyses.

17. Key Words	18. Distribution Statement
Driver Distraction, Real-time Driver Monitoring, Driver Feedback, Evaluation Protocol, System Specification, Acceptance, Behavior, Performance, Safety Benefits, National Advanced Driving Simulator	Document is available to the public from the National Technical Information Service www.ntis.gov

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified		166	

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

EXECUTIVE SUMMARY

Distraction represents an important driving safety problem and has received substantial attention recently. This attention reflects concern about a dramatic escalation of diversity and ubiquity of technology both built in and carried in to vehicles. The focus of this study is on the recent trend of using vehicle-based technology to combat distraction. It developed and assessed real-time distraction detection and mitigation systems to (1) guide technology development to enhance driver safety, and (2) identify potential evaluation techniques to characterize and assess this emerging technology. The more notable results of this study include:

- No single sensor, algorithm or countermeasure can be judged "optimal" for detecting or mitigating distraction because an algorithm may support one countermeasure well, while not supporting others; likewise a type of countermeasure may be more effective with some algorithms than with others. (Chapter 2)
- An approach was developed to describe detection and mitigation systems and organize the diversity of system designs. (Chapter 3)
- A protocol for assessing the ability of algorithms to identify different types of distraction was developed and its sensitivity was demonstrated. The sensitivity of the National Advanced Driving Simulator (NADS-1) for detecting the effects of distraction was validated. (Chapter 4)
- The ability to detect the effects of feedback on driver performance was demonstrated.
- Needs for future protocol refinements were identified. (Chapter 5)

Background and Objectives

According to NHTSA's *Distracted Driving 2009 Traffic Safety Facts*, driver distraction contributed to an estimated 16 percent of fatal crashes and 20 percent of injury crashes. Automobile and aftermarket manufacturers have begun introducing systems to reduce distraction-related crashes. These devices use advances in sensor technologies and algorithms to detect risk and warn drivers. Countermeasures take several forms. One form immediately redirects drivers' attention to the roadway. A second form is focused on preventing future distraction – it shifts driver attitudes and willingness to engage in distracting activities by providing them with feedback concerning the effect of distraction on their driving. These distraction detection and mitigation[1] systems vary significantly in their purpose, operation, capabilities and features, making them difficult to understand, evaluate and compare. Little is known about many aspects of their operation, effectiveness, and acceptability.

---

[1] Mitigation systems include adjusting collision avoidance systems such as forward collision warnings when they detect distraction, so some systems emphasize feedback more than others.

This report addresses these critical gaps by developing systematic methodologies for describing and evaluating systems, including several novel approaches to detect distraction and provide feedback.

This project has four objectives:

- Develop a standard *specification template* to describe distraction detection and mitigation systems;
- Develop a standard set of *performance metrics* embedded in a protocol for assessing distraction detection and distraction mitigation effectiveness;
- Provide a safety benefits framework for estimating the overall effect on driving safety; and
- Develop alternative distraction detection and distraction mitigation *design concepts.*

This research was sponsored by NHTSA as part of it overall program for the prevention of road traffic crashes for which driver distraction is a contributing factor.[2] In April 2010, NHTSA released *Overview of the National Highway Traffic Safety Administration's Driver Distraction Program*,[3] which summarized steps that NHTSA intends to take to "help in its long-term goal of eliminating a specific category of crashes- those attributable to driver distraction." NHTSA's Driver Distraction Program consists of four initiatives:

- Improve the understanding of the extent and nature of the distraction problem. This includes improving the quality of data NHTSA collects about distraction-related crashes along with better analysis techniques.
- Reduce the driver workload associated with performing tasks using both built-in and portable in-vehicle devices by limiting the visual and manual demand associated with in-vehicle device interface designs. Better device interfaces will help to minimize the amount of time and effort involved in a driver performing a task using the device. Minimizing the workload associated with performing non-driving, or "*secondary,*" tasks with a device will permit the driver to maximize the attention they focus toward the *primary* task of driving.
- Keep drivers safe through the introduction of crash avoidance technologies. These include the use of crash warning systems to re-focus the attention of distracted drivers as well as vehicle initiated (i.e., automatic) braking and steering to prevent or mitigate distracted driver crashes.
- Educate drivers about the risks and consequences of distracted driving. This includes targeted media messages, drafting and publishing sample text messaging laws for

---

[2] Information on NHTSA's efforts to address this problem can be found at www.distraction.gov/.
[3] NHTSA. (2010, April). Overview of the National Highway Traffic Safety Administration's Driver Distraction Program.(Report No. DOT HS 811 299) Washington, DC: Author. Available at www.nhtsa.gov/staticfiles/nti/distracted_driving/pdf/811299.pdf

consideration and possible use by the states, and publishing guidance for a ban on text messaging by Federal government employees while driving.

This work is part of the third initiative, keep drivers safe through the introduction of crash avoidance technologies. As stated above, automobile and aftermarket manufacturers have begun introducing systems to reduce distraction-related crashes. This research will help NHTSA promote systems that effectively reduce distraction-related crashes.

Specification Templates

A systematic approach to describe and compare distraction detection and mitigation systems was needed to facilitate assessments of their purpose and efficacy. Chapter 3 provides a template-based common language for describing and differentiating the type and quality of data that distraction detection systems need as input, what they produce as output, and the associated mitigation strategies. This approach identifies salient commonalities and differences between systems in terms of the system's purpose and functions without requiring more detailed proprietary information. The template was applied to current production and research systems, as well as to select algorithms and mitigations evaluated in Chapters 4 and 5.

Despite a lack of available (public) information from which to perform detailed analytical evaluations, the template descriptions provided useful distinctions between systems that differ in the type of distraction detected and mitigated, and intent and timescale of feedback (i.e., whether the feedback was intended to address current distraction or future distraction). Four key areas emerged that suggest important directions for future research and development. First, most systems do not distinguish between impairments (e.g., drowsiness, distraction), or types of distraction. Second, most systems use a small number inputs to detect distraction, and none of the existing systems combine data to identify both cognitive and visual distraction. Third, distraction detection algorithms can be described in the context of the countermeasure they support, which is important because countermeasure efficacy, driver acceptance, and the ultimate safety benefit depend on the match between the algorithm characteristics and the countermeasure. Fourth, most systems provide real-time feedback for immediate driving performance improvement. However, this approach may impose more workload on a driver in addition to the already highly demanding distracted driving situation.

Performance Metrics

Chapter 4 describes and applies a protocol consisting of evaluation methods, and measures, to compare the ability of vehicle-based systems to detect distraction. Chapter 5, assuming a common detection algorithm, describes the protocol adapted to evaluate a system's ability to mitigate distraction through driver feedback. The protocols provide systematic assessments of system efficacy, how the systems affect drivers, including driver acceptance.

Distraction detection assessment protocol

A protocol was developed and applied to candidate algorithms to evaluate their ability to distinguish between distracted and non-distracted drivers, and to identify the most promising (Chapter 4). The protocol consists of a data collection process that samples a selection of drivers, driving situations, and representative distractions designed to challenge the algorithms in a variety of ways to reveal their capabilities and vulnerabilities. A high-fidelity motion-based driving simulator, called NADS-1, equipped with separate eye-tracking and head-tracking systems, was used to collect data during baseline and distracted drives.

Data was collected from 32 drivers in a drive representative of a nighttime trip home from an urban entertainment district. Audio prompts were given to begin a series of three prompted distraction tasks (a "bug" task that required turning and reaching to follow a bug on a touch-screen display, an arrows identification task, and a voice-activated flight menu task). The arrows task required the driver to scan a display located to the right of the steering wheel and identify a target, whereas the flight menu task did not require the driver to divert his or her gaze from the road. These three tasks occurred during each of eight situations distributed across three driving environments or "drive segments": a two-lane urban segment; a four-lane divided interstate segment; and a two-lane rural highway with curves and gravel. Participants could delay initiating these tasks within each segment or to the end of the segment when the driver came to a stop in the drive. A self-paced distraction task (radio-tuning task) occurred in between the prompted distraction tasks. Participants were instructed to drive normally but that the tasks were urgent and task performance scores were provided in between driving environments and at the end of the drive.

The protocol was applied to four progressively more complex algorithms, all based on gaze measures. The least complex detected distraction when the driver's eyes were off of the road for more than 2 seconds in any 6-second interval. Complexity was added, for example, by including glances toward the mirrors as distinct from other glances away from the road. The most complex algorithm provided different outputs for visual and cognitive distraction, and allowed a degree of visual timesharing to occur while driving. Data was interpreted relative to evaluation criteria from signal detection theory to assess the algorithms' robustness across different distraction tasks and road segments.

Across all driving environments and both visual distraction tasks (bug and arrows), the most complex algorithm consistently performed better than the others. It performed best in detecting distraction during the arrows task. It was the only algorithm of the four capable of detecting cognitive distraction from the flight menu task, and it did so imprecisely, although at a rate substantially greater than chance. However, the least complex algorithm performed best in the urban environment, possibly because the most complex algorithm was inactive at low speeds, and it performed as well as the most complex algorithm in detecting distraction from the bug task. It yielded a high true positive rate, but also many false alarms when the driver performed the arrows task. Nonetheless, the protocol found that all of the algorithms succeeded in detecting visual distraction well above chance. It also demonstrated that the tradeoff between ensuring distraction detection and avoiding false alarms complicates the identification of the most promising algorithm for detecting distraction. High false alarm rates would likely undermine drivers' acceptance of a system that presents real-time feedback, but may result in less annoyance with a post-drive feedback system.

Distraction mitigation assessment protocol

A protocol was developed to assess the effect of alternative distraction countermeasures on driving performance, visual behavior, and attitudes toward distracted driving. This evaluation protocol was assessed by applying it to real-time visual feedback (flashing lights) presented on the left, center, and right side of the windshield with a heads-up display (HUD) and synchronized auditory alerts, and post-drive feedback comprised of a report card that described and evaluated the participant's performance and included a video playback of driving errors that occurred while he or she was distracted.

Thirty-six participants completed drives to collect baseline distracted driving data and distracted driving with feedback. One third of the participants experienced post-drive feedback at the end of their distraction drive and following each segment of their mitigation drive, one third experienced real-time feedback during their mitigation drive, and the remaining third served as a control condition and received no feedback. The primary dependent measures assessing the effects of feedback on driver performance and visual behavior included the duration of task engagement, lateral and longitudinal control, eye movements, and subjective data to assess the effect of the mitigations on drivers' performance, their awareness of distraction, vehicle control, and distraction task performance, and their willingness to engage in distractions while driving in the future.

The two mitigation approaches resulted in subtle but distinct differences in driver response. The post-drive feedback drivers (1) delayed engaging in the most intensive visual task (bug task) but did not then begin to perform it during a part of the drive that imposed lower demand, (2) increased attention to the roadway (shorter glances away) while engaged in visual distraction, and (3) improved lane keeping during the bug task and in the most visually demanding driving environments. However, it resulted in degraded lane keeping for two less visually demanding tasks (flight menu and radio). The real-time feedback decreased drivers' focus on the roadway and improved lane keeping only during the most challenging visual distraction and the most demanding driving environment. Overall, feedback provided a benefit in some cases (higher demand tasks and environments) but decreased performance in others. Additional work is needed to understand why this is the case. It was not possible to evaluate the effect of feedback on participants' awareness because of pre-existing differences among the groups that received the real-time, post-drive, or no feedback. Also, feedback received during the mitigation drive did not affect the intention to engage in distracting activities.

In summary, the protocol was able to show that a very limited exposure to post-drive feedback resulted in changes in engagement with the distraction tasks. The protocol also detected complex relationships between the mitigation systems, the tasks, and driving environments. The results point to some changes that could make a distraction feedback protocol more effective. These include changing the timing and structure of the distraction tasks to allow more flexibility in their engagement, and strategically locating high and low demand environments to make the protocol more sensitive to decisions to delay engagement in distracting tasks until a low demand environment is found. Further development of instruments to assess planned behavior and willingness to engage in tasks while driving is also recommended.

Safety Benefits Framework

The degree to which technology designed to mitigate distraction succeeds depends on its ability to reduce crashes and associated deaths, injuries, and property damage. Existing benefits analyses are insufficient to fully address this problem. A framework and method to estimate these benefits is proposed in Chapter 6.

The benefits of mitigation systems can accrue by discouraging drivers from *enabling* distracting devices, *engaging* in distracting activities, and *persisting* in distracting activities when distractions put them in crash-imminent situations. The specific steps associated with estimating a system's benefit include: identifying a representative sample of distraction-related crashes (e.g., from naturalistic driving studies); describing the mitigation system in sufficient detail to support an estimate of its ability to prevent crashes; defining the crash configuration in terms that describe the timing of the mitigation-triggering events relative to the time available to respond when an alert would occur and when the crash occurred; and estimating driver response time through models of driver attention in the seconds preceding a crash and drivers' strategic decisions regarding engagement in distracting activities and use of feedback. Probability of collision would estimate the contribution of the driver's response when given distraction feedback to overall system effectiveness. A critical challenge concerns how to incorporate the longer term effects of feedback on driver behavior and shifts in societal norms.

Implications and Next Steps

This study comprises one of NHTSA's vehicle-based initiatives to keep drivers safe through the use of distraction monitoring and warning systems. As vehicle manufacturers deploy first generation distraction systems based on real-time driver state monitoring, it provides a pair of assessment protocols to evaluate and compare the benefits of distraction systems. Consistent with the systems reviewed, his study did not attempt to distinguish distracted driving from other forms of impaired driving such as alcohol-impaired and drowsy driving. However, it identifies potential evaluation techniques to characterize and assess this emerging technology, provides a novel approach to detecting distraction, and provides recommendations for future protocol development.

As argued in Chapter 1, it is important to evaluate not only risky behavior and risky outcomes, but also how these systems affect driver understanding of risky behavior. Using this paradigm for evaluation, refinements of the protocol, additional development needs and additional areas of research are necessary to finalize an effective overall assessment of distraction mitigation systems.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ACT-R	Adaptive Control of Thought-Rational
ADAS	Advanced Driver Assistance Systems
AIDE	Adaptive Integrated Driver-Vehicle InterfacE
AUC	Area Under the Curve
CAMP	Crash Avoidance Metrics Partnership
CWS	Crash Warning System
DAC	Driver Alert Control
DARPA	Defense Advanced Research Projects Agency
DRIVE	Dedicated Road Information Infrastructure for Vehicle Safety in Europe
ECG	Electrocardiogram
EEG	Electroencephalogram
EOFR	Eyes Off Forward Roadway
EWMA	Exponential Weighted Moving Average
FARS	Fatality Analysis Reporting System
fNIR	Functional Near Infrared
FPR	False Positive Rate
FOT	Field Operational Test
FRD	Field Relevant for Driving
FSR	Force Sensing Registors
GLM	General Linear Models
GPS	Global Positioning System
GSR	Galvanic Skin Response
HASTE	Human Machine Interface and the Safety of Traffic in Europe

HVAC	Heating, Ventilating, and Air Conditioning
HUD	Head-Up Display
IRB	Institutional Review Board
IVBSS	Integrated Vehicle-Based Safety Systems
IVIS	In-Vehicle Information System
LED	Light Emitting Diode
MaxOR	Maximum Odds Ratio
MDD	Multidistraction Detection
MRT	Multiple Resource Theory
NADS	National Advanced Driving Simulator
NCAP	New Car Assessment Program
NHTSA	National Highway Traffic Safety Administration
OEM	Original Equipment Manufacturer
PA	Pilot's Associate
PPV	Positive Predictive Value
PRC	Percent Road Centre
PVI	Pilot-Vehicle Interface
RMS	Root Mean Square
ROC	Receiver Operator Characteristic
RVSP	Risky visual scanning patterns
SAVE-IT	Safety Vehicle Using Adaptive Interface Technology
SD	Standard Deviation
SEEV	Salience, Effort, Expectancy, and Value
SeMiFOT	Sweden Michigan Naturalistic Field Operational Test
SHRP-2	Strategic Highway Research Program-2

TLC	Time-to-Line-Crossing
TPR	True Positive Rate
VTS	Visual Time Sharing
WCT	Warship Commander Task

CHAPTER 1. INTRODUCTION

Driver distraction is occurring with greater frequency as in-vehicle technology and carried-in devices become increasingly common and complicated (McEvoy et al., 2006; Utter, 2001, 2009). Consequently, distraction and inattention contribute to crash risk and are likely to have an increasing influence on driving safety. Analysis of police-reported crash data from 2008 shows that distractions account for 5,870 fatalities and an estimated 515,000 injuries (Pickrell & Ye, 2009). The challenges of detecting distractions at the crash site and reluctance of drivers to admit to being distracted make it likely that these statistics underestimate the magnitude of the problem. A recent naturalistic driving study found that distraction and inattention contribute to approximately 80 percent of crashes and that distraction contributes to approximately 65 percent of rear-end crashes (Klauer et al., 2006). The extent to which this generalizes from the small number of crashes that were observed in this study to the overall population of crashes remains a topic for debated, but there is cause for concern even if the contribution is a fraction of that observed in this study.

The rapid advances in wireless, computer, and sensor technology will confront drivers with a range of new distractions. Not only do drivers need to manage use of cell phones, CD players and navigation systems, they are increasingly faced with long text message "conversations" and searches through MP3 music catalogs that can extend beyond 30 seconds (Salvucci, 2007) and involve more than 15 glances (Chisholm et al., 2007). In the coming years, drivers may also be increasingly tempted to retrieve a broad variety of information from the Internet via hand held phones as well as through dedicated connections in the vehicle itself. Rapid changes in vehicle design illustrate this trend: 90 percent of all new vehicles are compatible with MP3 players (http://www.apple.com/ipod/car-integration/), all 2009 Chrysler vehicles have a wireless connection to the Internet (Bensinger, 2008), and several manufacturers introduced sophisticated Internet-enabled computers in vehicle consoles in 2010 (Vance & Richtel, 2010). These devices have the potential to make driving more enjoyable, and efficient, and may even mitigate drowsiness. Yet, they also have the potential to distract drivers. Helping drivers benefit from these devices and avoid distraction-related crashes represents an important challenge.

Although efforts are afoot at state and federal levels to regulate the use of certain devices, such as hand held cell phones, or distracting behaviors, such as the federal ban on texting by commercial truck and bus drivers, such legislation will likely lag behind the fast pace to technological change that is responsible for many distractions. A complementary approach that uses technology to detect and mitigate dangerous episodes of distraction, such as warnings based on long and frequent glances to an in-vehicle device, also has great promise in reducing the frequency and severity of distraction-related crashes (Donmez et al., 2008c). Such technological mitigations have been hampered by limitations of sensors and algorithms, but the increasing availability of sensor and computing technology have made more sophisticated systems possible. This report describes and evaluates approaches for detecting and mitigating distraction.

DISTRACTION AND ATTENTION IN DRIVING

Drivers are required to respond to a variety of changes in the driving environment and surrounding traffic relying predominantly on visual inputs to detect events ranging from braking lead vehicles, signs, and sharp curves, to illegal turns of oncoming vehicles and incline changes. Drivers also engage in a variety of secondary activities that require glances away from the road as well as higher cognitive functions such as planning and decision making. The concept of attention is of central importance in understanding how secondary tasks distract drivers.

The perceptual system presents a very powerful illusion of a stable, full detail, pictorial external world and gives us a strong subjective impression of seeing everything at all times. Thus, it is often forgotten that eyes are not like cameras that deliver a uniformly detailed, uninterrupted picture of the world. In fact, drivers can only attend to a relatively narrow slice of the driving environment at any time.

Attention can be defined as the selection and processing of certain information from the array of information available from the environment (James, 1890). It follows that inappropriate selection and inadequate processing of the information that is relevant for safe driving constitutes driver inattention. Inattention has a broader definition than distraction and includes all conditions in which a driver fails to focus on information critical for safe driving, whether it is due to competing tasks, drowsiness, or other cognitive impairments. Driver distraction refers to a particular kind of driver inattention: the inappropriate selection of information such that safety-relevant information is neglected. Therefore, distraction is a relational property in that it reflects inadequate attention to the road relative to roadway demands and an inability to shift attention to the road when these demands require it. Distraction is defined as:

> "…the diversion of attention away from activities critical for safe driving toward a competing activity." (Lee et al., 2008a, p. 34)

In this definition, the competing activity distinguishes distraction from other cognitive states that might diminish a driver's ability to drive safely, such as drowsiness, anger, or alcohol-related impairment. These cognitive states may accompany and amplify distractions, as in the case of anger amplifying the distraction associated with contemplating a recent argument with a spouse. Likewise, alcohol impairment may amplify distraction by diminishing the driver's capacity to shift attention to and respond to the roadway when demands arise (Rakauskas et al., 2008). Distraction alone, or in combination with various cognitive states, poses an important challenge to driving safety.

EVALUATING TECHNOLOGY THAT ENHANCES DRIVERS' ATTENTION TO THE ROAD

One approach to addressing the challenge of distraction is detecting and mitigating distraction through technological interventions. This general approach can be implemented in many ways, and the success or failure of each intervention hinges on a range of factors, including the driving environment, road conditions, driver characteristics, sensor sensitivity, and the underlying theoretical assumptions of the algorithm that integrates the sensor data. Differences in the physical and computational configurations of distraction detection systems lead to profiles of detection that differ in content, accuracy, and reliability. Approaches to distraction mitigation present an array of options regarding feedback content, feedback timing, and feedback representation. How these elements of distraction detection and mitigation and their many possible combinations are implemented opens the door to a proliferation of potential systems. Currently, there is no standard way of describing these systems or their elements, measuring their performance, or estimating their overall safety benefit. This lack of standardization has negative implications for government agencies, manufacturers, and consumers who will be confronted with a new generation of safety systems that are difficult to understand, compare, and regulate.

The aim of this research is to fill this gap by creating a standard way of describing distraction detection and mitigation systems (specification templates), developing a standard way of assessing system performance (assessment protocols and metrics), and providing a framework for estimating the overall effect on driving safety (safety benefits framework). The specification templates, assessment protocols, and assessment metrics will be demonstrated by applying them to novel ways of detecting distraction and providing distraction-related feedback (design concepts). Each of these outcomes helps to guide technology development so that it enhances driving safety in a way that is acceptable to drivers.

This project is based on a layered approach to distraction detection and mitigation system evaluations. Figure 1 shows the nested constraints that affect the performance of the distraction detection and mitigation systems at different scales: to affect culture change a system must first have appropriate and sufficient sensor signals, a robust algorithm, and relevant and acceptable countermeasures that change drivers' performance and behavior. Each layer has its own methodological approach, set of evaluation criteria, and protocol requirements, from part-task simulator evaluations to assess sensors to naturalistic field operations tests to assess cultural change. The associated time scale varies considerably from a few weeks to conduct sensor evaluations to years to conduct and analyze naturalistic driving data.

There is the potential to extrapolate an approach from one level and apply it to another level. Partial assessments of culture change could use driving simulation technology, and qualitative methodologies could provide insight to changes in driving performance and behavior. While the simulator-based evaluation protocols[4] presented in this report primarily assess algorithm and driving performance, measurements of driver attitude are also employed to assess whether short exposure to a system in a driving simulator results in measurable shifts in driver attitude that ultimately may be reflected in changes in driver behavior and associated crash risk.

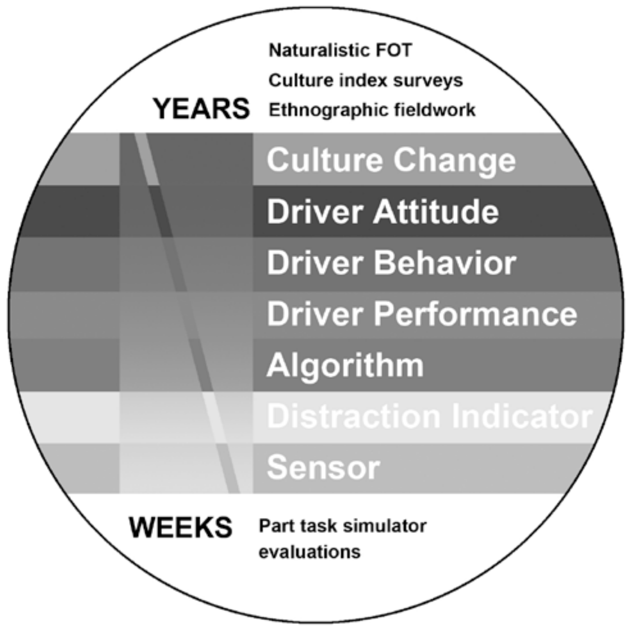


Figure 1 Layers of evaluation to assess the effectiveness of distraction detection and mitigation systems span evaluations of the sensor to the cultural impact of safety systems.

The following chapter describes the mechanisms of driver distraction and develops a theoretical framework that can help identify variables for detecting driver distraction. This chapter also

---

[4] The choice of simulation as a tool for these evaluations was made to provide a level of experimental control and safety that is not feasible in on-road and test track studies. The primary advantages of simulation over other methods for this effort include precise control of the experimental conditions, ease of implantation of distraction tasks, and a safe environment in which to engage in demanding distraction tasks. The trade-offs include a lower face validity, which we attempted to mitigate through use of the full motion NADS-1 which features an actual vehicle cab, and the potential for drivers to be more willing to engage in tasks if they perceive lower risk in the simulator. This greater willing ness to engage in non-driving tasks provides a benefit for evaluating effects of distraction, but potentially could lead to an underestimation of the effects on systems designed to mitigate the effects of distraction on driving.

outlines general approaches to mitigating distraction, which should be considered in selecting variables and developing distraction detection algorithms. The subsequent chapters provide templates for describing distraction detection and mitigation systems as well as demonstrations of evaluation protocols applied to novel distraction detection and mitigation systems. The final chapter concludes with a framework for estimating the benefits associated with distraction detection and mitigation systems.

CHAPTER 2. THEORETICAL AND EMPIRICAL CONSIDERATIONS FOR DISTRACTION DETECTION AND MITIGATION

Effective distraction detection and mitigation systems depend on a range of considerations. These range from the theoretical—how distraction is conceptualized—to the practical—the capabilities of particular sensors, such as those that provide eye tracking or lane position information. The value of these sensors depends on how sensitive they are to the behavioral signature of distraction relative to practical considerations such as cost and reliability. For example, pupil response is a sensitive measure of cognitive effort, but it is unlikely that any eye-tracking system for a production vehicle will have the precision needed to capture this information at a cost that is practical for manufacturers and consumers in the near future. Appendix A describes a range of distraction countermeasures and considerations for distraction detection sensors and algorithms. The effectiveness of systems in detecting and mitigating distraction depends on their robustness in functioning under a variety of roadway conditions, how they relate to the underlying mechanisms of distraction, and to the available sensor and algorithm technology. The specific topics considered in detail in Appendix A include:

- Measures of driver distraction and sensor technology tradeoffs;
- General types of countermeasures; and
- Countermeasure considerations in sensor selection.

Central to developing distraction detection and mitigation strategies concerns how it is conceptualized. A common perspective regarding distraction is that the driver is an information processing system, and that processing more than one stream of information at once compromises the response to one or more streams of information (i.e., dual-task interference). This information processing perspective captures important cognitive constraints regarding the consequences of diverting attention from critical driving tasks. However, such a perspective does not consider how and when drivers choose to engage in distracting tasks, but rather it considers drivers as passive recipients of tasks. The information processing perspective assumes that drivers act in response to the demands of driving and competing activities rather than actively manage these demands. Considering drivers as active controllers of these demands provides a more complete account of the mechanisms responsible for distraction and strategies for its mitigation.

Driving, as a control process, has been described in terms of three levels: operational, tactical, and strategic (Michon, 1985; Ranney, 1994; Sheridan, 1970). The operational level concerns the lateral and longitudinal control of the vehicle and occurs at a timescale of milliseconds to seconds. Tactical control concerns the choice of lanes and speed, and occurs at a timescale of seconds to minutes. Strategic control concerns decisions regarding routes and travel patterns and occurs at a timescale of minutes to weeks. Distraction can emerge from any of these three levels of control when competing activities interfere with activities critical to safe driving (Lee et al., 2008b; Lee & Strayer, 2004). At the operational level drivers control resource investment; at the tactical level they control task timing; and at the strategic level they control exposure to potentially demanding situations (Lee, 2010).

Distraction-related mishaps result from a breakdown of control at any one level, and from the accumulation of control problems that compound as they propagate across levels. Distraction-related crashes result not only from dual-task interference, but also from drivers' failure to manage distractions by either delaying or interrupting competing activities to maintain attention to the road. Considering driving as a multilevel control process identifies mechanisms of distraction and potential mitigations at each level of control that might not be considered with the more conventional information processing description of driver distraction.

Appendix A provides a theoretical framework that could be applied to mitigate distraction by predicting, identifying, and summarizing distraction indicators and providing feedback to the driver to induce positive behavioral changes. A critical outcome of this analysis is that no single sensor, algorithm or mitigation can be judged "optimal"; instead the real value of each depends on how it is combined with others. The summary tables in this section describe the characteristics of a mitigation system that are needed to support an evaluation of the combined system. There are several important implications for detecting and mitigating distraction, and evaluating candidate systems:

- Distraction occurs when the driver's capacity to respond to driving is compromised due to competing demands from the roadway and the secondary task. This implies that the entire driver–vehicle–environment relationship should be considered.
- Visual and manual (reaching), visual (reading), and cognitive distractions are qualitatively different types of distraction and may require different sensors, sensor combinations, and algorithms to detect.
- Measures of driver distraction cover a spectrum of inputs, including driver control inputs, vehicle state, body, head movement, eye movement, and physiological indicators. Unobtrusive sensors such as steering, eye glance, and lane position sensors, provide particularly promising estimations of distraction, especially when combined together.
- Algorithms need to match the type of distraction to the requirements of the mitigations supported (e.g., predict, identify, or summarize distraction indicators).
- Algorithms that detect distraction prior to vehicle state changes such as sudden braking may be particularly beneficial if they are sensitive and timely.

- Because of its unobtrusiveness, relative feasibility, and applicability for different timelines, enhanced feedback shows particular promise.

CHAPTER 3. TEMPLATE-BASED APPROACH FOR DESCRIBING EXISTING SYSTEMS

The previous chapter discussed mechanisms of driver distraction, as well as the strengths and weaknesses of different measures used to detect it. The chapter also provided a general discussion of the diverse approaches to mitigate distraction and the interdependence of detection and mitigation. This chapter attends to the problem of describing existing and future systems that detect and mitigate distraction. Currently, there is no systematic way to describe and compare systems, making assessments of their purpose and efficacy difficult. A common language is needed to describe and differentiate between the type and quality of data that distraction detection systems need as input, what they produce as output, and what mitigation strategies they support. This chapter develops and applies a template-based approach to describe distraction detection and mitigation systems in a systematic and consistent manner. The first section outlines the rationale for a template-based description, the second describes the template development and refinement, the third applies the template to two distraction detection algorithms evaluated in Chapter 4 and the two distraction mitigation systems evaluated in Chapter 5, and the final section describes the implications of the template application for benefits analysis and system evaluation.

RATIONALE AND CONSIDERATIONS FOR A TEMPLATE-BASED DESCRIPTION

The benefit of a template-based approach is that it identifies notable commonalities and differences between systems, and therefore provides government agencies, automotive companies, and consumers with a common language for describing distraction mitigation systems and understanding their functional and operational characteristics. Templates communicate system functionality without divulging proprietary information. This gives government agencies a valuable tool to evaluate the efficacy and potential problems of the rapidly evolving products manufacturers are developing. For example, templates provide NHTSA and the automotive industry with a catalog of features and functions of distraction detection and mitigation systems that can describe why a system will enhance driving safety and the overall benefits the system might provide. Templates also can capture system shortcomings, such as the sensor and algorithm limitations discussed in Chapter 2, which ultimately diminish countermeasure efficacy. Consumers also benefit from these templates because they inform purchasing decisions, and could have the positive ancillary effects of raising driver awareness of distraction and creating accurate driver mental models of mitigation system operation. Ultimately, a template-based description could form the basis for New Car Assessment Program (NCAP) ratings that differentiate highly capable distraction mitigation systems from less capable systems.

Already systems designed to detect and mitigate impairment are on the market or exist as advanced prototypes, including Saab’s ComSense, Volvo’s Driver Alert Control, Delphi’s SAVE-IT system, and Lexus's Driver Monitoring System. These systems detect changes in driver state with a general focus on drowsiness, distraction, or both, and intervene with different strategies. They use different sensors, integrate data differently, and serve different purposes, ultimately producing unique profiles of detection accuracy and supporting different types of distraction mitigation. Without a standard template to describe these systems, it is difficult to compare them. The following sections describe a framework—the abstraction hierarchy—for developing templates for distraction detection and mitigation systems that link concrete details of the system to a more abstract description of the system purpose. Applying the template generated by this framework identifies the key differences between various distraction detection and mitigation systems.

TEMPLATE DEVELOPMENT AND REFINEMENT

Many different approaches could be used to structure a template for describing distraction detection and mitigation systems. Here we use a systems engineering framework—the abstraction hierarchy—to define the general classes of information and organize the information according to input, transformation, and output. Applying these templates to existing systems revealed challenges associated with data availability, and identified opportunities to simplify the template form based on the use of publicly available information.

The Abstraction Hierarchy Framework for Describing Distraction Detection and Mitigation Systems

The abstraction hierarchy provides a natural organizing framework for describing complex systems systematically (Rasmussen, 1983). The abstraction hierarchy is a means-end structure that links a description of the physical system with the system's purpose and its functional description. Applied to distraction mitigation systems, the abstraction hierarchy identifies three levels of description: intentional—why the system is designed; functional—what the system does; and physical—how the system is configured. Such an approach makes it possible to describe the same system in very specific concrete terms (e.g., IR video-based eye-tracking system) and in more abstract, functional terms (detecting attention to the road and inside the vehicle). This is important because it allows for comparisons across systems that might use different sensors and low-level data, but might be very similar in other ways.

The abstraction hierarchy has several important qualities for describing distraction detection and mitigation systems. One of the most important is that it can help identify multiple ways of achieving the same ends. For example, a detection system might estimate dangerous levels of distraction based on gaze while another might use steering inputs. The purpose of the system is similar even though the functional means of achieving it are different. Likewise, two systems might have very different physical characteristics, but similar functional properties, such as laser- and radar-based systems for detecting vehicles ahead. Further, this framework could be used to identify common mechanisms that undermine the performance of detection and mitigation systems. For example, detection systems that are based solely on eye movement inputs might mistake scanning at intersections as instances of distraction. Overall, the abstraction hierarchy provides a basis for identifying common elements and generalizing performance expectations from one system to another (Bisantz & Vicente, 1994; Rasmussen, 1985).

Relationship between Distraction Detection and Mitigation

Although it is useful to consider them separately, distraction detection and mitigation components are often coupled. Information about the sensor and system characteristics associated with timeliness and accuracy of distraction detection might be tightly coupled with mitigation characteristics, such as timeliness for concurrent countermeasures (intended for use while driving), and the understandability and specificity of post-drive feedback. Some distraction mitigations (e.g., auditory alerts) may only require binary input indicating whether or not a driver is distracted. Other concurrent or post-drive feedback may need to present more detailed data concerning the severity or type of distraction, or where it occurred. Graded input, required for feedback that would permit drivers to continually monitor their level of distraction (e.g., a visual alertness display), may provide more opportunity to understand the context that precedes an alert, thus facilitating driver understanding of why an alert was presented, increasing trust (Lee & Lee, 2007) and possibly recalibrating drivers' risk comprehension related to distracting activities. Measures that immediately mitigate distraction demand shorter detection times than feedback provided at the end of the drive (Donmez et al., 2009). The importance of the timing of distraction detection depends on the mitigation strategy.

Type of distraction is also important. For example, the attention redirection approach (e.g., Fuchs et al., 2008; Engstrom & Victor, 2009) to distraction mitigation requires the detection algorithm to provide data that corresponds to the relevant type of distraction, because attention needs to be redirected differently to mitigate visual and auditory or cognitive distraction. Whereas it may be possible to prevent or mitigate visual distraction through an attention redirection display that draws the driver's visual attention back to the road center, this strategy would not be appropriate for cognitive distraction, which instead may require the display to prompt the driver to reinstate an adequate scanning pattern and alert the driver to hazards, such as bicycle riders, in the visual periphery (Reyes & Lee, 2008). Some augmented cognition strategies determine the sensory or cognitive channel for feedback taking advantage of unused resources when the resources required for a particular warning modality are unavailable (Kincses, 2006). Supporting this type of mitigation would require detection algorithms to identify the sensory modality or spatial/verbal information code of the resource-limiting distraction. Some distraction mitigation strategies need very specific indications of the type of distraction being detected.

Specification Template Development

The application of the abstraction hierarchy begins with a definition of system boundaries. Dividing a complex system into component parts can manage complexity and helps highlight issues that might be masked in considering the system as an undifferentiated whole. Although often coupled, an obvious way to divide the overall system is in terms of distraction detection and distraction mitigation.

Table 1 shows the nine classes of information that define the specification template for distraction detection systems. The rows derive from the abstraction hierarchy and the columns focus on the fundamental sensing, processing, and action elements that provide input to the distraction countermeasure. The upper row considers the system in terms of its purpose. This purpose-based or intentional perspective describes the system in terms of why it was developed and represents the intended capacity and the fundamental assumptions regarding its operation. The middle row considers the system in terms of what it achieves and represents the functions supported by the system. The bottom row considers the system in terms of how it is implemented and represents the physical characteristics of the system. These classes of information provide a complete description, ranging from why the system was built to how it was configured.

Table 2 provides the definitions of the mitigation parameters. Distraction mitigation takes the output of the detection process as input, linking the mitigation strategy with the purpose of detection (Table 1). The primary intentional inputs are related to the distraction type that requires mitigation, and the conditions under which the mitigation is intended to operate, such as a specified minimum speed.

At the functional level of abstraction, the same descriptors (driver, vehicle, environment, task, integrated system) apply to the inputs for both the mitigation and detection functions. They are not identical inputs, however, because the detection inputs are transformed before becoming mitigation inputs. The functional transformation of the mitigation can tailor it to specific circumstances by taking additional factors, such as traffic demand, into account. The functional description of the mitigation output indicates the type of feedback and how the mitigation achieves its purpose. The physical output, in contrast, consists of display characteristics or the interface with other vehicle systems.

Table 1 Classes of information for a distraction detection template

Level of abstraction	Input	Transformation	Output
Intentional—Why	Assumed operating conditions (e.g., not urban driving)	Theoretical framework and fundamental assumptions (e.g., assess degree of multiple resource theory information overload)	Countermeasure supported (e.g., post-drive feedback) Performance criteria (e.g., optimal Beta)
Functional—What	Information streams (e.g., gaze position and steering angle)	Algorithm (e.g., glance away from the road more than two sec)	Driver state (e.g., attentive or distracted)
Physical—How	Sensors (e.g., video camera)	Processing hardware (e.g., processing capacity of onboard computer)	Output data streams (e.g., binary indicator)

Table 2 Classes of information for a distraction mitigation template

Level of abstraction	Input	Transformation	Output
Intentional—Why	Assumed operating conditions (e.g., not urban driving)	Real time Post drive	Real-time feedback Post-drive feedback Performance criteria (e.g., timing, integration)
Functional—What	Information streams (e.g., attentiveness level)	Algorithm (e.g., driving situations and time aggregation)	Mitigation (e.g., feedback resolution and mode)
Physical—How	Wired or wireless link to mobile device	Processing hardware (e.g., specificity, resolution)	Output data streams (display or interface)

Application and Refinement of Detailed Distraction Detection and Mitigation Templates

Table 3 and Table 4 define the detailed draft distraction detection and mitigation template variables, respectively, at the three levels of abstraction. The tables also include a summary of the most important differences between systems revealed by the templates' application and a subjective estimate of the utility of each template variable for comparing systems and estimating their benefits.

The draft templates were applied to several distraction detection and mitigation systems that are in production or exist as advanced prototypes: Saab's ComSense, Volvo's Driver Alert Control, the SAVE-IT system, Lexus's Driver Monitoring System, Mercedes-Benz's Attention Assist, and Toyota's Wakefulness Level Judging System. The initial application of the draft template and concomitant textual summaries for each of these systems was intended to identify the key variables that capture differences among the systems reviewed, and to identify any gaps in the template categories that could be addressed with its revision (for examples of the textual summary, see Appendix B). Physical measurements were not conducted.

The preliminary state and proprietary nature of this technology limit the available information particularly at the lower levels of the hierarchy, but all systems that at least had general information available at the highest level of abstraction were described. Several complementary information collection strategies were used to describe the systems. Patents provided the most detailed information about distraction detection systems, including information about system data sources, transformations, and outputs. However, patents do not necessarily describe specific production systems, so it was not possible to obtain a comprehensive system description from this source. In addition, the delay between the patent process, research and development, and the introduction of the technology to the market further limits the ability to link information in patents to the final system. Trade and academic journals and web-based content, conference proceedings, government reports, and manufacturer websites were also included in order to complement and verify information gathered during the patent review. The confluence of information supports a better description of the system than any one source.

Table 3 Summary of the distraction detection template and its application

Template element	Label	Definition	Outcome of application and utility
Intentional—Input	Assumed operating conditions	According to the manufacturer, driving conditions where the system is expected to work and situations where it is not.	Roadway conditions (weather, road markings), vehicle interior conditions (illumination, glasses), and vehicle state (speed limitations) are important classes of operating conditions that distinguish the systems. Range of application across operating conditions represents a critical variable for benefits estimation Utility: Good
Intentional—Transformation	Theoretical constructs	Underlying theoretical assumptions of the algorithm, such as the constructs used to define the type of distraction being detected. These include information overload and attention as a resource compared to attention as a dynamic focus jointly defined by the driver and the distracting technology. These theoretical constructs define the underlying assumptions of the system design.	Most systems do not focus on distraction, but apply to a range of impairments (e.g., drowsiness). This criterion should be refined to address this distinction more precisely. Utility: Good
Intentional—Output	Performance criteria	The objective function for evaluating the system quality. This could be d' and optimal Beta.	Little or no publically available information for this criterion. Utility: Poor
Intentional—Output	Countermeasure supported	This represents the ultimate purpose of the distraction detection system. Some systems might detect distraction as a byproduct of their primary purpose, such as a drowsiness detection system; others might address visual/manual distraction and neglect cognitive distraction.	Real-time distraction prevention (workload management), attention redirection and alerts, adaptation of active safety system, cumulative feedback (trip report, SAVEIT). A central differentiating factor that is tied to the type of mitigation supported by the system. Utility: Very good

Table 3 Summary of the distraction detection template and its application (Continued)

Template element	Label	Definition	Outcome of application and utility
Functional—Input	Information streams	Driver, vehicle state, and environmental variables that form the algorithm input. These are functionally relevant transformations of raw data, such as gaze to the roadway.	Eye lid movement, pupil tracking, gaze determination, face orientation (Lexus); combination of vehicle data (speed, turn signal status, cruise control state, steering wheel angle data, brake pedal position, throttle pedal position, lane position, etc.), environmental data (headlamp status, wiper status, defroster status, GPS data, lane width, etc.), driver data (gaze position, pupil diameter, heartbeat), current task data (radio, phone status), additional data through integration with active safety systems) (Saab); vehicle and environmental data only (lane position, road geometry (Volvo, current models); vehicle, environment and driver (Volvo, prototype); vehicle and environmental data, especially steering, (Mercedes); head pose and environment/driving task demand (SAVEIT); head movement, eyelid closure, duration and frequency of blinks (Seeing Machines). This variable highlights the additional capacity afforded by systems integrated into the vehicle compared to aftermarket systems that are not able to draw upon vehicle state data. Utility: Very good.

Table 3 Summary of the distraction detection template and its application (Continued)

Template element	Label	Definition	Outcome of application and utility
Functional—Transformation	Information combinations	Combination of information streams that form functionally useful descriptions of driver state. This includes time window size over which information is combined and data smoothing to accommodate noisy data.	The relationship between eye gaze and head movement, taking into account the peripheral vision of the road ahead for adaptive warnings (such as when looking in the rear-view mirror or turning a corner) (Saab); planned path, planned deviation (Volvo); unspecified combination of vehicle signals, PERCLOS, eyes on Road Center Point, etc. (Volvo prototype); actual and theoretical course angle comparison (Mercedes); PERCLOS (Seeing Machines). Utility: Good
Functional—Transformation	Algorithms	The functional relationship between the information streams and the output driver state estimation. This includes whether or not the algorithm is generic or adjusted to the individual driver and the time window size used in algorithm performance.	Largely unavailable due to proprietary concerns and may not be a useful differentiator. Utility: Poor
Functional—Output	Driver state	The type and degree of driver impairment detected. Some algorithms may simply report a binary level of generic impairment, whereas others might report graded levels of impairment associated with specific types of distraction.	Not clearly defined in available information, but is a useful differentiator. Utility: Good

Table 3 Summary of the distraction detection template and its application (Continued)

Template element	Label	Definition	Outcome of application and utility
Physical—Input	Sensors	The physical sensors used as input to the system, such as video cameras for gaze, head pose, or lane position estimation.	Inconsistently available, and where available, it may not provide a useful distinguishing characteristic. Utility: Poor
Physical—Transformation	Processing hardware	The computer hardware that supports the algorithms.	Not available. Utility: Poor
Physical—Output	Signal	The data from the distraction detection system, either nominal, ordinal or continuous signal, and its time-varying characteristics	Meaningful only in the context of the mitigation strategy. Utility: Poor for distraction detection alone

Table 4 Summary of the distraction mitigation template and its application

Template element	**Label**	**Definition**	**Outcome of application and utility**
Intentional—Input	Driver State	The driver states and their consequences to which the mitigation will be applied.	Systems mitigate different driver states related to inattention, including visual distraction, fatigue, and drowsiness. Driving demand may condition the countermeasure deployment. Utility: Good
Intentional—Transformation	Mitigation Strategy	The approach to mitigation, for example an advising strategy in which the driver is alerted to the distracted state or a distraction mitigation strategy that directly addresses the source of distraction.	Although most systems mitigate inattention through real-time driver feedback, other strategies have been used. Utility: Fair
Intentional—Output	Performance criteria	The objective function for evaluating the system quality. It could include metrics related to salience/authority, timeliness, understandability, and acceptability.	Little information for this criterion is available, limited to research prototypes. Utility: Poor
Intentional—Output	Countermeasure supported	This represents the ultimate purpose of the distraction mitigation system. Some systems passively advise the distracted driver, who can then take an appropriate action, while others actively mitigate the distraction.	Most passive countermeasures are intended to provide real-time feedback, but a variety of approaches are taken in active countermeasures – attention redirection, IVIS management, and Integrated/Adaptive Crash Warning Systems (CWS) occur in both research and production systems Utility: Very good

Table 4 Summary of the distraction mitigation template and its application (Continued)

Template element	Label	Definition	Outcome of application and utility
Functional—Input	Information streams	Driver and vehicle state variables that govern when and how the mitigation will occur. These are functionally relevant transformations of raw data and detection algorithm output.	Driver input includes visual distraction, fatigue (Mercedes), control (Volvo DAC). Vehicle input includes driving demand (Saab ComSense, SAVE-IT). Integrated systems provide traffic risk (Lexus/Toyota) and lane departure input (Volvo Research 2). Utility: Very Good.
Functional—Transformation	Information combinations	Combination of information streams that result in functionally useful mitigations. This includes the possibility of modality augmentation/shifting, as well as information combinations that constrain the mitigation to occur under conditions where it will function well.	Mitigations may combine detection input with information indicating that an object is in the vehicle's path before activating (Lexus). Mitigations that reduce IVIS demand (lock-out strategy) may combine detection input with IVIS status (SAVE-IT). Mitigations may also combine detection and driver inputs to cancel an alert (Saab AttenD). Utility: Good
Functional—Transformation	Algorithms	The functional relationship between the information streams and the output mitigation.	All systems use thresholds to convert input to graded or binary output. Some employ algorithms to operationally define constructs such as "control" (Volvo Driver Alert Control), "emergency" (Saab ComSense), or "risk" (Lexus/Toyota). Another distinction among systems is whether some conditions inhibit alerts (Saab AttenD). Also, the algorithms that support post-drive feedback (Seeing Machines, SAVE-IT) would differ from those that support real-time feedback. Utility: Good

Table 4 Summary of the distraction mitigation template and its application (Continued)

Template element	Label	Definition	Outcome of application and utility
Functional—Output	Mitigation	Corresponding to algorithms that report a binary level of impairment, some driver feedback may only consist of issuing a single alert. Feedback may be presented in various modes and the vehicle may actively respond to address the driver's distracted state.	The current outputs are generally not graded, with the exceptions of Volvo Driver Alert Control, Saab AttenD (3 levels), and SAVE-IT. Feedback is presented in visual, auditory (including two examples of voice feedback), and vibration modalities. Several mitigations pair a vehicle response with driver feedback. The systems differed in the specific driver state that they mitigated and in the specificity of driver feedback. Mitigations are further distinguished by whether they include a vehicle response, including crash preparations when a driver is distracted and about to crash, or delaying phone calls when traffic demands full attention. Utility: Very Good
Physical—Input	Information Streams	The mitigation can use physical sensors and vehicle system interfaces for driver, vehicle, environment, task, and integrated system data.	The physical input streams to the mitigation reflect sensor and algorithm performance but are not often described in our sources. Utility: Poor
Physical—Transformation	Processing hardware	The computer hardware that supports the algorithms, triggers, and conditions, and specifies the mitigation.	Not available. Utility: Poor
Physical—Output	Display or interface	The physical display and vehicle interface parameters are a product of the distraction mitigation strategy. Strategies of distraction prevention, distraction mitigation (including driver feedback), and CWS adaptation employ different driver and vehicle interfaces.	The primary distinction among the physical outputs is the mitigation strategy that they support because their benefits depend on the success of the strategy. The mitigation system interface with the driver and/or vehicle supports mitigation, prevention, or CWS adaptation strategies. There is little detailed information available about visual display intensity and location, or the tonal qualities of auditory alerts. Utility: Fair.

Simplified Template Describing Both Distraction Detection and Mitigation

From the perspective of creating templates that clearly distinguish between systems, the abstraction hierarchy levels describing the countermeasure emerged as perhaps the most distinguishing feature. The importance of this component suggests that a template that uses it as the primary organizing principle might provide a more understandable description of system functionality. It also better represents the high degree of coupling between effective distraction detection and mitigation subsystems. Such an organization would make some criteria useful that are not when applied only to distraction detection. For example, the last entry in the table "Physical Output" provides little value for describing distraction detection systems, but could have great value in describing an integrated distraction detection and mitigation system. However, insufficient detail was available about the attributes of physical output such as physical location dimensions, luminance, and tonal frequencies so it was not included in the simplified template.

Other levels were also instructive, either for their ability to distinguish between systems or their lack of utility (Tables 6 and 7). The functional properties distinguish distraction detection systems well. Some systems track head position and/or eye movements, others use driver bio-behavioral information (features of the eyes, face, head, etc.), and some also derive general indices of driver performance (e.g., control). These differences suggest that using different types of sensors (CCD camera, vehicle data sensors, environmental sensors, etc.) and inputs (eye and/or head position, eye glance location, steering angle, etc.) might be an important differentiator. Little information about the algorithms and processing hardware at the physical level was found, so these considerations are not used to distinguish between the systems.

Based on these and other outcomes of the template test applications, a simpler and more concise template was created (Figure 2). Appendix C describes this template and provides examples of its application. It condenses the elements from Table 3 and Table 4 to describe detection and mitigation as an integrated system. The template adopts the elements of the distraction detection template at the intentional and functional levels of abstraction, but includes a greater focus on the information related to mitigation. The redundancy of driver state as the detection output (i.e., driver state) and mitigation input allowed for elimination of the mitigation input category (even though other inputs such as speed may only affect the mitigation). Any environmental or other inputs were assumed to be received by the detection component. These modifications reflect the challenges noted in the "Outcome of application and utility" column in Table 6 and Table 7. This is particularly true in the "Physical output" element. In the revised template, this element refers to output of the mitigation strategy and so it describes the specific nature of the feedback the driver receives, such as icon color of a visual display, and type of auditory alert.

The simplified template form also changed the open-ended format represented in the full template to a standardized form with pre-defined variables in order to apply standard terminology to distinguish between meaningful system characteristics. The simplified template variables were identified through an analysis of key system differences that have been described in publicly available sources.

Manufacturer: Volvo
System Name: Driver Alert Control

Model Year: 2008/2009/2010 S80, V70, XC70 (US), 2009 XC90 (US)
2010 XC60 (US), XC70 (US, NZ), S80 (US)

☐ **Aftermarket** *or* ☒ **OEM**
☒ **Production** *or* ☐ **Research**

☒ Standard configuration ☐ Reconfigurable Notes:

Detection - Purpose

Input	Applicable?			Requirement					Notes
Road Conditions	☒ Yes	☐ No	☐ Unknown	☒ Good visibility	☒ Visible Road Markings	☒ High Speed Roads	☐ ____	☐ ____	
Vehicle Interior	☐ Yes	☐ No	☒ Unknown	☐ Ambient Illumination	☐ Sound Level	☐ ____	☐ ____	☐ ____	
Vehicle State	☒ Yes	☐ No	☐ Unknown	☒ Speed	☐ Traffic Demand	☒ Imminent Collision	☒ Lane Marking	☐ ____	system activated at speeds > 40 mph, remains active until speeds < 37 mph
Driver	☐ Yes	☒ No	☐ Unknown	☐ Head Position	☐ Glance	☐ Posture	☐ ____	☐ ____	
____________	☐ Yes	☐ No	☐ Unknown	☐ ____	☐ ____	☐ ____	☐ ____	☐ ____	
Transformation	**Applicable?**			**System Type**					**Notes**
Derivative System	☒ Yes	☐ No	☐ Unknown	☒ Drowsiness Fatigue	☐ CWS	☐ ____	☐ ____	☐ ____	
Detection Approach	☒ Yes	☐ No	☐ Unknown	☐ Visual Search	☒ Vehicle Control	☐ ____	☐ ____	☐ ____	
____________	☐ Yes	☐ No	☐ Unknown	☐ ____	☐ ____	☐ ____	☐ ____	☐ ____	
Output	**Applicable?**			**Supported**					**Notes**
Driver State	☒ Yes	☐ No	☐ Unknown	☐ Visual Distraction	☐ Cognitive Distraction	☒ Inattention	☐ ____	☐ ____	Inattention: fatigue and drowsiness distraction detection as a byproduct
____________	☐ Yes	☐ No	☐ Unknown	☐ ____	☐ ____	☐ ____	☐ ____	☐ ____	

Manufacturer: Volvo **Model Year:** 2008/2009/2010 S80, V70, XC70 (US), 2009 XC90 (US) ☐ **Aftermarket** *or* ☒ **OEM**

System Name: Driver Alert Control 2010 XC60 (US), XC70 (US, NZ), S80 (US) ☒ **Production** *or* ☐ **Research**

Detection - Function

Input	Applicable?	Information Streams	Notes
Driver Data	☐ Yes ☒ No ☐ Unknown	☐ Pupil ☐ Eye Gaze ☐ Head Pose ☐ ☐	
Vehicle Data	☒ Yes ☐ No ☐ Unknown	☒ Longitudinal Control ☒ Lateral Control ☐ CWS ☐ ☐	
Environmental Data	☐ Yes ☒ No ☐ Unknown	☐ GPS Location ☐ Headlights ☐ ☐ ☐	
Task Data	☐ Yes ☒ No ☐ Unknown	☐ Audio Status ☐ Phone Status ☐ IVIS Status ☐ ☐	
______	☐ Yes ☐ No ☐ Unknown	☐ ☐ ☐ ☐ ☐	
______	☐ Yes ☐ No ☐ Unknown	☐ ☐ ☐ ☐ ☐	
Transformation	**General Algorithm Information**		
Algorithm	Machine Learning? ☐ Yes ☐ No ☒ Unknown	*Notes:* Planned deviation measures (MPD) are compiled by forming a mean over the planned deviation measures. If this measure exceeds a threshold, the state of distraction is detected	
Output	**Applicable?**	**Driver State**	**Notes**
Driver State: Type	☒ Yes ☐ No ☐ Unknown	☐ Visual Distraction ☐ Cognitive Distraction ☒ Inattention ☐ ☐	
______	☐ Yes ☐ No ☐ Unknown	☐ ☐ ☐ ☐ ☐	
______	☐ Yes ☐ No ☐ Unknown	☐ ☐ ☐ ☐ ☐	

Manufacturer: Volvo | **Model Year:** 2008/2009/2010 S80, V70, XC70 (US), 2009 XC90 (US) / 2010 XC60 (US), XC70 (US, NZ), S80 (US) | ☐ **Aftermarket** *or* ☒ **OEM**

System Name: Driver Alert Control | ☒ **Production** *or* ☐ **Research**

☒ Standard configuration ☐ Reconfigurable Notes:

Countermeasure-Purpose

Output: Concurrent	**Applicable?**			**System Type**					**Notes**
Distraction Feedback	☒ Yes	☐ No	☐ Unknown	☐ Distraction Prevention	☒ Distraction Mitigation	☐ Collision Mitigation	☐ ____	☐ ____	
In-Vehicle Information Management	☐ Yes	☒ No	☐ Unknown	☐ Distraction Prevention	☐ Distraction Mitigation	☐ ____	☐ ____	☐ ____	
CWS Adaptation	☐ Yes	☒ No	☐ Unknown	☐ Adaptive Passive Warning	☐ Shift to Active if Distracted	☐ Adaptive Active Assistance	☐ ____	☐ ____	
Attention Redirection	☐ Yes	☒ No	☐ Unknown	☐ Forward Hazard	☐ Peripheral Hazard	☐ Speed Control	☐ ____	☐ ____	
____	☐ Yes	☐ No	☐ Unknown	☐ ____	☐ ____	☐ ____	☐ ____	☐ ____	
____	☐ Yes	☐ No	☐ Unknown	☐ ____	☐ ____	☐ ____	☐ ____	☐ ____	
Output: Post-Drive	**Applicable?**			**System Type**					**Notes**
Behavioral Change	☐ Yes	☒ No	☐ Unknown	☐ Driver	☐ Fleet Manager	☐ ____	☐ ____	☐ ____	
____	☐ Yes	☐ No	☐ Unknown	☐ ____	☐ ____	☐ ____	☐ ____	☐ ____	
____	☐ Yes	☐ No	☐ Unknown	☐ ____	☐ ____	☐ ____	☐ ____	☐ ____	
____	☐ Yes	☐ No	☐ Unknown	☐ ____	☐ ____	☐ ____	☐ ____	☐ ____	

Manufacturer: Volvo **Model Year:** 2008/2009/2010 S80, V70, XC70 (US), 2009 XC90 (US) ☐ **Aftermarket** *or* ☒ **OEM**

System Name: Driver Alert Control 2010 XC60 (US), XC70 (US, NZ), S80 (US) ☒ **Production** *or* ☐ **Research**

Countermeasure-Function									
Output: Concurrent	**Applicable?**			**Specification**					**Notes**
Distraction Alert Display Timing	☒ Yes	☐ No	☐ Unknown	☒ Real Time	☐ Delayed	☐	☐	☐	
Distraction Alert Specificity	☒ Yes	☐ No	☐ Unknown	☒ Visual Distraction	☐ Cognitive Distraction	☒ Inattention	☐	☐	
Attention Feedback Presentation	☒ Yes	☐ No	☐ Unknown	☒ Discrete Alert	☒ Continuous Level	☐	☐	☐	Continuously displayed graded concentration level feedback (physically displayed as 5 bars). A binary aural and visual (coffee cup) and verbal (Time for a Break) warning is displayed if it reaches a threshold.
Distraction Alert Modality	☒ Yes	☐ No	☐ Unknown	☐ Tone	☒ Voice	☒ Visual	☐ Haptic	☐	
Distraction Alert Response Timing	☐ Yes	☒ No	☐ Unknown	☐ Milliseconds	☐ Seconds	☐	☐	☐	
Distraction Alert Resolution	☒ Yes	☐ No	☐ Unknown	☒ Binary	☒ Graded	☐	☐	☐	
In-Vehicle Information Management	☐ Yes	☒ No	☐ Unknown	☐ Vehicle Status	☐ Audio	☐ Telecom	☐	☐	
CWS Adaptation	☐ Yes	☒ No	☐ Unknown	☐ Delay if Attentive	☐ Adapt Passive Alert Intensity	☐ Adapt Active Assistance Force	☐	☐	
Attention Redirection Modality	☐ Yes	☒ No	☐ Unknown	☐ Tone	☐ Voice	☐ Visual	☐ Haptic	☐	
____________	☐ Yes	☐ No	☐ Unknown	☐	☐	☐	☐	☐	Continuously displayed graded concentration level feedback (physically displayed as 5 bars). A binary aural and visual (coffee cup) and verbal (Time for a Break) warning is displayed if it reaches a threshold.
Output: Post-Drive	**Applicable?**			**Specification**					**Notes**
Cumulative Feedback	☐ Yes	☒ No	☐ Unknown	☐ Quantitative	☐ Incident Replay	☐	☐	☐	

Figure 2 Simplified template describing distraction detection and mitigation systems

APPLICATION OF TEMPLATE TO DISTRACTION DETECTION AND MITIGATION SYSTEMS

The template was used to describe and compare two detection algorithms evaluated in Chapter 4, representing the two extremes of algorithm complexity: the simple *Eyes off forward roadway* algorithm (Klauer et al., 2006) and the sophisticated *multidistraction detection* (Victor, 2010) with modifications by NADS for implementation in the countermeasure effectiveness evaluation (Chapter 5). Template specifications for the *Eyes off forward roadway* and *multidistraction detection* algorithms can be found in Appendix C. Although similar in approach – both algorithms are designed to detect distraction using eye-based measures – the analytic assessment shows differences in the intent of the algorithms, their requirements, the countermeasures supported, and the algorithms' functionality. The *Eyes off forward roadway* algorithm was derived using odds ratios to determine what threshold for eyes off the forward roadway leads to a statistically significant increase in crash risk (Klauer et al., 2006). The algorithm identifies a cumulative glance away from the road of two seconds within a 6-second running window as visual distraction. The required inputs are limited to measurements obtained from the driver, specifically eye glance measures.

The *multidistraction detection* algorithm (Victor, 2010) provides drivers with real-time alerts that correspond to an array of risky scanning behaviors associated with distraction. It is based on the eye glance measurement of percent road center (PRC). Although both algorithms use only driver data as primary inputs, the *multidistraction detection* algorithm has several measures of driver state (eye glance, head pose, and weight distribution), creating layers of redundancy to compensate for sensor signal quality issues. When eye glance data is unavailable, the algorithm uses head pose data to calculate PRC. The *multidistraction detection* algorithm also uses vehicle state inputs (i.e., speed) to adjust thresholds for algorithm variables. Below 25 miles per hour, the algorithm is not activated, and once activated the vehicle must maintain 23 miles per hour to remain engaged. Vehicle speed is also used to adjust the road center cone: the road center cone is widened in low speed environments where drivers may be more actively scanning the roadside. The size of the road center cone also adjusts to the sensor signal, increasing from 10 to 20 degrees when sensor input shifts from eye glance to head pose signals.

The primary distinction between the two algorithms is the type of countermeasure supported: the *Eyes off forward roadway* algorithm can only support countermeasures for visual distraction, whereas the *multidistraction detection* algorithm supports mitigations for both visual and cognitive distraction. The latter algorithm actually identifies three types of distraction: (1) visual distraction from a single long (3 second) glance away from the roadway; (2) visual distraction from a history of glances away from road center (glances fall below a PRC of 60% within a 17.3-second running window); and (3) cognitive distraction (glances rise above a PRC of 83% within a 60-second running window).

The *multidistraction detection* algorithm was developed for production applications, and therefore is designed to be robust and reliable. To improve acceptability, its algorithm has mechanisms to reset the calculations of distraction when its criteria are not met (e.g., minimum speed requirements) or after distraction is indicated so repeated nuisance alerts are not issued. A visual time sharing (VTS) PRC window is also calculated to improve the consistency and reliability of distraction detection by resetting the visual and cognitive PRC windows when glances return to road center after a brief time off the road (PRC less than 65% in a 4 second window). Further details about these algorithms can be found in Chapter 4.

The template was also used to describe and compare the alternative countermeasure feedback evaluated in Chapter 5: real-time and post-drive feedback. Both distraction countermeasures used the modified multidistraction detection algorithm described above and in more detail in Chapter 4. The two countermeasures differ primarily in the time scale and intent of their feedback, with the real-time countermeasure providing concurrent feedback intended to mitigate distraction through visual and auditory alerts, whereas the post-drive countermeasure provided retrospective feedback designed to prevent future instances of distracted driving by showing drivers their level of distraction, distraction related driving performance decrements, measures of inattention, as well as video replays of their distracted driving. The distinction between distraction mitigation and prevention suggests the differences in purpose between the two approaches: to prompt an immediate change in performance or to present a pattern of behavior as a means to coach drivers to change their behavior and attitudes about distracted driving.

The specification templates clearly present the differing complexity of the two example algorithms. The templates describe substantial differences in purpose and functionality even between algorithms that are both based on eye glance variables. The multidistraction detection algorithm template includes variables that enhance system robustness when sensor signal quality is poor, and in situations where the driver's gaze would be expected to leave the road center in the course of normal driving. The countermeasure portion of the template indicates that the system provides feedback for the risky scanning behaviors engendered by engagement in both visual and cognitive secondary tasks. The templates do not explicitly capture the possibility that future mitigations will provide distinct real-time feedback for instances of distraction (long glances and distracted periods (a history of glances). The benefits revealed through a template can be difficult to assess because a good design concept (as shown in the template) can be poorly implemented. In this case, complexity may lead to confusion: drivers may not distinguish between the different eye glance patterns indicated by the algorithm and may disregard alerts because their meaning is not easily understood. Algorithms that change their performance based on speed or sensor signal inputs may be interpreted as unpredictable and inaccurate by drivers who fail to decipher patterns of algorithm performance in different driving situations and environments. Although simplistic, the *Eyes off forward roadway* algorithm is intuitive, and may more easily support changes in behavior, attitude and culture because the associated alerts are more easily understood.

IMPLICATIONS OF A TEMPLATE-BASED DESCRIPTION OF DISTRACTION DETECTION AND MITIGATIONS SYSTEMS

Complete descriptions of distraction detection and mitigation systems are not available, and the templates highlight the gaps. A greater level of transparency is needed to fully understand the potential benefits of distraction detection and mitigation systems, and depending on the future relationship between government and industry, the templates may need to adjust to address levels of information available. As the application of the templates to production and research systems and to example algorithms and alternative feedback systems indicates, distinctions between systems are possible but there is a limit to what can be done analytically. Even with full transparency, there would continue to be uncertainty about how systems actually work. Empirical evaluations may help fill this gap.

Even though many specific elements of the systems remain poorly defined, the template descriptions provide useful distinctions between type of distraction detected and mitigated, and intent and timescale of feedback. Two dimensions emerged that suggest important directions for future research. The first concerns the degree to which the algorithm is specific to distraction or a type of distraction: most do not distinguish. The second concerns the use of multiple measures to detect distraction: most do not.

The purpose of the systems differed from detecting and mitigating distraction in that most systems are intended to support better attention to the road independent of the cause of inattention. For example, Volvo's system estimates impairment associated with differences between the vehicle's path and the planned path. These differences indicate degraded vehicle control and could reflect distraction, drowsiness, or alcohol impairment. No production system distinguishes between cognitive distraction and visual distraction; the *multidistraction detection* algorithm is unique in that regard, although the technology developed by SeeingMachines also approaches this level of specificity. Highly precise characterization of eye movements could make it possible to identify drowsiness using PERCLOS measures, detect visual distraction associated with the time the drivers' eyes are off the road, and detect cognitive distraction associated with gaze concentration. The degree to which systems identify specific sources of impairment may be valuable for some distraction countermeasures and not others. For some countermeasures, such as those associated with adjusting parameters of collision warning systems, the specificity of the algorithm may not matter. For others, such as the post-drive feedback evaluated in Chapter 5, drivers might benefit from more specific information regarding the source of impairment so that they can adjust their behavior accordingly. Impairment specificity could be an important gap in distraction detection algorithms, but its importance depends on the countermeasure the algorithm supports.

The degree to which the algorithm integrates multiple variables to estimate distraction-related impairment also varied: most use a very limited set of variables. Some systems, such as the Lexus Driver Monitoring System, combine driver state data with collision warning data regarding crash threats, but only use a video-based indicator of driver gaze to assess distraction. The Lexus system does not use steering or accelerator modulation data to estimate distraction. None of the existing systems combine data to identify particular types of distraction, such as visual and cognitive. Again, the *multidistraction detection* algorithm fills this gap in the design landscape. A more specific distraction detection system could be quite valuable for certain countermeasures, such as post-drive feedback, and the integration of multiple measures could make this specificity possible.

Beyond these specific themes, a more general conclusion emerged: distraction detection algorithms need to be described in the context of the countermeasure they support. Countermeasure efficacy, driver acceptance, and the ultimate safety benefit depend on the match between the algorithm characteristics (e.g., specificity and sensitivity) and the countermeasure characteristics (e.g., structure, formatting, and timing of distraction-related feedback to the driver).

In general, distraction mitigation is supported by providing real-time feedback for immediate driving performance improvement. All these systems issue acoustic feedback. Some of them combine different distraction alert modalities such as acoustic and visual (Volvo's Driver Alert Control and Mercedes-Benz's Attention Assist) to enhance driver feedback reception. Volvo's and Saab's prototypes consider haptic modality of alert as well. This combination of different feedbacks is aimed at expanding mitigation of different types and degrees of inattention. For instance, drowsiness, cognitive, and visual distraction could be successfully supported by acoustic and haptic alerts but not by visual alone. However, a visual modality of alert can be used for feedback grading: Volvo rates driver attentiveness on a five-bar scale. Another application of the real-time mitigation developed by Seeing Machines (Driver State Sensor) is system integration into fleet management for later analysis or communication.

Real-time distraction prevention is implemented through workload management functions: when the current state of a driver or driving environment is considered highly demanding, the system interrupts a phone call (Saab's ComSense), applies emergency braking (Lexus's Driver Monitoring System) or corrective steering (Toyota's Wakefulness Level Judging System). The former two systems also use acoustic alerts to draw driver attention to the increased workload. In all of these cases, the negative impact of the immediate feedback is that it may impose more workload on a driver in addition to the already highly demanding situation (Donmez et al., 2008)

One important application of the templates is benefits estimation, which concerns assessing the potential impact of a system on driving safety and efficiency. The template's functional description of systems included in templates can support this application by linking system capabilities to the crash mechanism. For example, if visual and manual distractions represent the predominant contributors to distraction-related crashes, then a system that detects visual and manual distraction might provide a large benefit. If, on the other hand, cognitive distraction associated with hands-free cell phones is the predominant contributor and the system does not detect cognitive distraction, then the associated benefit might be correspondingly smaller. Chapter 6 describes how a template-based description of distraction mitigation systems identifies how such systems influence the multiple behaviors associated with distraction-related crashes and how such influences combine to enhance driving safety.

CHAPTER 4. DESCRIPTION AND APPLICATION OF A PROTOCOL FOR EVALUATING DISTRACTION DETECTION ALGORITHMS

Chapter 3 demonstrated that several systems are deployed that might detect distraction. The proprietary nature of algorithmic components limits the publically available information about these systems; hence, empirical evidence is needed to understand their capabilities. Currently there is no uniform method for assessing and comparing algorithms. Understanding the distraction potential of in-vehicle devices is a similar challenge, but several standard protocols for measuring distraction have emerged. For example, the lane change and shutter tasks have both been used to assess demand of secondary tasks (Mattes & Hallén, 2009). A similar protocol is needed to assess the ability of vehicle-based systems to detect distraction. This assessment protocol would provide an empirical basis for assessing the capabilities and vulnerabilities of algorithms. It would identify the most promising algorithms the interventions might support.

The purpose of this chapter is to define and apply an evaluation protocol to promising distraction detection algorithms. It focuses on assessing the ability of algorithms to detect distraction assuming the algorithm receives valid sensor data. The protocol requires a data collection process that samples a selection of drivers, driving situations, and representative distractions. Data collected from this process are then reduced and interpreted relative to evaluation criteria. The chapter describes each element of this protocol. Appendix V details a separate algorithm approach based on visual-motor coordination that was conducted parallel to this effort, but which is not discussed further in this chapter.

A central requirement of such a protocol is that the scenarios and secondary tasks are sensitive to distraction. The driving situation must place a sufficient demand for attention on the driver so that consequences of distraction can be observed in degraded vehicle control and attention to the roadway. This sensitivity to distraction is particularly important for evaluating countermeasures that the distraction detection algorithms support. The experiment described in Chapter 4 validates this protocol using driving situations that include secondary tasks and then the data are used to reveal algorithm capabilities and vulnerabilities.

The following sections describe the evaluation protocol and the results of its application:

- Data collection protocol, including participants, methodology, and procedure;
- Protocol sensitivity to distraction;
- Algorithm assessment for detecting distraction; and
- Conclusions and recommendations for the evaluation protocol and for algorithm performance.

Additionally, it should be noted that a parallel effort for the detection of alcohol and drowsiness related impairment is underway as part of NHTSA's Advanced Countermeasures for Multiple Impairments (ACMI) program.

DATA COLLECTION PROTOCOL FOR ALGORITHM EVALUATION

The data required for algorithm evaluation comprises a sample from representative drivers, in representative driving scenarios, performing representative distraction tasks.

Participants

Thirty-two participants balanced for gender age 25 to 50[5] were recruited for this study. Due to dropouts, another 14 participants were recruited to complete 32 sets of valid data. Among the dropped participants, seven withdrew due to simulator sickness, 5 were dropped due to bad eye-tracker data, 1 was dropped due to data collection tablet problems, and 1 was dropped due to simulator issues. Details of the screening criteria for participants, as well as other material used in recruitment and screening, are found in Appendices D, E, and F.

Among the 32 participants whose drive data was used, 17 were 25 to 34 years old, and 15 were 36 to 50; 16 were male, and 16 were female; 16 were white/Caucasian, 4 were Asian, 1 was black/African-American, and 1 was American Indian/Alaska Native.

Apparatus and Driving Scenarios

The experimental drives were conducted in a high-fidelity, motion-based driving simulator, the NADS-1. The simulator included a Chevy Malibu cab equipped with eye-tracking hardware, active feel on steering, brake, accelerator pedal, and a fully operational dashboard.

The algorithm evaluation drive represented several common environments, included events that challenge distraction detection algorithms, and contained distracting tasks that are likely to challenge the driver. The constraints governing the system were reviewed and compared to the planned driving environment to ensure a sufficient match as to make the evaluation useful (see Table 5). As can be seen by comparing the systems in the top rows of the table with the planned driving scenario environment in the last row of the table, the planned drive provides good coverage and meets the needs of this evaluation.

Participants completed the evaluation drive twice, once with distractions and once without. The drive begins in an urban setting that includes a transition to an urban arterial, then continues onto an interstate, and ends on a rural road. The order of these three segments was the same for all drivers and for both the baseline and distraction drives. During each segment, participants complete three prompted secondary tasks. The distraction tasks were presented as a cluster that could be completed in a 75-90 second time-window. All participants experienced all three distractions in all three driving

---

[5] This age group has a stable low crash risk per mile driven and was chosen to minimize within-group variability that might reduce the statistical power of the comparisons.

environments. However, if a participant deferred the start of a task, the allowable duration was extended by the amount of deferment.

Participants were instructed that these tasks represent urgent activities and they should complete as many as possible while driving as they normally would. If they deferred a task, a cue to initiate it would be repeated every 10 seconds. Drivers were also provided with task performance feedback to motivate good performance. When drivers were not engaged in these three "forced paced" or prompted distraction tasks, they were presented with a simple self-paced visual/manual task that required manual inputs to reduce the loudness and numerically displayed level of white noise. This task represented tuning the radio to minimize static. If no inputs were made, the loudness of the static would reach an uncomfortable level once per driving environment. Unlike the prompted distractions, the participants could often respond to the radio task or not, although most responded frequently. This simple visual/manual task was presented (without feedback) whenever the prompted secondary tasks were unavailable to assess drivers' willingness to engage across a wider range of driving situations.

Table 5 Driving scenario constraints by distraction system

Systems	Active Speed Range			Road Geometry					Environmental Conditions				Traffic Density	
	<35 mph	35-55 mph	>55 mph	Straight	Curve	Hill	Inter-section	Inter-change	Fog	Rain	Snow-covered roads	Poor lighting	Low	High
Volvo's Driver Alert Control (Prototype)	X	√[6]	√	√	?	?	√	?	X	?	X	X	?	?
Volvo's Driver Alert Control	X	√[1]	√	√	√	√	√	√	X	?	X	X	√	√
Saab's Driver Attention Warning System (AttenD)	√	√	√	√	√	√	X	√	√	√	√	√	√	√
Saab's ComSense	√	√	√	√	√	√	√	√	?	?	?	?	√	√
Lexus' Driver Monitoring System	√	√	√	√	√	√	X	√	?	?	?	√	√	√
Mercedes-Benz's Attention Assist	X	√[7]	√	√	√	√	√	√	?	?	?	√	√	√

[6] 40 mph and higher

[7] 50 mph and higher

Toyota's Wakefulness Level Judging System	?	√	√	√	√	√	√	√	X	?	X	X	√	√
Seeing Machines' Driver State Sensor	√	√	√	√	√	√	X	√	√	√	√	√	√	√
Delphi's SAVE-IT System	√	√	√	√	√	√	X	√	X	?	X	?	√	√
Planned Driving Environment	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	Y	Y	Y

√ = condition accommodated by the noted system
X = condition is not accommodated by the noted system
? = not known
Y = Included in planned drive
N = Not included in planned drive

Figure 3, Figure 4, and Figure 5 show the three road segments of the experimental drive – urban, interstate, and rural driving environments, respectively – as well as the location of the numbered scenario events. The road segments are associated with the corresponding events that represent a specific environment.

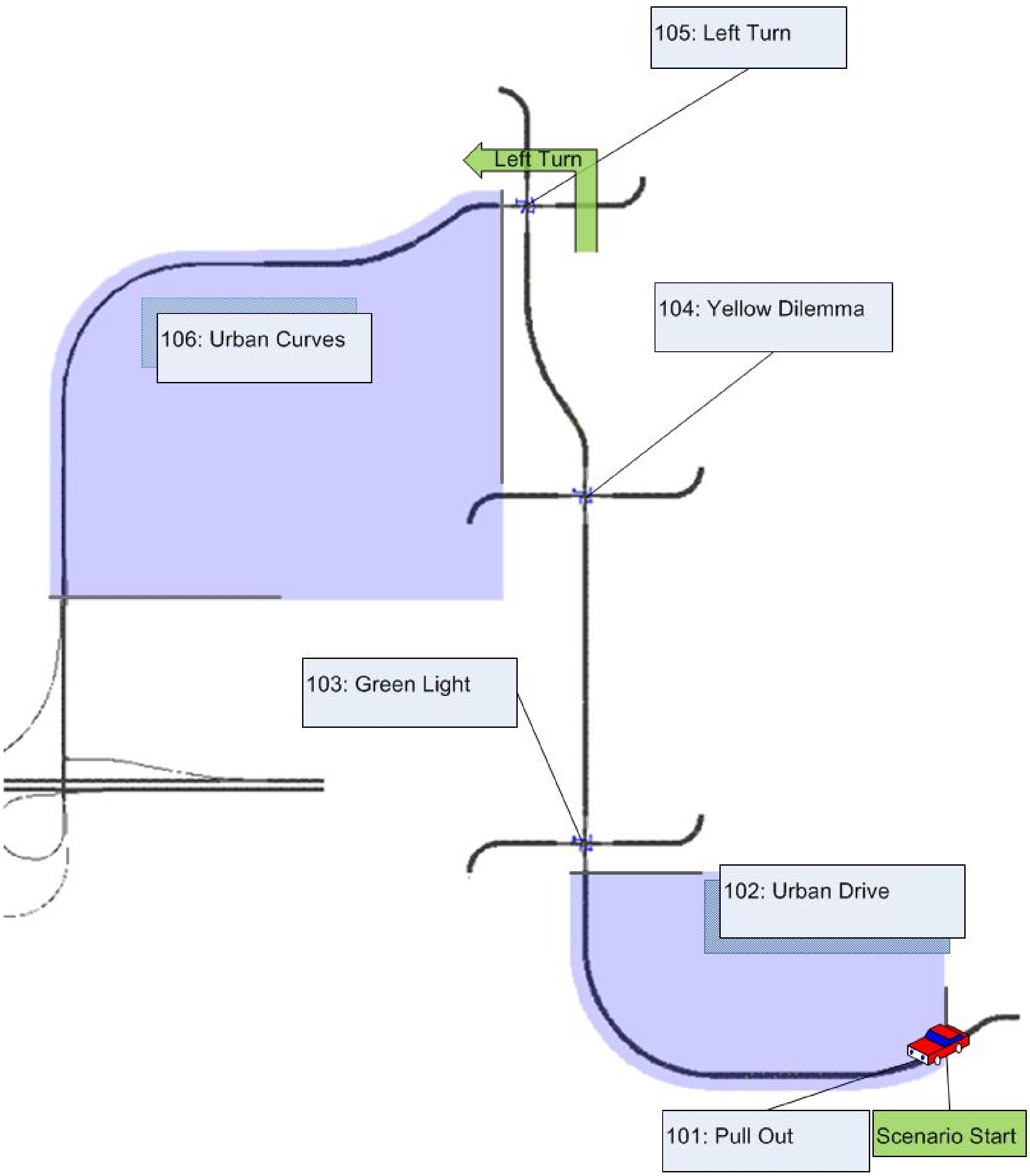


Figure 3 The urban segment of the drive.

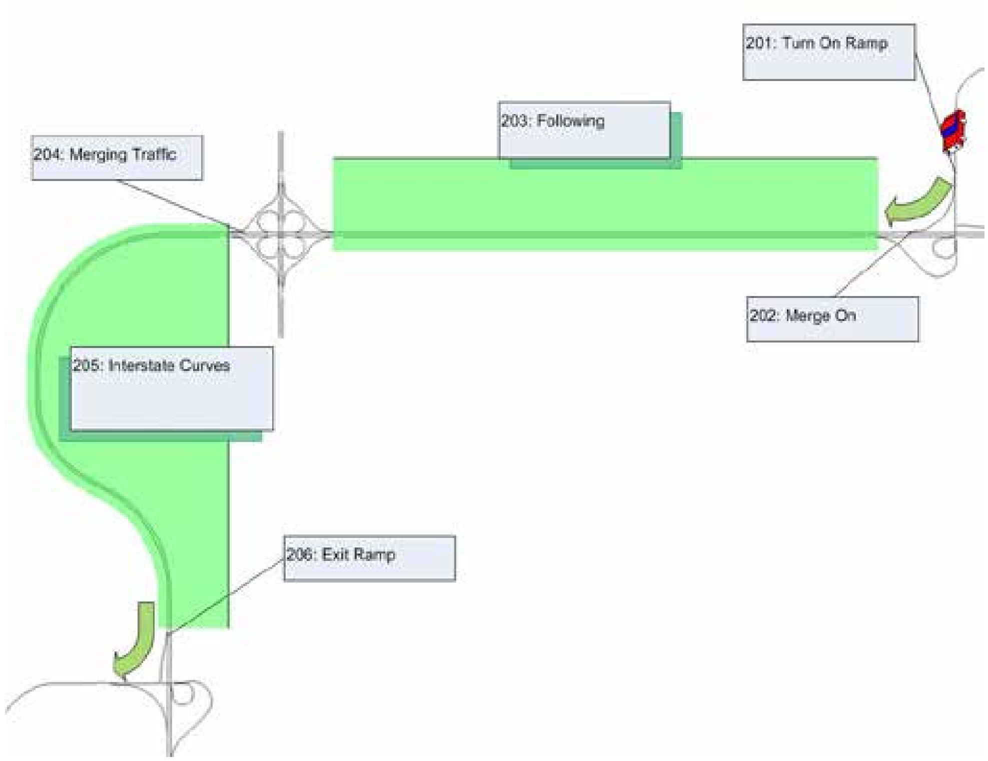


Figure 4 The interstate segment of the drive.

Figure 5 The rural segment of the drive.

Each participant engaged in the set of distraction tasks eight times during each distraction condition drive. The tasks occurred in an urban drive (2), in a less dense urban environment with curves and high density housing (1), while following trucks on the interstate (1), during interstate curves (1), in a transition between light and dark rural road (1), on a dark rural road (1), and on a gravel road (1). Table 6 summarizes the locations of the eight sets of three prompted tasks. Table 7 lists the events planned for the self-paced simple visual/manual "radio" task. Appendix G describes the data collection protocol in more detail, including the scenario events, distraction tasks, and their challenges to the driver.

Table 6 Summary of prompted tasks for the three segments of the drive.

Environment	Prompted Distractions	Events Included
Urban	1st Urban Interaction Menus Arrows Bug	Urban Drive (102) Green Light (103)
	2nd Urban Interaction Arrows Bug Menus	Yellow Light Dilemma (104)
	3rd Urban Interaction Menus Arrows Bug	Urban Curves (106)
Interstate	1st Interstate Interaction Menus Arrows Bug	Following (203)
	2nd Interstate Interaction Menus Arrows Bug	Interstate Curves (205)
Rural	1st Rural Interaction Menus Arrows Bug	Turn Off Ramp (301) Lighted Rural (302) Transition to Dark (303)
	2nd Rural Interaction Arrows Bug Menus	Dark Rural (304)
	3rd Rural Interaction Menus Arrows Bug	Gravel Road (306)

Table 7 Summary of planned engagement with the self-paced radio task.

Environment	Events Included
Urban	Pull-out (101) Green Light (103) Left Turn (105)
Interstate	Turn On Ramp (201) Merge On (202) Merging Traffic (204) Exit Ramp (206)
Rural	Dark Rural (304) Transition to Gravel (305) Gravel Road (306)

Distraction Tasks

Three secondary tasks were chosen to reflect distracting activities in which drivers currently engage, such as reaching to the backseat or adjusting the radio, as well as future distractions that a distraction detection algorithm should detect. Based on the current trajectory of innovations for in-vehicle internet-based technologies and the proliferation of wireless "carried-in" devices that drivers use in vehicles, the specific activities drivers might engage in are likely to change quickly in the coming years. For this reason, generic tasks were prioritized over specific tasks that are linked to a particular technology so that the results are more likely to accommodate the rapidly changing array of distractions that will confront drivers.

In the experimental design that follows, distraction type has three levels: a reaching task (bug), a visual/manual task (arrows), and a cognitive task (menu). The reaching task required drivers to reach to the back passenger side seat and follow a moving display with their finger. The visual/manual task was based on the arrow task used in the HASTE project (Engström et al., 2005), and presented drivers with a series of matrices of arrows on a three-inch diameter LCD touch screen. Participants had to review and discern whether or not a target arrow pointed in a particular direction was present in a field of distracter arrows. In the cognitive task, drivers traversed an interactive voice response menu that required them to respond to prompts from the system based upon information they were given concerning a fictional flight to determine if the flight was on time. The self-paced radio task did not contribute to the protocol sensitivity analysis or algorithm evaluation except to indicate task engagement throughout the drive. Appendix G contains a detailed description of each distraction task.

Experimental Design and Independent Variables

The protocol's overall sensitivity to distraction was analyzed using a 3×8×2 within-subjects experimental design that permitted comparison of driving performance during engagement in the three distraction tasks presented in the distraction drives and at comparable intervals during the baseline drives. The order of the baseline and distraction drives was counterbalanced across participants. For the analysis of task engagement, a 3x3 within-subjects experimental design was used to compare engagement in the three types of distraction task in the three driving environments during the distraction drive.

Table 8 and Table 9 show the composition of the three distraction task orders used in the experiment and their pairing with the scenario events, respectively. The order of the events and distraction tasks were the same for all participants. The order drivers perform tasks can affect their performance of the tasks, particularly if some tasks occur on the same stretch of road. To guard against this confounding, the tasks were counterbalanced using a Latin square as show in Table 8 and these counterbalanced orders were distributed across the events as shown in Table 9. Any evaluation protocol will need to consider such counterbalancing to avoid confounding the effect of the task with the road situation.

Table 8 Order of distraction tasks

Order	1st	2nd	3rd
I	Arrows	Bug	Menu
II	Bug	Menu	Arrows
III	Menu	Arrows	Bug
IV	Arrows	Menu	Bug
V	Menu	Bug	Arrows
VI	Bug	Arrows	Menu

Table 9 Pairing of events and order of distraction tasks

Driving Event	Start Event ID	Order
Urban Drive	102	III
Yellow Dilemma	104	I
Urban Curves	106	IV
Interstate – Truck Following	203	II
Interstate Curves	205	III
Lighted Rural	302	VI
Dark Rural	304	V
Gravel Rural	306	IV

Procedure and Participant Instructions

After providing informed consent (Appendix H), each participant completed a demographic questionnaire that assessed their driving history, habits of interaction with distracting devices, and beliefs in their own capability as safe drivers (Appendix I), then watched a PowerPoint presentation (Appendix J) that described the simulator cab and the tasks they were to perform during their drives. Participants then completed three drives; an eight-minute practice drive, an experimental drive performing distracting tasks, and another experimental drive with no distractions (the latter two in a counterbalanced order). The practice drive served to acclimate the participant to the simulator and to provide practice performing the distraction tasks.

After driving the urban, interstate, and rural segments, participants completed a visual-analog scale assessing their subjective workload and performance (lateral and longitudinal control) (Appendix K) for each distraction type. Standard simulator realism and wellness surveys (Appendices L and M) were also administered after the drives, as was a post-drive survey (Appendix N). A debriefing statement requesting that participants not discuss their participation with others until the end date for the data collection was provided to encourage participants to not share strategies they may have developed to perform the tasks while driving with other potential participants (Appendix O).

An incentive system (score) was used to encourage the participants to engage in the distracting tasks. The incentive was a function of overall task performance, including the time to initiate the distraction task, continuous attention to the task, and response accuracy. The experimenter provided scores out of 100 points to participants at the end of the three road segments in the drive with distraction tasks. Participants were instructed to complete as many tasks as possible while driving as they normally would. See Appendix G for more details about the incentive score.

Dependent Variables

Table 10 is a list of driver performance measures that might be sensitive to distraction. The table also summarizes the findings regarding their sensitivity to driver performance of the distraction tasks.

Table 10 Summary of metrics sensitivity to distraction

Metrics		Description	Sensitivity to distraction
1	lp_avg	average lane deviation	Not sensitive to distraction
2	lp_sd	standard deviation of lane position	Increases with distraction Not sensitive to urban environment (events 104 and 106)

Metrics		Description	Sensitivity to distraction
3	lpn_sd	modified standard deviation of lane position, not including mean	Increases with distraction Not sensitive to urban or interstate environment (events 102, 104, and 203) but sensitive to rural environment
4	sp_avg	average speed	Decreases with distraction Not sensitive to urban environment

Table 10 Summary of metrics sensitivity to distraction (Continued)

Metrics		Description	Sensitivity to distraction
5	sp_max	maximum speed	Sensitive only to two events: increases with distraction in urban environment (event 102), but decreases for rural environment (event 301)
6	sp_min	minimum speed	Decreases with distraction Sensitive to interstate environment
7	sp_sd	standard deviation of speed	Increases with distraction for interstate environment (events 203 and 205). Not sensitive to urban and rural environments
8	spn_avg	average speed, with respect to speed limit	Decreases with distraction Not sensitive to urban environment
9	spn_sd	modified standard deviation of speed, with speed limit subtracted	Increases with distraction Sensitive to interstate

Metrics		Description	Sensitivity to distraction
			environment
10	ster_freq	frequency content in the steering signal, measured at the 3dB power drop cutoff	Increases with distraction Sensitive to all events but not to rural turn off ramp (event 301)
11	steer_H	steering entropy, as defined in Boer, Rakauskas, Ward, & Goodrich, 2005. Baseline and reference data calculated from 60 sec windows in baseline drive	Increases with distraction Sensitive to all events
12	steer_sd	standard deviation of steering wheel angle	Increases with distraction Sensitive to all events except 301 (turn off ramp) and 304 (dark rural)
13	str_rr_avg	average steering reversal rate in a 6 second moving window. Based on 6 degree change in adjacent peaks, zero crossing not required	Increases with distraction Sensitive to all events
14	str_rr_sd	standard deviation of steering reversal rate in a 6 second moving window	Increases with distraction Sensitive to all events

Data Reduction and Verification

The data was analyzed with a four-phase approach similar to that used in the IMPACT project (Lee et al., 2010). The first phase was a review of the data to remove any spurious data points and resolve any inconsistencies using data visualization. The driving data from road segments in which the driver failed to engage in the task were treated as missing for all subsequent analyses.

The second phase used an analysis of variance within the SAS general linear models procedure to test the primary hypotheses concerning sensitivity to driver distraction and task engagement. The statistical model for sensitivity included drive type, scenario event and type of distraction as within-subject variables. The statistical model for task engagement included road segment and type of distraction as within-subject variables. The dependent measures associated with the self-paced visual/manual radio task were analyzed by road segment and drive type. Linear models for the variance decomposition analysis assessed the relationship between subjective and objective performance of the drivers. Overall, these analyses assess the sensitivity of the driving scenario and data analysis to distraction. Where appropriate post-hoc t-tests were used.

The third phase of analysis evaluated the real-time detection algorithms. The algorithms were evaluated using a signal detection paradigm to assess how many distractions are correctly identified (hits), how many are not detected (misses), how many instances are incorrectly identified (false alarms), and how many are correctly rejected (correct rejection). Hits and misses were defined by the presence or absence of distracting tasks from the distracted drive; whereas, false alarms and correct rejections were defined by corresponding locations in the baseline drive where the distractions were not present. The resulting analysis was portrayed on Receiver Operator Characteristic (ROC) plots. Similar to the algorithm analysis for the IMPACT project (Lee et al., 2010), the primary measure of algorithm performance is the Area Under the Curve in the ROC plot (AUC) with higher values indicative of better performing algorithms.

The fourth phase of analysis involved data exploration to identify ways to improve the algorithms. This began with a review of the residuals—instances of misses and false alarms where the candidate algorithms failed to predict the state of the driver. Additional data sources were explored to account for the residuals, which might also improve algorithm performance.

PROTOCOL SENSITIVITY TO DISTRACTION

Several factors were considered when assessing the sensitivity of the test protocol to the effects of the experimental distraction manipulations. The initial step was to ensure that the manipulation, in this case distraction, resulted in data collected with the driver in the correct state. After ensuring protocols were effective in collecting the necessary data, the next step was to examine the sensitivity of dependent measures in identifying impaired performance and the relative sensitivity of the events within the experimental protocol. The primary comparisons of interest were between the baseline and distraction conditions within a roadway segment. Because the dependent measures were not consistent across road segments, it was not possible to consider the interaction between road segments and distraction types. Follow-up analyses included examining the effect, if any, that the order of the baseline and distracted drives played in sensitivity to distraction effects, and how participants traded-off driving performance against task performance.

Task Engagement

To assess the extent to which drivers engaged as expected with the distraction tasks, a time history was plotted for each participant over the duration of the drive. Figure 6 shows a time history with horizontal lines indicating the portion of the drive over which the driver was engaged in each type of task, and vertical lines indicating when each task became available to the driver. For each task, engagement is measured from the driver's first interaction with the task to the last interaction with the task. In the event that the driver disengages with the task before it's completed, the line would stop at the point of the last engagement even though the task continued to be available. Open areas on the timeline for each participant indicate portions of the drive in which the driver was not engaged in a task. Open areas between a vertical line and the start of the horizontal line of the same color indicate task deferrals by the driver. The radio task display changed about every 7 seconds and if the person engaged with the task at every change, then engagement was interpreted as continuous.

The figure shows that most participants had a relatively uniform engagement in the tasks across the drive. The gaps in task performance were almost exclusively the result of disengagement with the radio-tuning task. As anticipated, drivers performed the primary distraction tasks (menu, bug, and arrows) with little if any deferral of either the initial or subsequent tasks.

Drivers 3 and 5 provide a good contrast in how drivers approached the radio task. Both drivers remained engaged throughout the primary distraction tasks; however, Driver 3 is continuously engaged throughout the drive with very short breaks in the radio task, whereas Driver 5 frequently disengaged from the radio task. This difference in engagement highlights a difference in approach adopted by participants. Some participants such as Driver 5 appeared to periodically adjust the radio when they noticed it deviating significantly from the nominal value; whereas more participants such as Driver 3 attempted to minimize the deviation from the nominal value by making frequent adjustments whenever a change was noticed. Individual differences associated with the degree of engagement with the radio task have substantial implications for assessing the effect of the mitigation systems.

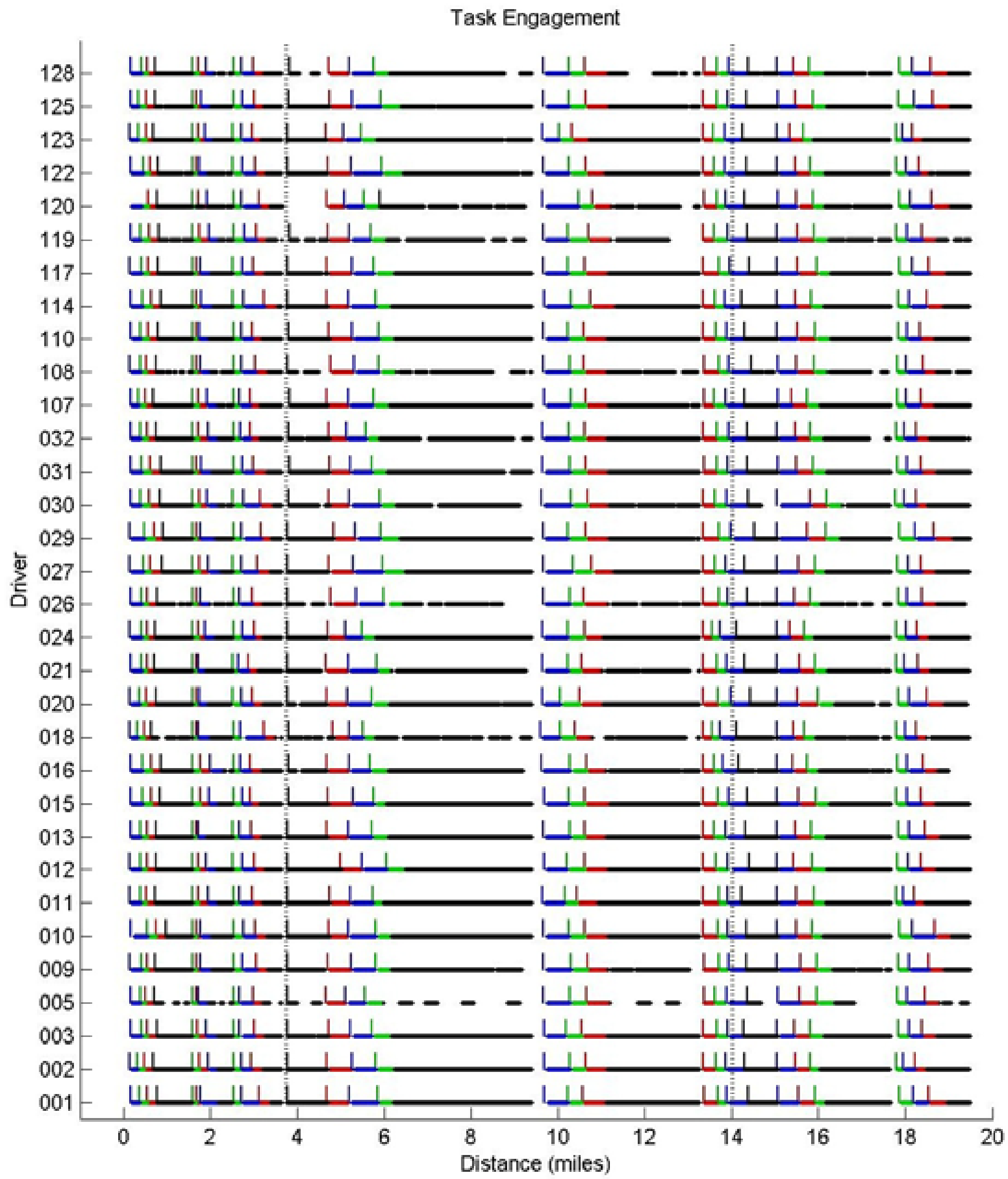


Figure 6 Task engagement by driver. Red indicates engagement with the bug task. Green indicates engagement with the arrows task. Blue indicates engagement with the menu task. Black indicates engagement with the radio task. Open areas indicate no engagement with any task.

The average deferral time of the three prompted tasks ranged from two to 3 seconds across the road segments, but with no systematic differences ($F(2, 604) = 0.9$, $p = 0.41$). Figure 7 shows that the variability associated with each mean is relatively consistent with the exception of menu task deferral in the urban area where the standard error is nearly twice that of the others. Even with these differences in variability, drivers tended to begin the primary distraction tasks as soon as they became available regardless of the task or driving environment.

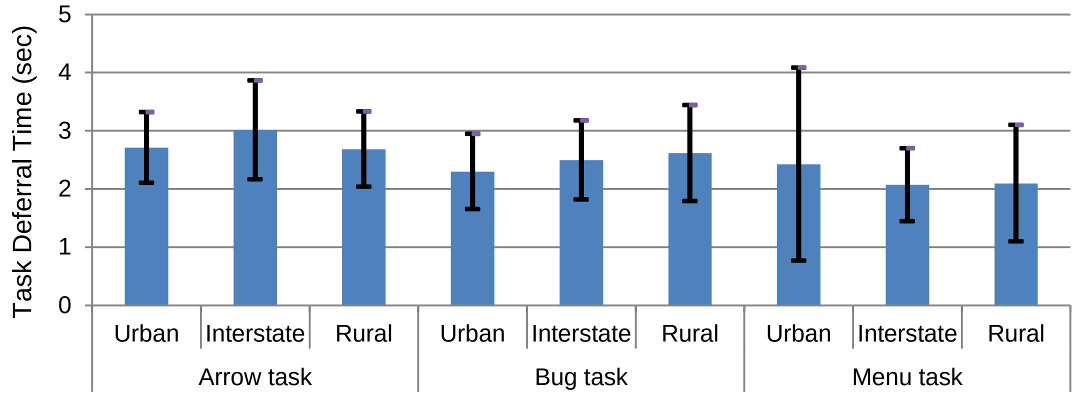


Figure 7 Task deferral time means with standard error bars for task by road segment.

Drivers also differed in how well they performed the tasks. They performed the arrow and bug task better than the menu task, see Figure 8. Drivers received scores above 80 percent for both the bug and arrows task across all three driving environments, but received scores in the 70 percent range for the menu task. Similar to task deferral, performance was not substantially different across the road segments ($F(2, 604) = 1.89$, $p = 0.15$).

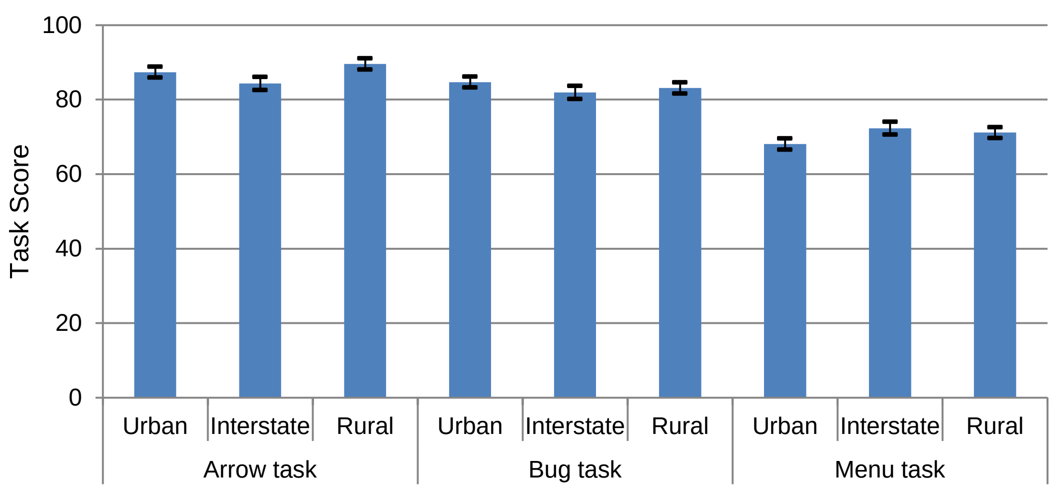


Figure 8 Task score means with standard error bars for task by road segment.

Examining task performance across events and tasks for individual drivers can provide additional insight into the variations in task performance. Figure 9 shows differences in task score performance based on z-scores for each driver for each task in each of the eight driving environments. There are more instances where the task performance is significantly worse than mean performance, but relatively few where performance is significantly better. The greatest deviation from the mean occurs in the urban area; whereas performance in the rural area is least variable. No consistent pattern of performance emerges when looking at individual drivers either across tasks or events. For example, when looking at Driver 125, task score is better than average for 7/8 of the menu task, 0/8 for the bug task, and 3/8 for the arrows task. This shows that drivers perform better on some tasks than others, such as the menu versus the bug task, and that this differential performance is not consistent across drivers.

Overall, several conclusions can be clearly drawn concerning task engagement:

- Drivers consistently engaged in the tasks over the duration of the drive;
- Drivers initiated the tasks within 2-3 seconds after they were presented;
- Drivers engaged in the tasks as evidenced by their task scores; and
- Task performance was not uniform across tasks, across drivers, or across drivers performing particular tasks.

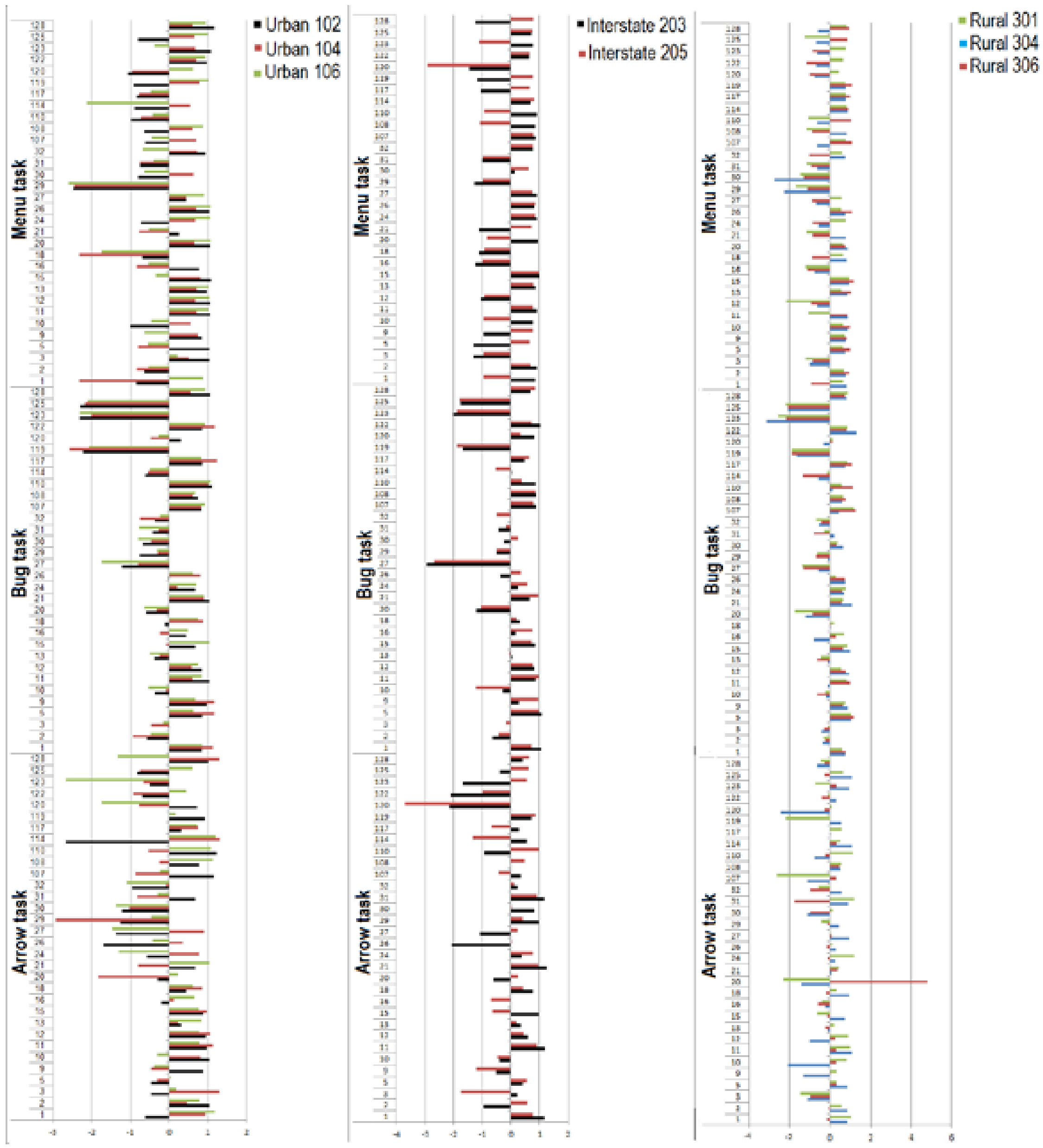


Figure 9 z-scores by participant and task across events.

Sensitivity to Distraction

Having established that drivers engaged in the distracting tasks fairly evenly over the course of the drive, we can now focus on whether the protocol demonstrated sensitivity to the effects of those distractions. Distraction effects can be assessed in terms of uniformity of the change across driving events and in terms of total impact on driving as indicated by the driving performance that are affected. Sensitivity of the protocol to distraction is important both to show distraction influenced driving performance and also to identify driving performance measures that might be used in an algorithm to detect distraction.

Both driving performance and gaze variables are considered as candidates for inclusion in distraction detection algorithms. This sensitivity analysis focused on both the relative value of different metrics in detecting distraction and the driving events most sensitive to those effects.

As stated above, Table 10 defined the measures that were analyzed to determine the effect of distraction, and also summarizes their sensitivity in terms of direction of change and situations in which they were not predictive. Measures that were most sensitive were those associated with steering, such as number of steering reversals, steering frequency, and steering entropy. Standard deviation of lane position also showed promise, but was not as sensitive as the steering measures. Figure 10 shows the relatively consistent effect of distraction on these four measures across the primary distraction events. Steering frequency was the only measure to provide an instance where distraction shifted the mean in the opposite direction (in the Rural Turn Off Ramp event), and in that case, the event was short and followed a right-hand turn. Distraction caused a clear decrement in performance on a variety of measures.

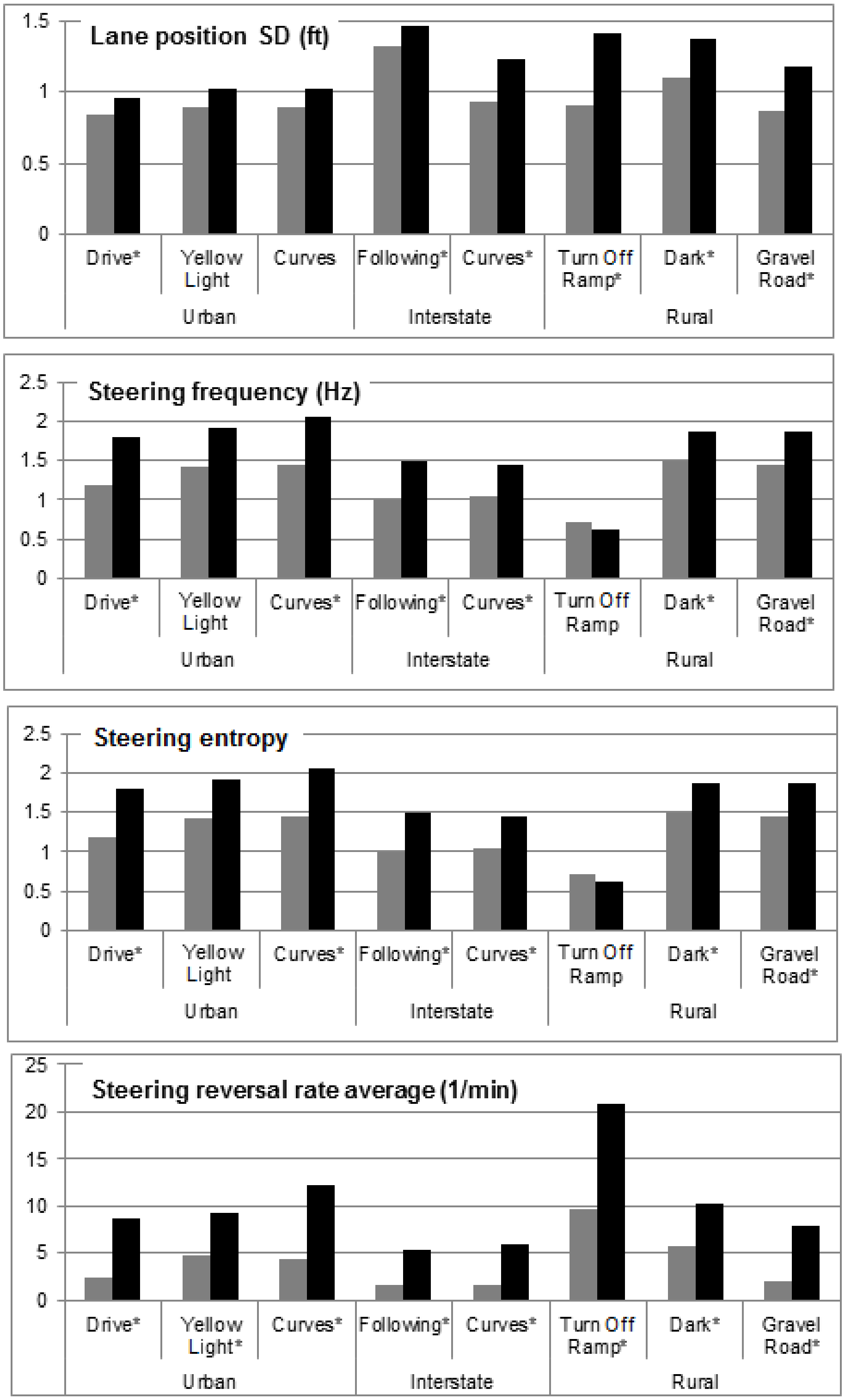


Figure 10 Effects of distraction on standard deviation of lane position, steering frequency, steering entropy and steering reversals. Black bars indicate the presence of distraction. Light grey is baseline driving with no distraction. Statistically significant differences are indicated (*).

Having established that several measures are sensitive to the effect of distraction, it is important to consider if this effect occurred uniformly across drivers. Table 11 shows that the effect of distraction is not consistent across all drivers. Only one measure, steering entropy, showed that distraction produced a trend in the expected direction in at least 60 percent of the drivers in each event. Steering metrics, e.g., steering entropy, steering reversal rate, and steering frequency, followed by standard deviation of lane position were the most consistent across drivers and scenario events. Speed metrics were the most inconsistent. The events on the interstate (Following and Curves) and on the rural paved road (Lighted Rural, Transition to Dark and Dark Rural) showed the most consistent effects of distraction. The effect of distraction in the urban environment was less consistent.

Table 11 Percent of drivers who showed the expected change in driving with distraction for each metric and event. The expected directions are indicated by an arrow in the first column. Unshaded cells indicate less than 60 percent of drivers. Light grey indicates 60-74.9 percent. Medium grey indicates 75-89.9 percent. Dark grey indicates > 90 percent

Metric / Event	Urban				Interstate		Rural				
	Drive	Green light*	Yellow light	Curves	Following	Curves	Turn off ramp	Lighted rural*	Transit. to dark*	Dark	Gravel road
lp_sd ↑	71.9	46.9	62.5	68.8	75.0	93.8	87.5	84.4	78.1	87.5	81.3
lpn_sd ↑	56.3	53.1	68.8	71.9	78.1	84.4	78.1	71.9	68.8	87.5	81.3
sp_avg ↓	37.5	46.9	59.4	37.5	71.9	84.4	84.4	71.9	90.6	75	53.1
sp_min ↓	59.4	46.9	50.0	37.5	68.7	78.1	40.6	78.1	90.6	62.5	50.0
sp_sd ↑	78.1	65.6	46.9	62.5	81.3	75.0	12.5	71.9	87.5	62.5	50.0
spn_avg ↓	37.5	46.9	59.4	37.5	71.9	84.4	84.4	71.9	90.6	75	53.1
spn_sd ↑	75.0	46.9	40.6	59.4	71.9	75.0	75.0	68.8	84.4	75.0	37.5
ster_freq ↑	96.9	62.5	87.5	100.0	100.0	100.0	9.4	100.0	62.5	100.0	93.8
steer_H ↑	96.9	78.1	96.9	100.0	100.0	96.9	93.8	100.0	87.5	100.0	100.0
steer_sd ↑	81.3	71.9	90.6	90.6	100.0	90.6	37.5	90.6	59.4	68.8	87.5
str_rr_avg ↑	96.9	40.6	93.8	100.0	100.0	96.9	93.8	65.6	62.5	96.9	100.0
str_rr_sd ↑	93.8	40.6	81.3	100.0	100.0	96.9	75.0	62.5	62.5	90.6	96.9

Note: * indicates events where the set of distraction tasks was continued from the previous event.

In summarizing the findings:

- The most sensitive metrics involved steering with steering entropy the most consistent.
- Standard deviation of lane position, while sensitive, was less effective in the urban environment.

- Urban events tended to provide less consistent effects than the interstate and rural environments.

Influence of Task Order on Metric Sensitivity

The analyses accounted for the effects of the order of the drives (distraction first versus baseline first) by including it in the statistical model; however, it is important to consider the role that order effects such as practice and fatigue play when studying distraction. Figure 11 shows that for some metrics, such as standard deviation of lane position (lp_sd) and steering entropy (steer_H), considering order removed noise from the data and provided a more accurate description of the influence of distraction. However, for other metrics such as standard deviation of steering wheel angle, (steer_sd) effects that appeared strong were in fact artifacts associated with the order of the trials. These results emphasize the need of an evaluation protocol to counterbalance the order of events and tasks.

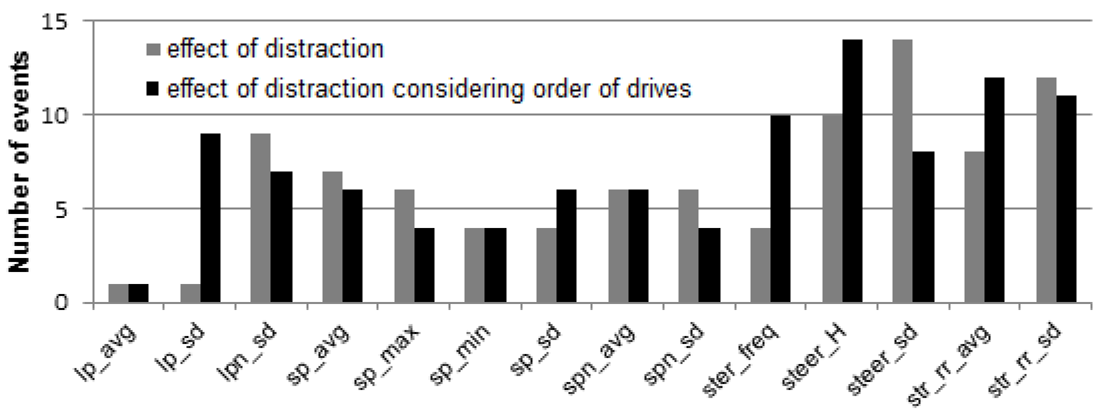


Figure 11 Summary of metric sensitivity across events.

Driving and Task Performance Trade-offs

The relationship between performance on the distraction tasks and driving performance can reveal the influence of distraction. To examine this relationship, normalized task performance was compared to the four most sensitive driving metrics from the prior analysis (standard deviation of lane position, steering frequency, steering entropy, and steering reversals). Better performance on distraction tasks is expected to reflect greater distraction, and greater distraction is expected to have a greater effect on driving performance.

A higher normalized task score represents increased task performance, whereas a higher driving performance value represents increased degradation. The expected relationship indicating that driving performance was traded for task performance would be represented by a line from bottom left to top right; however, this relationship was not found for any task in any of the three driving environments (Figure 12).

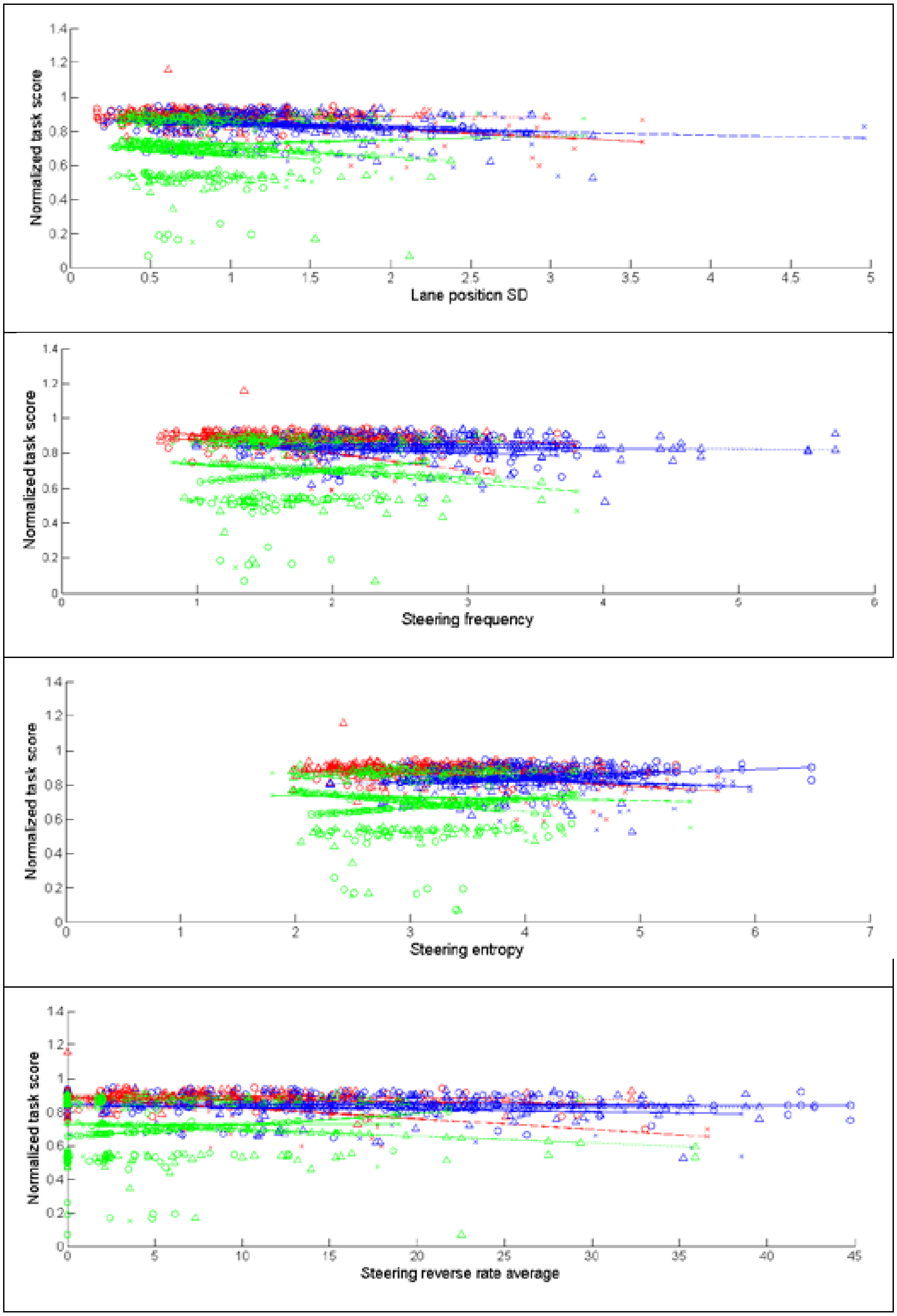


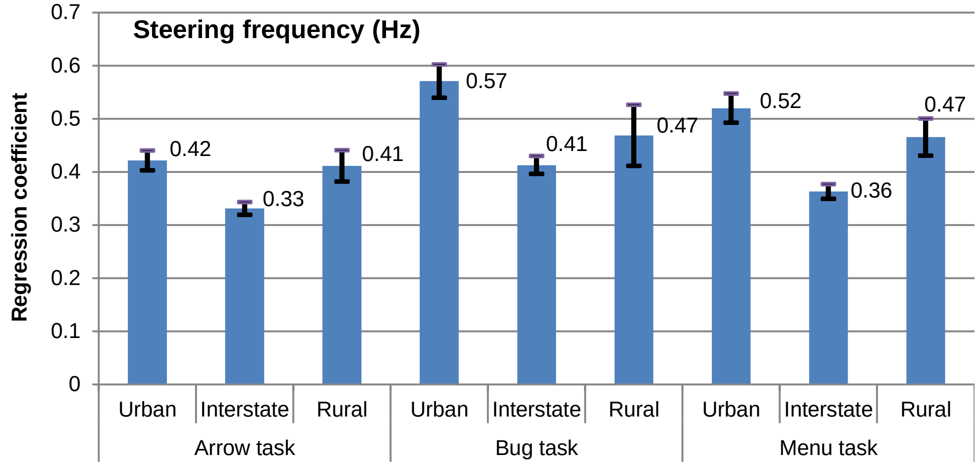
Figure 12 Relationship between driving performance metrics and task performance. The lines represent regressions for each set of data. Red represents arrows task. Blue represents bug task. Green represents menu task. Circles represent urban. Triangles represent rural. X's represent interstate.

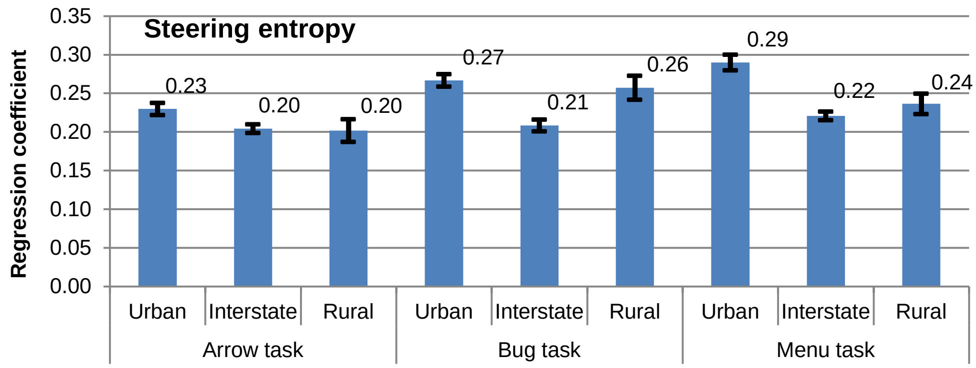
This leads to two potential conclusions, (1) that regardless of how well drivers perform a secondary task, other factors such as individual differences in driving style or secondary task capacity govern driving performance; (2) that there is a relationship between task performance at the individual level that is obscured when data are analyzed at the group level. This relationship warrants further inquiry because how participants tradeoff task performance and driving has important implications for assessing tradeoff strategies.

The preceding analysis examined the tradeoff at the group level. To evaluate how drivers trade off driving performance for task performance at the individual level, regression models were built for each driver: *normalized task score* = a + b * *driver performance*. An indicator of the tradeoff strategies that drivers may have developed – increased task performance at the cost of driving performance, increased task performance and driving performance, or increased driving performance at the cost of task performance, was estimated through regression coefficient b. The analysis showed that drivers do not increase their driving performance at the cost of task performance (Figure 13). However, the relationship between driving and secondary task performance is significantly different across drivers (Table 12). Road type (urban, interstate or rural) affected only the relationship between steering behavior and task performance.

Table 12 Summary of regression coefficient analysis results

Factors	DF	F-value			
		Lane position SD	Steering frequency	Steering entropy	Steering reversal rate average
Subject	31, 124	**2.24**	**2.93**	**3.82**	0.71
Road	2, 124	0.81	**10.80**	**17.32**	0.72
Task	2, 124	**61.88**	**20.89**	**28.60**	**4.87**
Subject x Road	62, 124	1.06	1.09	**1.44**	0.76
Subject x Task	62, 124	**2.22**	**1.58**	**1.81**	0.93
Road x Task	4, 124	2.05	1.07	**3.98**	0.78


Steering frequency (Hz)
Regression coefficient
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.42
0.33
0.41
0.57
0.41
0.47
0.52
0.36
0.47
Urban
Interstate
Rural
Urban
Interstate
Rural
Urban
Interstate
Rural
Arrow task
Bug task
Menu task


Steering entropy
Regression coefficient
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0.23
0.20
0.20
0.27
0.21
0.26
0.29
0.22
0.24
Urban
Interstate
Rural
Urban
Interstate
Rural
Urban
Interstate
Rural
Arrow task
Bug task
Menu task

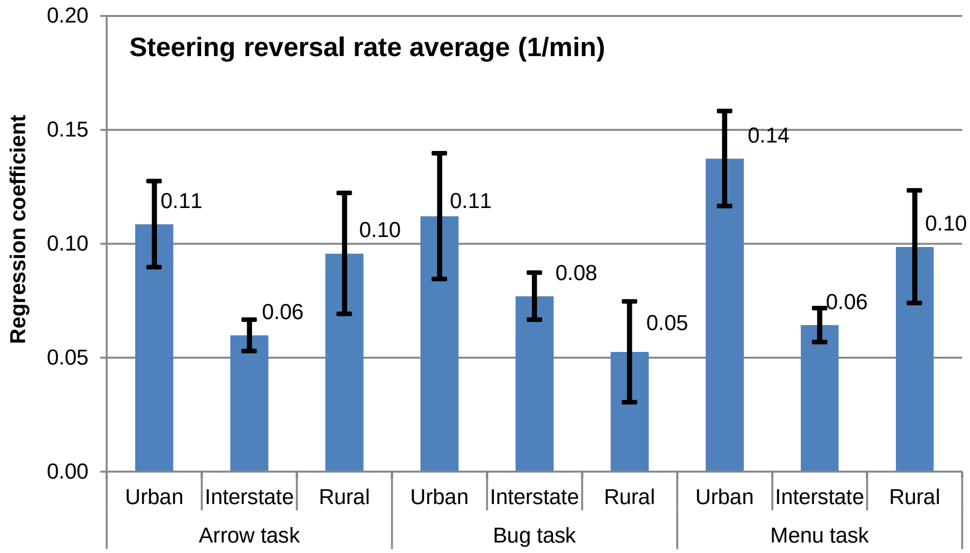


Figure 13 Regression coefficients' mean values (with standard error bars) for each task and road segment.

DISTRACTION DETECTION ALGORITHM EVALUATION

As demonstrated in Chapter 3, there are limitations to analytic assessments of distraction detection algorithms that empirical assessments are better able to address. This algorithm evaluation considers the performance of four progressively more complex algorithms to assess the degree to which additional variables and additional combinations of variables improve distraction detection performance. The algorithms all are based on gaze measures. They are assessed using metrics from signal detection theory that are applied to the algorithms to assess their robustness across different tasks and road segments. The evaluation concludes with an assessment of driving performance measures that might be combined with gaze measures to improve distraction detection performance.

Four Progressively More Complex Algorithms

Eyes off forward roadway (Klauer et al., 2006). As described in Chapter 3, this algorithm defines visual distraction as a cumulative glance away from the road of 2 seconds within a 6-second running window. Because the original algorithm defined the 6second window assuming an identifiable action (i.e., lead vehicle braking) occurred during the fifth second within the 6-second window (Klauer et al., 2006), it is unusable as a real time distraction detection algorithm. To compare it with the other algorithms under study, a six-second window is used to accumulate glance duration away from the forward roadway.

Risky visual scanning patterns (Donmez et al., 2007, 2008). This algorithm considers the history of glances and considers both the duration of the current glance and the cumulative glances away from the road to define risky visual scanning patterns. It has been used to provide feedback to mitigate distraction during drivers' interactions with in vehicle systems (Donmez et al., 2007, 2008). Levels of distraction are identified using Equation 1:

Equation 1

$$\gamma = \alpha * \beta_1 + (1 - \alpha) * \beta_2$$

where α is 0.2, β_1 is the current glance duration away from the road, and β_2 is the cumulative glance duration away from the road within the last 3 seconds (Donmez et al., 2007, 2008). γ (or risk) is considered moderate at values above 2 seconds, and high at values above 2.5 seconds (Donmez et al., 2007, 2008). However, the current implementation of the algorithm does not distinguish between moderate and high levels of distraction. Once the algorithm reaches a set threshold, the driver is considered distracted.

AttenD (Kircher et al., 2009; Kircher et al., 2009). Similar to the *risky visual scanning patterns* algorithm, the *AttenD* algorithm considers long glances away from the road as hazardous, and uses a buffer to represent the amount of road information the driver possesses. The buffer begins at 2 seconds and is decremented over time as the driver looks away from the field relevant for driving (FRD) (Kircher et al., 2009). The FRD consists of the intersection between a circle corresponding to a visual angle of 90° and the vehicle windows. The FRD excludes the mirrors. Once the driver looks back at the FRD, the buffer increases until a value of 2 seconds is reached. When the driver looks at the rear view mirrors or the speedometer (outside of the FRD) for one second or less, the buffer value remains constant. For such glances longer than one second, the buffer decreases at a value of one unit per second. In addition, when the driver looks at objects that are not relevant to driver safety (i.e., radio, cell phone, HVAC), the buffer decreases at a value of one unit per second. After the buffer has decreased, there is a latency of 0.1 seconds before the buffer increases again with attention to the road. The absolute minimum buffer value is zero and the absolute maximum buffer value is two. When the buffer reaches a value of zero, the driver is considered distracted.

Similar to the two aforementioned algorithms, the *AttenD* algorithm relies solely on eye movements to detect driver distraction. However, it differentiates itself as it distinguishes three glance categories: glances to the forward roadway, glances necessary for safe driving (i.e., at the speedometer or mirrors), and glances not related to driving. This algorithm also uses the world model of the vehicle instead of the more general approaches in the *Eyes off forward roadway* and *Risky visual scanning patterns* algorithms (Kircher et al., 2009).

multidistraction detection (Victor, 2010). The *multidistraction detection* algorithm was developed to identify distraction in real time and to give drivers alerts that correspond to risky scanning behavior associated with both visual and cognitive distraction. It relies on the notion that drivers should spend a certain amount of time glancing towards the road center area. The road center area is defined as a circle of 10 degrees radius centered on the road center. (The road center is defined as the most frequent gaze angle during normal driving.) With the road center area identified, it is possible to give three types of alerts: (1) single long glance—when drivers glance away from the road for 3 seconds; (2) visual distraction—when drivers' glances fall below a percent road center (PRC) of 60 percent within a 17.3-second running window; (3) cognitive distraction—when drivers glances rise above a PRC of 92 percent within a 60-second running window. The running PRC windows were initialized when the vehicle speed reaches 50 kilometers/hour (31.1 miles/hour). When the speed falls below 47 kilometers/hour (29.2 miles/hour), the PRC windows are reset to 80 percent and the PRC calculation is paused until the speed reaches 50 kilometers/hour again. In addition to the two PRC windows (for visual and cognitive distraction), a third PRC window is also calculated to improve reliability; it is called the visual time sharing (VTS) PRC window. This separate PRC calculation relies on a 4-second running window. When a sink is detected (a PRC value below 65%) followed by a rise (a PRC value above 75%), then the visual and cognitive distraction PRC windows are reset to 80 percent. The VTS window is also reset (to 75%) when the vehicle speed falls below 47 kilometers/hour (Victor, 2010). The *multidistraction detection* algorithm is the only algorithm to detect and distinguish between both visual and cognitive distraction.

A MODIFIED MULTIDISTRACTION-DETECTION ALGORITHM FOR THE NADS

The multidistraction detection algorithm was selected for use in this study; however, it was modified for use in the NADS-1 simulator in the context of the experiment. There are several reasons for using a modified, rather than the original algorithm. First and foremost is the fact that all the details of the commercial algorithm were not available for implementation. Secondly, constraints of the experiment imposed by the scenario events motivated some changes in algorithm parameters. Finally, the sensor suite feeding the algorithm was expanded to include not only an eye tracker, but also a separate head tracker and seat pressure sensors as well, with the intention of improving the robustness of the system. A flowchart of the modified algorithm is shown in Figure 14.

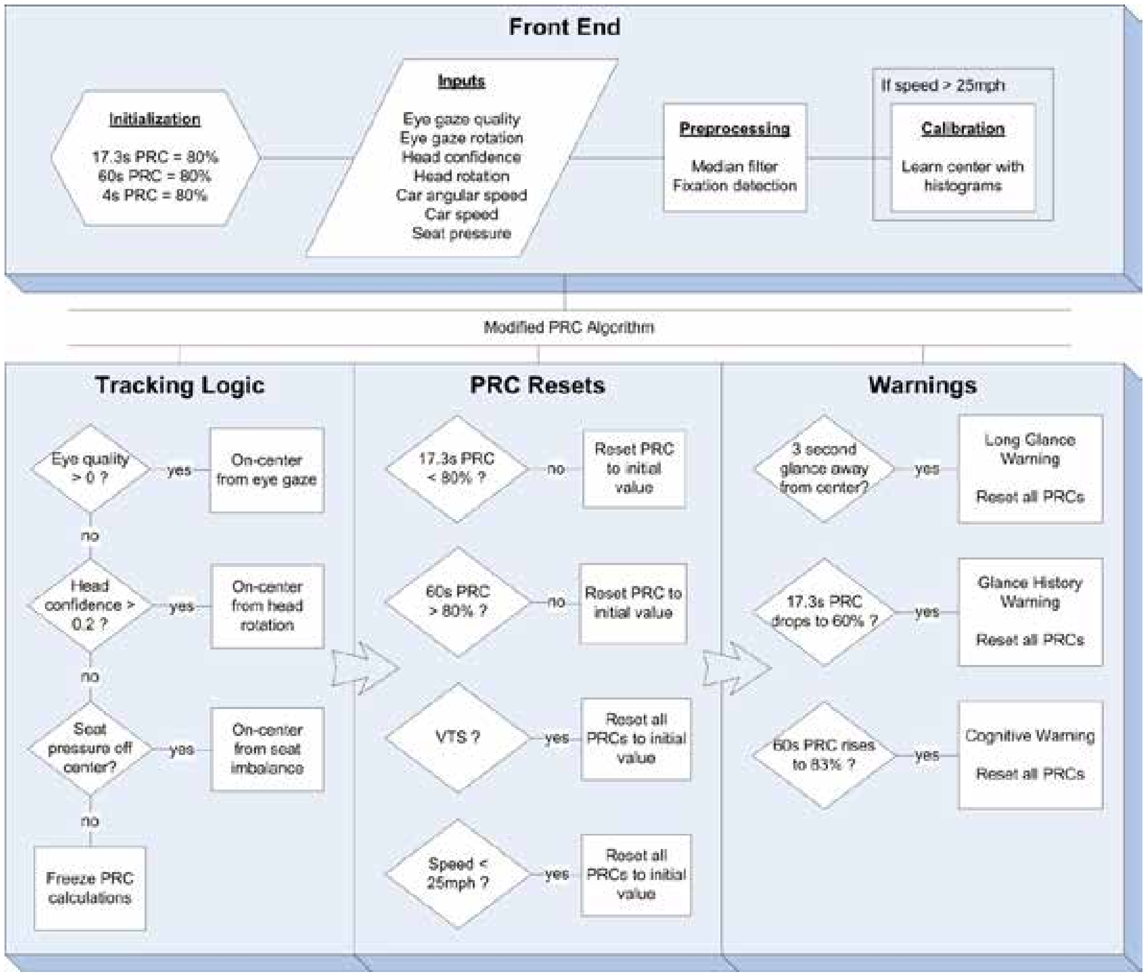


Figure 14 Modified NADS algorithm flowchart.

The vehicle cab used in this protocol was equipped with an eye tracker, a separate head tracker, and seat pressure sensors. The modified algorithm was developed to incorporate data fusion between the various sources of data. Eye tracker gaze data was used if it was of sufficient quality. If not, the system used the head tracker data. If neither the eye tracker nor the head tracker was accurate, then the seat sensor data was used. The difference between the left and right seat sensors was used to determine lateral shifts of the driver associated with reaching. If such a shift is sensed above a set threshold, it was interpreted as a glance away from the road center. If no weight shift was sensed, then no conclusion was drawn and the algorithm paused until tracking resumed or there was a weight shift.

The size of the road center cone was enlarged from 10 to 20 degrees when the head tracker was used. Additionally, the car's angular yaw rate was used to shift the center cone to the left or right. Finally, the size of the cone was also enlarged at the low end of the working speed. Each of these modifications to the center cone size and position were also intended to enhance system robustness in situations where the driver's gaze would be expected to leave the road center in the course of normal driving.

Figure 14 shows several states in which PRC windows are reset. Resets are used as a mechanism to saturate the maximum value of the medium length window (17.3 s) to 80 percent, and the minimum value of the long window (60 s) to 60 percent. Resetting all PRC windows is also the mechanism used to "freeze" the algorithm in a known, initial state when the speed drops below the minimum threshold. Additionally, whenever a VTS event is detected, all the PRC windows are reset.

The same three warnings are used in the modified algorithm. Whenever a warning is triggered, the PRC windows are reset. Notice that there are two parameter changes from the original algorithm. The warning threshold for the cognitive warning has been lowered to 83 percent. The reason for this was to have a greater chance of reaching the threshold during the cognitive distraction task in the experiment. The second change was to lower the speed threshold. The speed threshold actually consists of two speeds separated by a small amount. This *hysteresis* prevents the algorithm from quickly switching on and off as the speed crosses a single value. The speed thresholds for switching on and off have been lowered from [31.1 mph, 29.2 mph] to [25 mph, 23 mph]. The reason for this change was that the first section of each drive took place in an urban environment that had a lower speed limit.

The implementation of the finer points of the algorithm, such as data preprocessing and calibration, were created from descriptions available in theses (Victor, 2005; Larsson, 2003) and a patent application (Larsson & Victor, 2008). The degree to which all details match the commercial applications was constrained by available documentation and the time allowed for development.

Signal Detection Theory Criteria for Algorithm Assessment

The objective of the analyses was to assess how well the algorithms distinguish between distracted drivers and non-distracted drivers, and to demonstrate a protocol that can be used to assess other algorithms. To assess how well each algorithm predicts distraction (either visual or cognitive), criteria from signal detection theory were used because it provides a theoretical basis for discriminating between a signal and noise (Stanislaw & Todorov, 1999). In this case, the "signal" is when the driver is distracted and "noise" is when the driver is not distracted. Each algorithm can either indicate distraction or not (a bimodal classification). Taken together, there are four possible outcomes for each measurement: a true positive—the algorithm correctly indicates distraction; a true negative—the algorithm correctly indicates no distraction; a false positive—the algorithm indicates distraction when it is in fact not present; and a false negative—the algorithm does not indicate distraction when it in fact is present. In each instance, a true instance of distraction is defined as when a participant is engaged in a task, i.e. the time period from the beginning to the end of a task. Future definitions of distraction should describe situations with problematic scanning behavior or impaired responses to events. However, as evaluation protocol, distraction defined by task engagement provides a well-defined basis for assessing distraction detection algorithms.

Each algorithm was evaluated at different distraction thresholds (the multidistraction detection algorithm used in this analysis was the version modified by NADS). For the *Eyes off forward roadway* algorithm, values from 0 to 6 seconds with 0.3 second intervals are used as thresholds. The *Risky visual* *scanning patterns* algorithm was assessed at 0.1 seconds to 6 seconds, with 0.3 second intervals. The *AttenD* algorithm was assessed at thresholds ranging from 0 seconds to 2 seconds with 0.1 second intervals. The *multidistraction detection* visual distraction alert was assessed at thresholds ranging from 0 to 80 percent in 4 percent intervals. The *multidistraction detection* cognitive distraction alert was assessed at thresholds ranging from 80 to 100 percent in 1 percent intervals. These thresholds reflect the range of distraction recorded by each algorithm.

The algorithm predictions at each distraction threshold form the basis for the receiver operator characteristic (ROC) curves. These are plotted for each of the algorithms using the true positive or "hit rates" and the false positive or "false alarm rates" for each of the thresholds (Fawcett, 2006; Stanislaw & Todorov, 1999). The true positive rate is calculated by dividing the number of true positives by the sum of the true positives and false negatives. The false positive rate is calculated by dividing the number of false positives by the sum of false positives and true negatives. As the protocol is densely packed with distraction events, both the number of true positives and false negatives are only accumulated during the distracted drive. Accordingly, only false positives and true negatives were accumulated during the baseline drive. The false positive rate is calculated by examining the portion of the baseline drive that corresponds with the task engagement period during the distracted drive, as indicated by distance. During each task engagement period, if an instance of a true positive was identified (during the distracted drive) or an instance of a false positive was identified (during the baseline drive), then the entire period was classified as a hit (for the distracted drive) or a false alarm (for the baseline drive). As a result, since there are 24 task engagement periods, there are 24 resulting periods used to calculate the true positive and false positive rates.

The area below the ROC curve (AUC) is calculated to determine if the algorithm correctly classified distraction at a probability higher than that due to chance (Fawcett, 2006). Chance performance corresponds to a curve along the diagonal of the graph, with an AUC of 0.50. The algorithm with the highest AUC value indicates the approach that correctly identifies distraction with the highest probability. The AUC is a common classification metric that is insensitive to class skew that can bias simple measures of percent correctly classified. In addition to the ROC curves and AUC values, algorithms are assessed according to their positive predictive value (PPV indicates precision) and their accuracy.

Receiver Operator Characteristics by Task

To evaluate how each algorithm detected visual and cognitive distraction associated with different tasks, each of the three tasks was analyzed separately. Figure 15 shows the ROC plots for each algorithm for each task.

All four algorithms performed similarly in detecting engagement with the bug task. This reflects the nature of the bug task, which requires participants to look to the backseat. All four algorithms performed significantly better than chance (indicated by the dashed line in Figure 15). Even with high true positive rates near 0.75, the corresponding false positive rates were no larger than 0.25. This indicates that each of the algorithms can accurately identify distraction at a reasonable rate (0.75) with false alarms occurring at a rate of only 0.25.

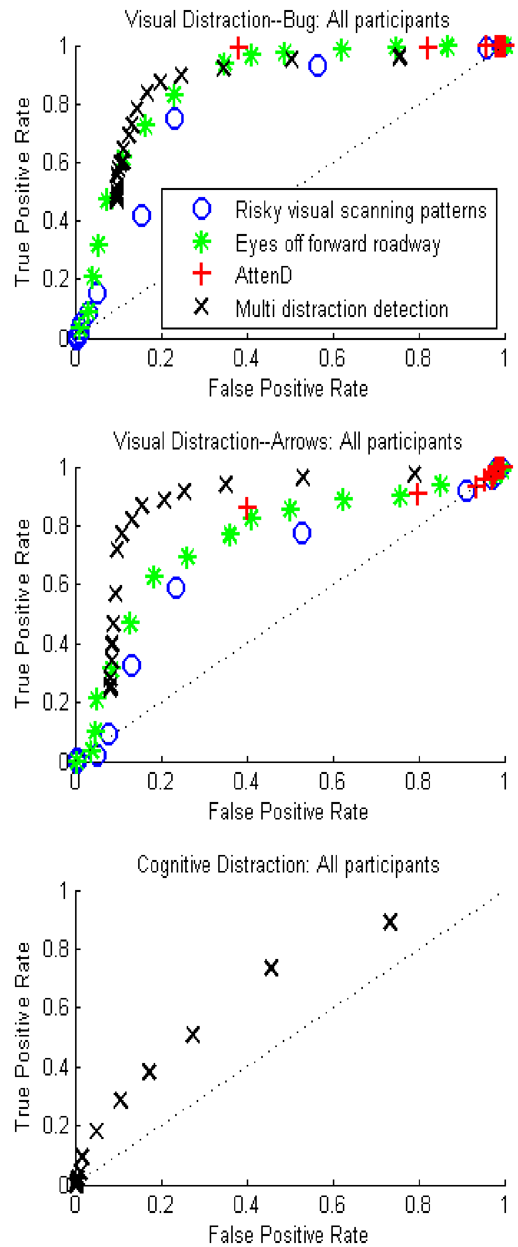


Figure 15 ROC plots for each algorithm separated by task; only the multidistraction detection algorithm was intended to detect cognitive distraction.

All four algorithms again performed much better than chance with the arrows task. At the same time, compared to the bug task, substantial differentiation between the algorithms can be seen. Here, the *multidistraction detection* distinctly outperforms the other algorithms. The *AttenD* algorithm constantly yields high true positive rates, but at the expense of high false alarm rates, which exceed 0.4. The two least complex algorithms (*Eyes off forward roadway* and *Risky visual scanning patterns*) perform in a similar manner.

Because the *multidistraction detection* algorithm was the only algorithm intended to predict cognitive distraction, there was no comparison. Comparing the results of the *multidistraction detection* algorithm to the other three algorithms would not yield equivalent comparisons as the other three algorithms are essentially detecting eyes off the road, which generally does not occur with cognitive distraction. Detection of cognitive distraction is greater than chance (indicated by the dashed line) and consistently yields higher true positive rates than false positive rates at each threshold.

A clear difference can be seen between the algorithms because not all of the ROC curves reach points where both the true and false positive rates are one and zero. Failure of the ROC curve to reach the upper or lower corner indicates that even at the extreme thresholds, the algorithm will not catch or misclassify all instances of distraction.

Further differentiation between the algorithms can be seen in the area under the ROC curve (AUC; Table 13). All algorithms performed well above chance (AUC=0.5). In addition, the AUC values analytically confirm what Figure 15 shows: the *multidistraction detection* algorithm generally outperforms the other algorithms.

Table 13 Area under the curve for each of the algorithms and each of the task (RVSP=Risky visual scanning patterns; EOFR=Eyes off forward roadway; MDD=multidistraction detection).

		Algorithms			
		RVSP	EOFR	AttenD	MDD
Tasks	Arrows	0.669	0.752	0.711	0.870
	Bug	0.779	0.866	0.803	0.860
	Menu	n/a	n/a	n/a	0.675

Because AUC values do not completely describe algorithm performance (i.e., the AUC does not distinguish between an algorithm that suffers from a high miss rate or a high false alarm rate), accuracy and precision can be useful metrics (Table 17 and Table 18). Additionally, for production systems which use a particular threshold that makes the AUC not relevant, the system would be described by accuracy and precision.

Table 14 Accuracy for each algorithm and task (RVSP=Risky visual scanning patterns; EOFR=Eyes off forward roadway; MDD=multidistraction detection; TPR=True positive rate; FPR=False positive rate).

		Algorithms			
		RVSP	EOFR	AttenD	MDD
Maximum accuracy					
Tasks	Arrows	0.676 TPR=0.587 FPR=0.234	0.721 TPR=0.625 FPR=0.184	0.728 TPR=0.857 FPR=0.401	0.855 TPR=0.867 FPR=0.156
	Bug	0.761 TPR=0.753 FPR=0.230	0.799 TPR=0.829 FPR=0.230	0.805 TPR=0.992 FPR=0.822	0.836 TPR=0.871 FPR=0.199
	Menu	n/a	n/a	n/a	0.639 TPR=0.734 FPR=0.457
Minimum Accuracy					
Tasks	Arrows	0.486 TPR=0.023 FPR=0.051	0.493 TPR=0.992 FPR=1	0.495 TPR=0.961 FPR=0.971	0.497 TPR=1 FPR=1
	Bug	0.495 TPR=1 FPR=1	0.495 TPR=1 FPR=1	0.495 TPR=1 FPR=1	0.495 TPR=1 FPR=1
	Menu	n/a	n/a	n/a	0.495 TPR=1 FPR=1

Accuracy is a measure of how often the algorithm gives a correct or true classification and was calculated using Equation 2:

Equation 2

$$Accuracy = \frac{True\ postives + True\ negatives}{True\ positives + False\ negatives + True\ negatives + False\ positives}$$

Precision is a measure of exactness in terms of how correct the algorithm is in classifying distraction and was calculated using Equation 3:

Equation 3

$$Precision = \frac{True\ positives}{True\ positives + False\ positives}$$

In the case of precision, values were aggregated across only those participants with instances of both true and false positives as calculating precision without any true or false positives will result in values of infinity.

Based on these values, both the *Eyes off forward roadway* and *Risky visual scanning pattern* algorithms typically have lower accuracy and precision. The *AttenD* and *multidistraction detection* algorithms perform in a like manner, with the *multidistraction detection* algorithm having higher accuracy and precision in most instances.

Table 15 Precision for each algorithm and task (RVSP=Risky visual scanning patterns; EOFR=Eyes off forward roadway; MDD=multidistraction detection; TPR=True positive rate; FPR=False positive rate).

		Algorithms			
		RVSP	EOFR	AttenD	MDD
Maximum precision					
Tasks	Arrows	0.678 TPR=0.587 FPR=0.234	0.722 TPR=0.693 FPR=0.262	0.673 TPR=0.857 FPR=0.401	0.906 TPR=0.719 FPR=0.098
	Bug	0.777 TPR=0.753 FPR=0.230	0.880 TPR=0.615 FPR=0.113	0.736 TPR=0.992 FPR=0.383	0.887 TPR=0.780 FPR=0.145
	Menu	n/a	n/a	n/a	0.662 TPR=0.734 FPR=0.457
Minimum precision					
Tasks	Arrows	0 TPR=0 FPR=0.004	0 TPR=0 FPR=0.004	0.498 TPR=0.965 FPR=0.975	0.499 TPR=1 FPR=1
	Bug	0 TPR=0 FPR=0.004	0.172 TPR=0.031 FPR=0.011	0.498 TPR=1 FPR=1	0.498 TPR=1 FPR=1
	Menu	n/a	n/a	n/a	0 TPR=0 FPR=0.004

Receiver Operator Characteristic by Road Segment

In addition to evaluating the algorithms by task, they were also evaluated by road segment. For this evaluation, both the arrows and bug tasks were grouped together into visual distraction and cognitive distraction was still analyzed separately. Only 31 participants were included in these analyses as one participant was removed due to incomplete task performance data. Figure 16 shows the ROC plots for each algorithm separated by road segment and task type.

While detecting visual distraction during the urban section, all four algorithms performed similarly. The speed limit in the urban section was slightly above the speed limit threshold of the multidistraction detection algorithm. The three simpler algorithms did not utilize the speed threshold; however their performance was still similar to the multidistraction detection algorithm. The speed threshold used in the *multidistraction detection* algorithm is a design feature that reduces false alarms, but also increases misses. With that said, each algorithm still performs reasonably well and well above chance.

Differentiation between the algorithms can be seen in the interstate section while detecting visual distraction. As the speed limit during this road section is significantly higher than the speed threshold, all algorithms were capable of detecting distraction. The *multidistraction detection* algorithm performed quite well by yielding true positive rates near 1.0 while corresponding false alarm rates were only at 0.2. The *AttenD* and *Eyes off forward roadway* algorithms also performed well, yielding true positive rates near 0.8 while corresponding false positive rates were at 0.4.

The rural segment revealed a similar ordering of algorithm performance. At the same time, each algorithm performed better in the rural environment than in both the urban and interstate sections. This could reflect the speed threshold during the urban segment and the fact that only six tasks were presented during the interstate segment (as compared to nine tasks during the urban and rural segments). The visual demands of the environment, such as pedestrians, traffic, and storefronts, drew drivers' eyes away from the forward roadway and were another factor contributing to the poorer performance in the urban segment. These glances away from the forward road might contribute to false alarms.

The *multidistraction detection* algorithm's ability to detect cognitive distraction resulted in similar performance over the three road segments. This was confirmed by examining the area under the curve values presented in Table 16 as all values were between 0.6 and 0.7.

However, the area under the curve values show clear differences between the algorithms when detecting visual distraction. The *multidistraction detection* algorithm consistently yielded values above 0.8. The *Eyes off forward roadway* algorithm also performed well across the range of road segments and yielded values between 0.7 and 0.9. The *AttenD* algorithm had area under the curve values between 0.7 and 0.8 while the *Risky visual scanning patterns* algorithm had values between 0.6 and 0.8. The last line of Table 16 shows the range of algorithm performance across the three segments varied relatively little and that *AttenD* and *multidistraction detection* were the most robust.

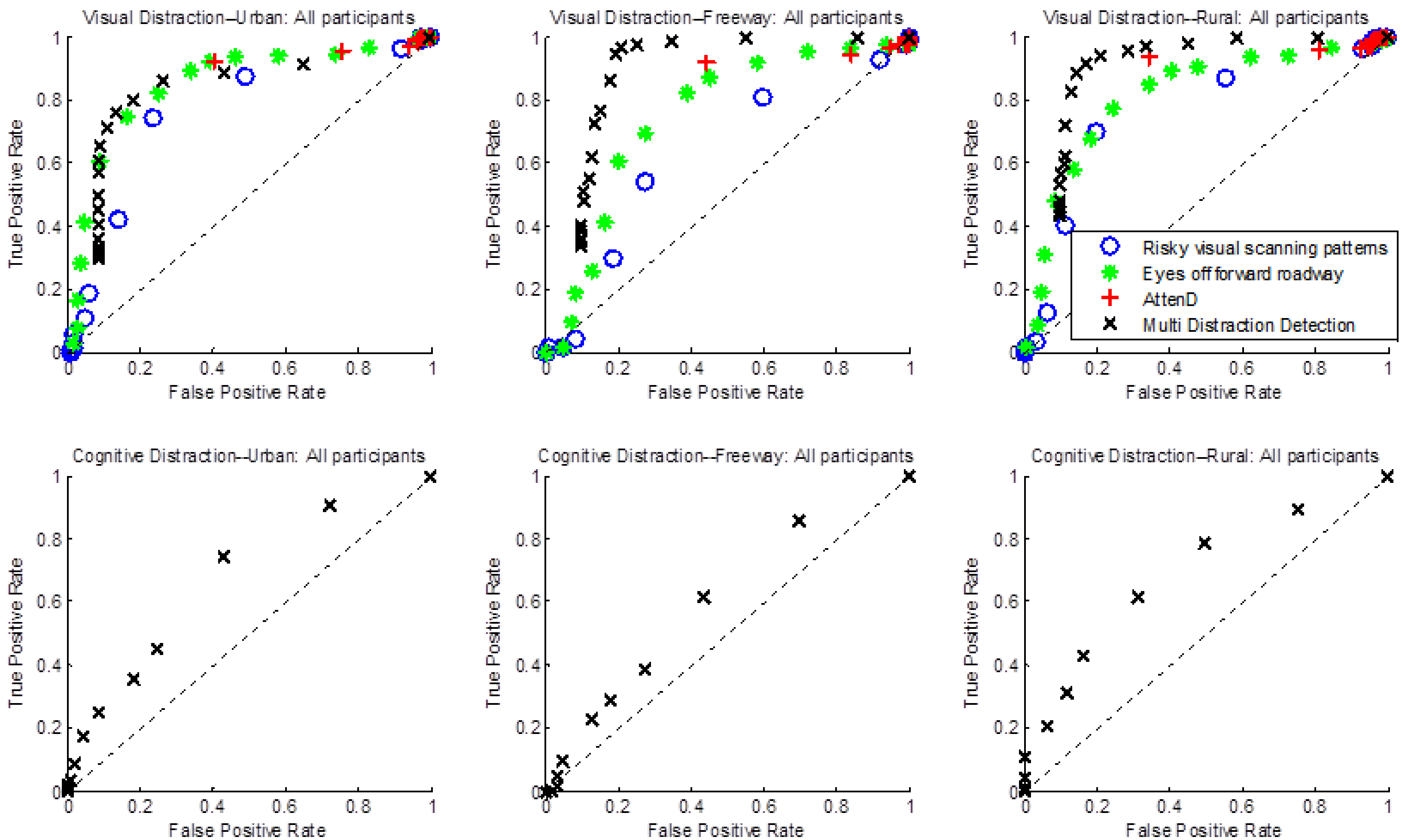


Figure 16 ROC plots for each algorithm separated by road segment and task; only the multidistraction detection algorithm was intended to detect cognitive distraction.

Table 16 Area under the curve for each algorithms and road segment (RVSP=Risky visual scanning patterns; EOFR=Eyes off forward roadway; MDD=multidistraction detection)

		Algorithms				
		RSVP	EOFR	AttenD	MDD-visual	MDD-cognitive
Road	Urban	0.764	0.851	0.754	0.830	0.682
	Interstate	0.630	0.749	0.727	0.886	0.614
	Rural	0.756	0.813	0.786	0.896	0.697
	Range	0.126	0.102	0.059	0.066	0.083

The accuracy and precision for each of the algorithms (Table 17 and Table 18) follows the results for the area under the curve. The *multidistraction detection* algorithm consistently yields the highest accuracy and precision. Although the *Eyes off forward roadway* algorithm had promising area under the curve values, the *AttenD* algorithm often yields better accuracy and precision. The *Risky visual scanning patterns* algorithm consistently yielded the lowest values for both accuracy and precision.

Table 17 Accuracy for each algorithm and road segment (RVSP=Risky visual scanning patterns; EOFR=Eyes off forward roadway; MDD=multidistraction detection; TPR=True positive rate; FPR=False positive rate)

		Algorithms				
		RVSP	EOFR	AttenD	MDD-visual	MDD-cognitive
Maximum accuracy						
Road	Urban	0.753 TPR=0.742 FPR=0.237	0.793 TPR=0.747 FPR=0.161	0.761 TPR=0.925 FPR=0.403	0.812 TPR=0.758 FPR=0.124	0.656 TPR=0.742 FPR=0.430
	Interstate	0.633 TPR=0.540 FPR=0.272	0.718 TPR=0.823 FPR=0.387	0.738 TPR=0.919 FPR=0.444	0.879 TPR=0.968 FPR=0.210	0.589 TPR=0.855 FPR=0.694
	Rural	0.750 TPR=0.699 FPR=0.199	0.766 TPR=0.774 FPR=0.242	0.796 TPR=0.935 FPR=0.344	0.871 TPR=0.914 FPR=0.172	0.651 TPR=0.613 FPR=0.312
Minimum Accuracy						
Road	Urban	0.495 TPR=0 FPR=0.011	0.500 TPR=1 FPR=1	0.500 TPR=1 FPR=1	0.503 TPR=1 FPR=0.995	0.500 TPR=0 FPR=0
	Interstate	0.480 TPR=0.040 FPR=0.081	0.483 TPR=0.016 FPR=0.048	0.488 TPR=0.976 FPR=1	0.500 TPR=1 FPR=1	0.492 TPR=0 FPR=0.016
	Rural	0.500 TPR=0 FPR=0	0.500 TPR=0.995 FPR=0.995	0.500 TPR=1 FPR=1	0.500 TPR=1 FPR=1	0.500 TPR=0 FPR=0

Table 18 Precision for each algorithm, and road segment (RVSP=Risky visual scanning patterns; EOFR=Eyes off forward roadway; MDD=multidistraction detection; TPR=True positive rate; FPR=False positive rate)

		Algorithms				
		RVSP	EOFR	AttenD	MDD-visual	MDD-cognitive
Maximum precision						
Road	Urban	0.762 TPR=0.742 FPR=0.237	0.865 TPR=0.608 FPR=0.091	0.714 TPR=0.925 FPR=0.403	0.916 TPR=0.656 FPR=0.091	0.697 TPR=0.903 FPR=0.720
	Interstate	0.632 TPR=0.540 FPR=0.274	0.788 TPR=0.605 FPR=0.202	0.706 TPR=0.919 FPR=0.444	0.888 TPR=0.726 FPR=0.137	0.570 TPR=0.855 FPR=0.694
	Rural	0.796 TPR=0.699 FPR=0.199	0.837 TPR=0.677 FPR=0.183	0.758 TPR=0.935 FPR=0.344	0.906 TPR=0.720 FPR=0.113	0.636 TPR=0.613 FPR=0.312
Minimum Precision						
Road	Urban	0 TPR=0 FPR=0.011	0.129 TPR=0.027 FPR=0.016	0.500 TPR=1 FPR=1	0.501 TPR=1 FPR=0.995	0 TPR=0 FPR=0
	Interstate	0 TPR=0 FPR=0	0 TPR=0 FPR=0	0.493 TPR=0.976 FPR=1	0.500 TPR=1 FPR=1	0 TPR=0 FPR=0
	Rural	0 TPR=0 FPR=0	0.065 TPR=0.016 FPR=0.005	0.500 TPR=1 FPR=1	0.500 TPR=1 FPR=1	0 TPR=0 FPR=0

Considering the results of the ROC curves, AUC, accuracy and precision, it is apparent that certain tradeoffs exist when deciding which algorithm is best suited for distraction detection. For example, the *AttenD* algorithm consistently yielded high true positive rates, area under the curve values, accuracy, and precision, yet the lowest false positive rate was always near 0.4. Choosing this algorithm for distraction detection would ensure detection of distraction, but at the expense of a high number of false alarms. Drivers might not accept a system with such a high rate of false alarms, but the consequence of false alarms likely depends on the particular characteristics of the mitigation, an issue we return to in the following section.

Alert Latency

To further distinguish the algorithms, the time between the secondary task onset and when each algorithm first alerted (i.e. alert latency) was also analyzed. Note that alert latency refers to the time when an alert would have been issued; no alerts were actually issued in the present experiment. An algorithm that has a high area under the curve, high accuracy, high precision, and low alert latency would be ideal.

Alert latency was calculated for each of the 16 different task periods—because only one algorithm measured cognitive distraction, only the visual tasks were analyzed. Alert latency was averaged across all task periods to form one aggregate value. If any missing values were still present, multiple imputations were used to estimate the missing values using Bayesian linear regression.

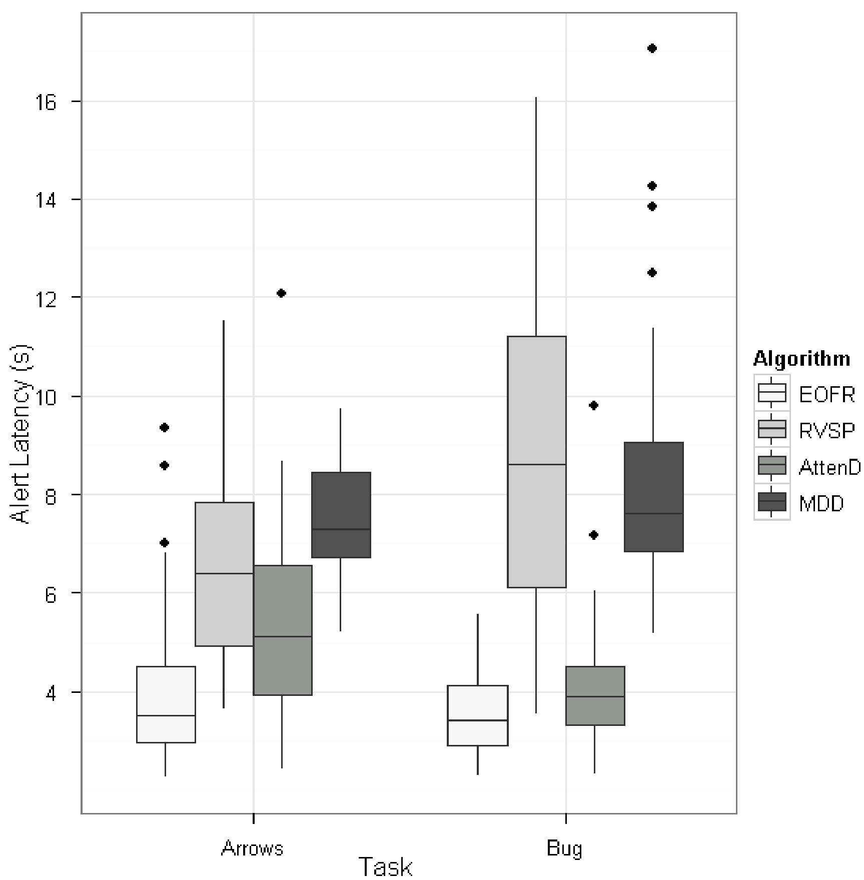


Figure 17 Box plot of each algorithm's alert latency by task

Figure 17 is a box plot of each algorithm's alert latency by task. A mixed model ANOVA shows significant effects of algorithm ($F(3,90) = 76.79$, $p < 0.0001$), and the task-algorithm interaction, $F(3,90) = 9.51$, $p < 0.001$. The *EOFR* algorithm had the lowest alert latency (M = 3.85), followed by the *AttenD* algorithm (M = 4.85), the RVSP algorithm (M = 7.78), and the MDD algorithm (M = 8.09). The significant interaction showed that alert latency was lower during the arrows task for the *RVSP* and *MDD* algorithms, and lower during the bug task for the *EOFR* and *AttenD* algorithms. However, the difference between the arrows and bug task was only significant for the RVSP algorithm ($F(1,30) = 2.89$, $p = 0.002$) and the *AttenD* algorithm ($F(1,30) = 7.85$, $p = 0.0088$.

Additional Indicators of Distraction beyond Eye Metrics

Although eye gaze metrics are very effective in detecting distraction, driving metrics might supplement them. Several powerful feature selection approaches can identify potentially useful indicators of distraction. Many feature selection approaches exist and no single approach is most suitable for all applications. However, a comparison of many feature selection techniques identified several particularly promising approaches (Hall & Holmes, 2003). One of these — wrapper-based—selects features based on algorithm performance when it is tested with different combinations of features. The feature combination that produces the best performance is selected for further algorithm development. Another approach—correlation-based feature selection—includes features that best predict distraction and removes features that are correlated with other selected features (Hall & Holmes, 2003). Such techniques often outperform more traditional approaches such as principal component analysis. These techniques are also less likely to capitalize on chance associations compared to step-wise regression.

The wrapper-based and correlation-based approaches were combined with two common data mining techniques—decision tree and naïve Bayes models to produce four feature selection methods that were applied to the data: wrapper-based using a decision tree, wrapper-based using a naïve Bayes algorithm, and correlation-based feature selection tested with both a decision tree and naïve Bayes algorithm. Each of these approaches was applied to the driving performance data in each segment of the drive (urban, interstate, and rural), identifying a set of distraction indicators for each segment. The indicators were based on data aggregated by event within each segment. The driving performance data included driver inputs, such as accelerator and steering wheel modulation, as well as driving performance indicators, such as speed and lane position. Table 10 and Table 22 summarize the variables considered. Note that these measures exclude gaze based indicators because gaze is much more sensitive to visual distraction and would dominate the other driver performance variables. A similar analysis of gaze measures produces AUC measures of 1.0, which confirms the use of gaze metrics in the four algorithms evaluated. The AUC values described in this section (unlike the feature selection results) are based on data from the entire drive and reflect the classification of the entire drive as one where the driver engaged in a distraction or not. This contrasts with the previously described algorithms that are detecting second-to-second engagement, which is a much more difficult classification problem.

Table 19 Additional variables included in feature selection.

Metrics		Description
1	acc_hold	throttle hold
2	acc_rr_avg	average accelerator pedal reversal rate in a 6 second moving window.
3	acc_rr_sd	standard deviation of accelerator pedal reversal rate in a 6 second moving window

Ten-fold cross-validation was used to assess the value of the selected variables. Cross validation splits the data into subsets or folds and the model is fit to all but one subset and then the model is tested on the withheld data. A 10-fold cross validation splits the data into 10 subsets and the overall worth value of the model, and in this case the value of a set of variables, is based on the average performance of the model on the withheld data. For the wrapper-based approaches an additional 10-fold validation process was nested within to select the set of variables that were then included and tested in outer cross-validation process.

The value of any set of distraction indicators depends, to some degree, on the algorithm in which they will ultimately be used. Some algorithms can extract predictive information from the interaction between variables (e.g., decision tree) that other algorithms cannot (e.g., naïve Bayes). The area under the ROC curve reflects how the combination of the variables and algorithm contributes to distraction detection performance. The value of the individual variables is calculated as weights that reflects their importance in the classification process. This weighting ranges from 0, where the variable is dropped because it has no systematic association with distraction, and 1, where it is strongly related.

Data for the performance of the four feature selection approaches were combined across the three segments of the drive (urban, interstate, and rural). Figure 18 shows that the driving performance measures were generally quite sensitive to distraction. An AUC value of 1.0 corresponds to perfect performance. The correlation-based feature selection and the naïve Bayes approaches performed best and had the least variation from segment to segment as shown in the thickness of the box plots.

Figure 19 summarizes the indicators of distraction selected by the four approaches within each event during the drive. The dots indicate variables selected and their weighting, and only events where there was a variable selected are included. The darker the dot the more approaches that selected the variable. Only variables with a weighting greater than .10 are plotted. The plot shows that steering-based indicators are more sensitive than accelerator or speed-based indicators, or even those based on lane position. Steering-based indicators are more often sensitive to distraction as indicated by the number of times they appear and are also more sensitive as indicated by their higher weighting. These results largely parallel the findings from the sensitivity analysis. The exception to this trend is that steering entropy emerged as highly sensitive in the sensitivity analysis, but did not emerge as one of the most indicative measures. One explanation for this result is that the sensitivity analysis is based on a single measure and its association with distraction, whereas this analysis considers combinations of measures.

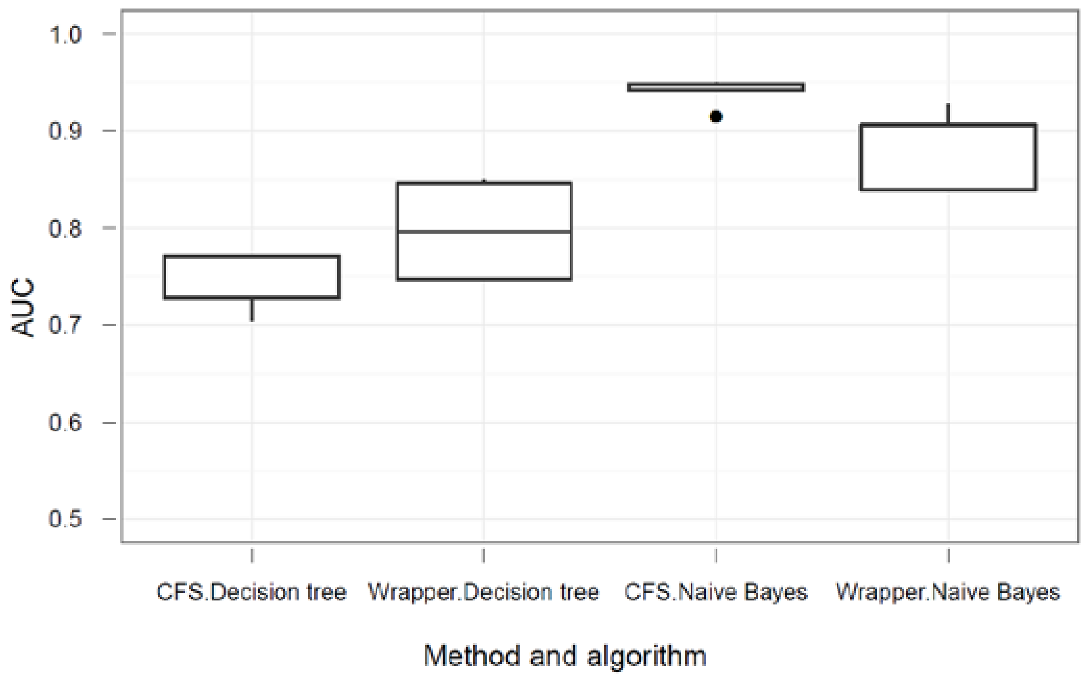


Figure 18 Area under the ROC (AUC) indicating the sensitivity of simple algorithms based on CFS (correlation-based feature selection) and wrapper-based feature selection.

Interestingly, steering measures are not best for all events: speed is more indicative for transition to dark (event 303). Of the steering measures the average and standard deviation of the steering wheel reversal rate were the most sensitive. Table 20 indicates the loadings of each variable within the correlation-based feature selection and naïve Bayes approaches. Similar to Figure 19, the standard deviation of lane position (lp_sd), steering wheel reversal rate (str_rr_avg), and the standard deviation of the steering wheel reversal rate (str_rr_sd) indicated distraction best across all three driving segments.

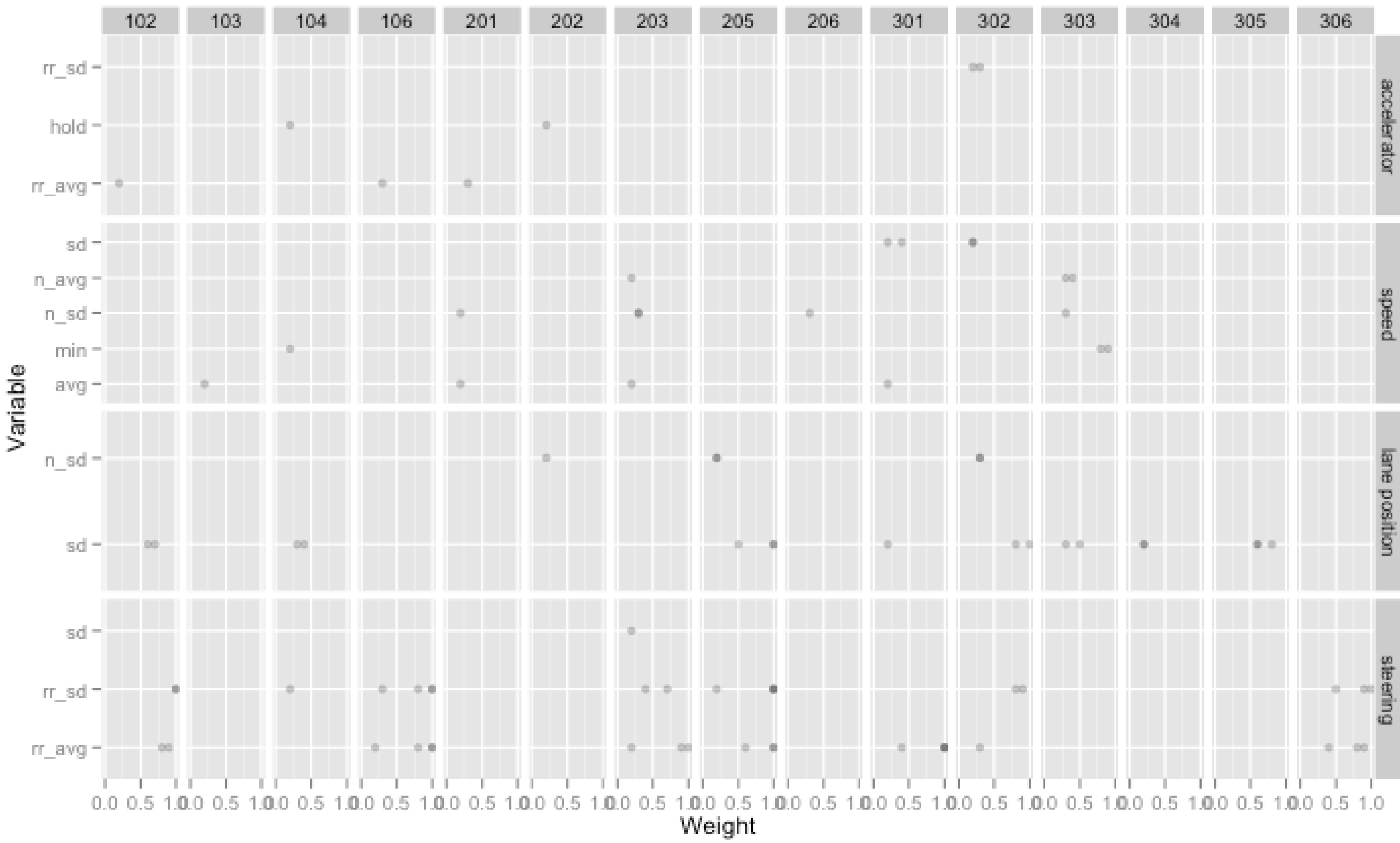


Figure 19 Importance of potential indicators of distraction. Events are labeled at the top of each column.

Table 20 Variable importance from the correlation-based feature selection approach with the naive Bayes algorithm. Dark grey indicates 0.9 or greater; medium grey indicates 0.5 à 0.8; light grey indicates 0.2 à 0.4; unshaded areas are <0.2.

		Urban (AUC: 0.915 +/- 0.081)			Interstate (AUC: 0.950 +/- 0.071)		Rural (AUC: 0.942 +/- 0.106)					
		102	104	106	203	205	301	302	303	304	305	306
Lane position	lp_avg											
	lp_sd	0.6	0.4			1.0	0.2	1.0	0.5		0.8	
	lpn_sd				0.2			0.3				
Speed	sp_avg				0.2							
	sp_max											
	sp_min								0.9			
	sp_sd						0.2	0.2		0.2		
	spn_avg								0.4			
	spn_sd				0.3							
Steering	ster_freq											
	steer_H											
	steer_sd											
	str_rr_avg	0.8		1.0	0.9	1.0	1.0	0.3				0.8
	str_rr_sd	1.0		1.0	0.7	1.0		0.8				1.0
Throttle	acc_hold*											
	acc_rr_avg*											
	acc_rr_sd*							0.3				

As steering metrics seemed to best indicate distraction, additional analyses were done which combined steering metrics with the MDD algorithm. More specifically, if the exponential weighted moving average (EWMA) of the magnitude of the steering wheel angle exceeded 2.5 degrees, the MDD visual warning was suppressed. If the EWMA exceeded nine degrees, its visual threshold was increased to 65 percent from 60 percent.

Adding the steering metrics to the MDD algorithm produced the following ROC plots for the arrows and bug task. Adding in steering metrics slightly improved the TPR on the left end of the plot. The AUC was significantly reduced for the arrows task (Figure 20) but increased for the bug task (Figure 21).

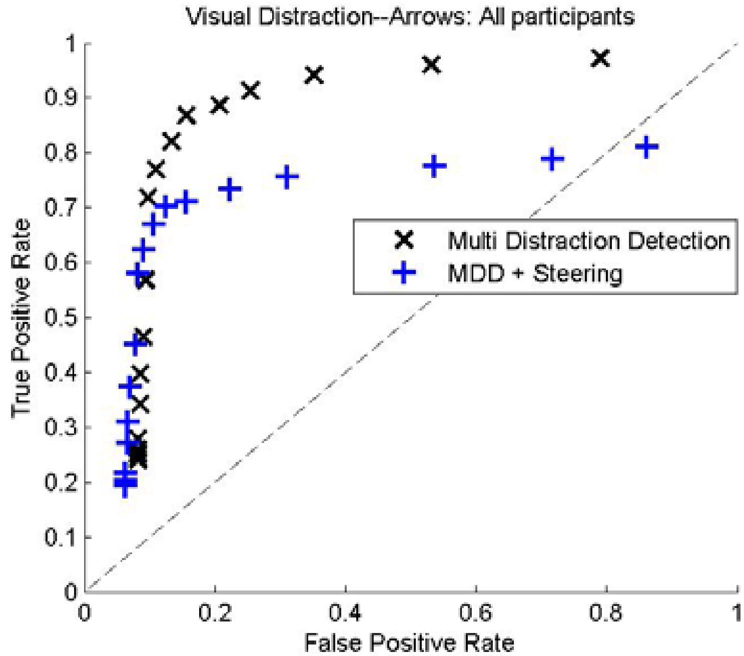


Figure 20 ROC plots showing performance of the multidistraction detection algorithm with steering wheel angle enhancement for the arrows task.

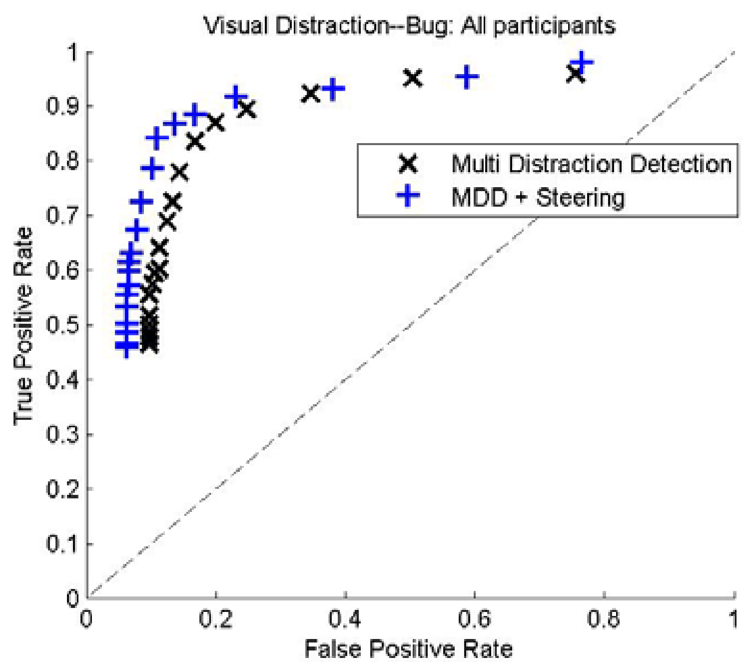


Figure 21 ROC plots showing performance of the multidistraction detection algorithm with steering wheel angle enhancement for the bug task.

CONCLUSIONS AND RECOMMENDATIONS

The results demonstrate that gaze-based algorithms can detect driver distraction in a reliable manner. They also point to the successful application of a protocol for assessing distraction detection algorithms. This section discusses the implications of these results for future algorithm and evaluation protocol development.

Implications for Algorithm Development

All four distraction detection algorithms show that the drivers' gaze can be used to indicate distraction. The protocol also demonstrated the vulnerability of these algorithms to the performance of a single sensor—an eye tracking system. This finding prompted the inclusion of a seat-based sensor of drivers' posture to compensate when the eye tracker cannot provide reliable data. Steering, lane position, and other driver performance metrics are also sensitive to distraction; these measures could be used as supplements to an algorithm that solely relies on eye movements.

Implications for Protocol Development

Perhaps the most important outcome of this analysis is an understanding that distraction and its detection cannot be considered independent of the driving environment. A dense urban environment includes situations where a driver's eyes are often not focused on the center of the road. The influence of visual demands associated with urban scenes can undermine algorithm performance, resulting in either a high rate of misses or a high rate of false alarms. This suggests that evaluation protocols require a simulator with a wide field of view and a detailed urban scene. Such a requirement might be difficult to satisfy on a test track.

A challenge with regard to use of the simulator was the attrition rate due to simulator sickness. This protocol added distraction tasks to the driving environments that were previously used for evaluating the effects of alcohol (Lee et al., 2010). The alcohol study had an attrition rate of less than 1 percent due to simulator sickness; whereas, this study had an attrition rate of 15.2 percent. The addition of these tasks clearly had a negative effect on simulator sickness rate. Due to the nearly continuous engagement in secondary tasks throughout the drive, many of which took the drivers eyes away from the roadway repeatedly, the primary concern is the rate at which the driver needed to engage in these tasks. To reduce simulator sickness, it would be recommended that the density of distraction tasks that take the drivers eyes away from the forward roadway be decreased.

A further challenge in assessing distraction detection algorithm performance concerns the cost of a false alarm and the value of correctly identifying a distracted driver. The cost of a false alarm depends on driver acceptance. False alarms might either disrupt drivers' attention to the road or undermine their acceptance of the mitigation system. To some extent these consequence depend on the mitigation system and suggest that algorithm assessment should be tied to mitigation assessment. The degree of this coupling and how it is addressed in a comprehensive evaluation protocol represents a critical issue. One approach to algorithm evaluation to partially address this issue is to include a cost-based performance metric that weights false alarms and misses differently.

Another challenge confronting distraction detection concerns the definition of distraction-related impairment. This protocol operationalizes distraction as occurring when the driver engages in a distracting task. This approach might overestimate the occurrence of distraction because it may be possible for drivers to engage in secondary tasks and not become distracted. A challenge for refining the assessment protocol is to identify an alternate way of operationalizing distraction.

Strategic Highway Research Program (SHRP2) and Algorithm Assessment

One approach to operationalizing distraction for algorithm assessment is to build on naturalistic data, such as that collected as part of the Strategic Highway Research Program (SHRP2). The SHRP2 data can help distinguish between distraction and simply secondary task engagement. Using these data instances of distraction can be identified by associating patterns of interaction and eye glances with crash and near-crash events. This association can be measured using an odds ratio to indicate how much more likely a crash or near crash is to occur when distraction is present compared to when there is no distraction. When an odds ratio is greater than one, it indicates that there is an increased crash risk relative to baseline driving.

Different levels and operational definitions of distraction could be defined by the performance of different algorithms. Because each definition of distraction might produce a different scale, the levels of distraction must be normalized for comparison purposes. A non-parametric procedure – the percentile – is an appropriate normalization approach since it could be applied for non-symmetric and skewed distributions (Liang, 2009).

The odds ratios and their 95 percent confidence intervals could be calculated for each normalized segment of the distraction distribution. The underlying assumption of this analysis is that the odds ratio reflects the accuracy with which distraction is estimated. It is assumed that a high level of distraction would correspond to a high crash risk and low levels of distraction would correspond to a low crash risk. A good definition of driver distraction as an inherently hazardous state should produce a steep, monotonic relationship between odds ratio and distraction level. Two indicative metrics for the algorithms' comparison would be considered: the slope of the odds ratio as a function of distraction level and the maximum odds ratio (MaxOR). The slope of the odds ratio indicates how quickly the odds ratio increased with distraction level; the greater the slope, the faster the odds ratio increased and the earlier the algorithm can identify increased crash risk. The MaxOR could indicate how sensitive the crash risk is to the estimate of distraction (Liang, 2009).

CHAPTER 5. DESCRIPTION AND APPLICATION OF A PROTOCOL FOR EVALUATING ALTERNATIVE FEEDBACK SYSTEMS

INTRODUCTION

Potential vehicle-based countermeasures used to mitigate distracted driving vary in the type of mitigations used, the intent of the mitigation, and the timescale of feedback. Examples of countermeasures based on real-time assessments of driver distraction include adaptations of crash warning systems, electronic communication or "infotainment" function lock-out, feedback intended to redirect driver attention, and continuous attention level feedback. Real-time feedback may be delayed in high workload situations. Retrospective, post-drive feedback can be provided when the vehicle is no longer in motion. Potential countermeasures can also have different intentions, ranging from mitigating only instances where distracted driving results in a crash-imminent situation to providing feedback that helps the driver better understand the risks involved in their distracted driving.

The effectiveness of countermeasures depends on several factors, including how well matched they are to the distraction detection algorithm, the quality of the sensor data, the appropriateness of the mitigation form, function, and timescale for the intent of the system, and its acceptability to drivers. Effectiveness also may depend on addressing drivers' willingness to engage in secondary tasks while driving through post-drive feedback that helps drivers judge their level of distraction and its impact on their driving performance.

This chapter describes a protocol for evaluating distraction mitigation systems. The protocol evaluates a mitigation approach's effect on driving performance, behavior that leads to decrements in driving performance, and driver attitudes toward distracted driving.

The protocol provides a best-case estimate of distraction detection effectiveness – the data used to estimate vehicle and driver state are more accurate than what would be collected during on-road driving. However, in some ways, the protocol may underestimate mitigation effectiveness – the greatest benefit of the mitigation might occur through changing drivers' behavior over several weeks or months, which the protocol cannot directly evaluate. Measuring drivers' perceptions of their driving performance and the performance of their peers while distracted, as well as their intention to engage in future distracting activities while driving, provides a subjective, complementary approach to assess the long-term effect of the mitigation system. To that end, this protocol assesses attitudinal changes that distraction mitigation might produce. Behavior changes that a mitigation system might encourage include: not engaging in distracting tasks while driving, engaging less frequently, shifting distracting tasks away from high-demand situations, reducing the frequency of dangerously long glances, and ensuring that drivers' attention is directed towards the road during critical events. This protocol is designed to measure the degree to which mitigations influence drivers to change their behavior as well as their influence on current driving performance. This approach allows for the benefits of each type of mitigation to be factored into the overall evaluation of the distraction detection and mitigation systems.

The protocol used to assess distraction detection was used for this protocol, with slight modifications detailed below. It includes a relatively uncommon, but very risky, secondary task (reaching to the backseat), a common but less risky task (adjusting the radio), and more artificial tasks to represent future in-vehicle entertainment and information systems.

The evaluation protocol was assessed by applying it to the real-time and post-drive feedback distraction mitigation approaches. The real-time feedback system has visual and audio components. The visual component is a head-up display (HUD) presented on the left, center, and right side of the windshield. The post-drive feedback system has two components: a report card of the participant's performance during the drive accompanied by video playback of a short drive segment depicting distracted driving.

DEVELOPMENT OF THE DISTRACTION FEEDBACK COUNTERMEASURE EVALUATION PROTOCOL

Participant Characteristics

Participants between the ages of 25 and 50 who have experience engaging in distracting activities while driving, including talking on a cell phone, sending or receiving text messages, sending or receiving emails, eating, or changing compact discs, were used to minimize within-group variability that might reduce the statistical power of the comparisons. They must also have adequate driving experience and possess a valid, unrestricted U.S. driver's license (with an exception for participants with a corrective lens restriction) for at least one year and have driven more than 3,000 miles per year. To ensure that simulator experience would not affect our results, participants could not have participated in a driving simulator study in the 12 months preceding their study visit, and they could not have participated in any studies conducted at NADS that used a similar simulation scenario. Participants were required to be in good general health, have no history of motion sickness, and not use special devices (e.g., spinner knobs, booster seats) while driving.

Because some of the real-time mitigations were auditory alerts, participants were required to have normal hearing or full correction of hearing problems with the use of hearing aids so that hearing loss would not affect their evaluation. Since this was not an evaluation of distraction detection, potential participants were excluded if they required corrected vision to drive and could not wear contact lenses to their study visit, due to limitations of the research-grade eye tracking system to accurately track drivers with eyeglasses.

Driving Scenario and Distraction Tasks

The driving scenario used in this study was the same as the one used in the evaluation of the distraction detection protocol (Chapter 4). As in that evaluation, it included typical driving events from urban, interstate, and rural roadways. In each of these driving environments, participants completed three types of prompted secondary tasks and one self-paced task: a task requiring a visually guided reach into the back seat (bug task); a visual/manual task using a touch screen display located at the top of the center console (arrows task); a complex cognitive task requiring the recall of information and navigation through a menu system (menu task); and a self-paced visual/manual task using a touch screen display in the lower center console (radio task). Each of these tasks required the participants to take their attention from the roadway and provided an opportunity for the system to attempt to mitigate that distraction. The distraction tasks indicate how the mitigation influences driver response to instructions to perform them. Appendix G includes detailed descriptions of the tasks.

The presentation of the bug, arrows, and menu tasks occurred during the 75- to 90-second intervals during each of the drive environments (urban, interstate, and rural), such that participants experienced each kind of distraction in each environment (see Table 9 in Chapter 4 for the events during which these distraction tasks were expected to occur). Tasks were designed to be completed in 3 to 6 seconds. When drivers were not engaged in the prompted distraction tasks, they were presented with the radio task, a self-paced task that requires them to adjust a setting to reduce the noise level (see Table 10 in Chapter 4 for the events during which this distraction task was expected to occur). Unlike the discrete presentations of the prompted tasks, the radio task was designed to be an ongoing task that drivers could attend to at their leisure. Timing and dynamics of this task match the data collection protocol for distraction detection.

The prompted tasks could be deferred, but no instructions about this were provided. A deferred task was defined as a delay in initiating the task after the participant was prompted to begin (by contrast, tasks that were initiated by the participant but not completed were considered incomplete, not deferred). Tasks could be deferred to any point within the drive segment; all deferred tasks not completed during a drive segment were presented to the participant to complete at the end of the segment. If a prompted task was deferred, the prompt to begin the task would continue to play every 10 seconds until the task was initiated. The tasks were made deferrable to allow participants the option to engage in tasks when they were more comfortable doing so, such as in lower demand areas of the driving environment (e.g., no curves, less traffic), or in the no demand situation at the break between driving environments.

Subjective Assessment

One of the greatest benefits from distraction mitigation systems may derive from their effect on drivers' attitudes toward engaging in distracting activities because a shift in attitude could ultimately reduce their willingness to engage in these tasks. Table 21 shows the questionnaires used in this study and summarizes the intent of each questionnaire. The questionnaires used in the distraction detection protocol (Chapter 4 and Appendices J-N) were supplemented with three additional questionnaires. Two scales were designed for this study to assess important contributors to attitudinal change: a planned behavior questionnaire (Appendix P) and a performance calibration questionnaire (Appendix

Q). Additionally, the post-drive questionnaire was designed to assess the degree to which participants found the mitigations to be acceptable (Appendices R and S).

Table 21 Questionnaire/Scale Purpose

Questionnaire/Scale	Citations	Purpose
Distraction Driving Survey		Demographic, driving history, and general health questions; Questions assessing frequency and comfort level of performing distraction tasks while driving
Performance Questionnaire	Horrey et al., 2008; Horswill et al., 2004; Agostinelli et al., 1995; Neal & Carey, 2004; Neighbors et al., 2004; Lesch & Hancock, 2004	Assess participants' awareness of driving performance decrements associated with secondary task performance; would expect to see a shift in performance ratings as participants' perceptions of decrements in driving performance while distracted become better calibrated with actual driving performance decrements
Planned Behavior Questionnaire	Ajzen, 1991; Horrey & Lesch, 2008; Horrey et al., 2009; Horrey et al., 2008	Predict participants' future actions and behaviors when engaging in distracting activities while driving
Post-Drive Questionnaire	Davis, 1989, 1993	Assess mitigation acceptance
Realism Questionnaire		Assess perceived realism of the simulator and simulated drive compared to the real world
Wellness Questionnaire	Kennedy et al., 1993	Evaluate signs of simulator sickness
Workload Scale		Participants' perceived workload and driving performance (lane departure and speed control) by distraction type and driving environment

The planned behavior questionnaire is based on the theory of planned behavior (Ajzen, 1991). As described in Lee and See (2004),

> [The theory of planned behavior] shows that behaviors result from intentions and that intentions are a function of attitudes. Attitudes in turn are based on beliefs. According to [the theory], beliefs and perceptions represent the information base that determines attitudes…An attitude is an affective evaluation of beliefs that guides people to adopt a particular intention. Intentions then translate into behavior, according to the environmental and cognitive constraints a person faces (pg. 53).

As such, the planned behavior questionnaire contained questions aimed at determining the participant's future actions and intentions to engage in distracting activities. Questions regarding frequency of talking on a cell phone, interacting with an entertainment system, eating/drinking, and looking at a map while driving replicate previous studies (Horrey & Lesch, 2008; Horrey et al., 2009). Texting while driving was added as it is also a common distraction. The questionnaire also assessed the participants' perceived need to engage in these activities while driving (Horrey & Lesch, 2008). These questions concerned specific situations in which drivers would encounter an urge to engage in distracting activities (i.e., after a phone rings, after missing a phone call, or after receiving a text). Participants responded to these questions along a continuum, similar to a previous study (Horrey & Lesch, 2008). In addition to frequency in engagement, it was also important to measure how drivers choose to defer engagement, if at all. As we are aware of no existing surveys that address this issue, these questions are original. Because the theory of planned behavior states that intention to perform behaviors can be predicted by social norms, participants were also asked to assess how often their peers will engage in certain behaviors.

Calibration of drivers' estimates of how distractions affect their driving performance with their actual driving performance is an important contributor to attitudes about distracted driving. Mitigation systems are likely to improve driving safety by helping drivers realize how much distractions put them at risk. Consequently, measures of drivers' risk calibration could be central to evaluating distraction mitigation systems. The performance questionnaire (Appendix Q) included four measures – task performance, lane position, speed, and eye glance behavior – as each of these measures can be accurately perceived and assessed by the driver. They can also be easily compared to actual driving performance without much data transformation. For each question that asks for an indicator of driving performance, participants are also asked to estimate the performance of a peer driver. The use of the word "peer" was chosen rather than "average" because some people associate "average" as a negative characteristic. The concept of asking for a peer's performance is modeled after similar driving studies (Horrey et al., 2008; Horswill et al., 2004) and studies from other fields (Klar et al., 1996; Neighbors et al., 2004; Perloff & Fetzer, 1986). By asking participants their opinions of a peer driver, their responses can be directly compared to the actual performance of drivers from the same study to see how well-calibrated drivers are to the social norm (Horrey et al., 2008), similar to studies related to binge-drinking behavior (Agostinelli et al., 1995; Neal & Carey, 2004; Neighbors et al., 2004).

Participants are asked to give the confidence levels of their performance estimates to determine if their confidence level corresponds to their driving performance (Lesch & Hancock, 2004). For example, it is expected that those with higher confidence ratings will be those who exhibit the smallest performance decrement in driving behavior while engaging in distracting tasks. The effect of overconfidence is also measured by asking whether the participant thinks a peer driver would behave in the same manner as the participant (Horswill, et al., 2004). In this case, overconfidence is exhibited when participants think they always perform better than their peers.

One approach to assessing risk calibration is a variance decomposition analysis of the "skill score," which quantifies the relationship between subjective ratings of driving performance and actual driving performance. Murphy's skill score is one measure of risk calibration. Previous research using Murphy's skill score uses a relatively high number of samples. The original analysis was applied to meteorological data with an *n* of nearly 800 forecasts —the exact values ranged from 775 to 870 (Murphy, 1988). In a more recent study that used a within-subject design to determine the effect of receiving automated aid on estimates of enemy threat level, 15 participants rated their judgments of 51 different scenarios to generate 51 separate skill scores for an *n* of 51 (Horrey et al., 2006). In the second half of the study, 12 participants each completed 43 trials yielding an *n* of 43. Participants generated estimates of their performances every 25 seconds yielding a large volume of data.

Because of participants' limited exposure to the distraction tasks and mitigation systems, the performance questionnaire from which skill scores are to be obtained was administered after each driving environment to increase *n* in order to provide enough data to calculate a reliable skill score. This was also done to help drivers avoid accurately remembering only their performance during the last portion of the drive (a recency effect). The decision to have repeated exposures to the performance questionnaire was weighed against two concerns: repeated measures could increase self-awareness, change willingness to engage, and affect driving performance, and these changes could diminish the observed benefit of the mitigation relative to the baseline.

The post-drive questionnaire was based on a user acceptance framework developed by Davis (1989, 1993) that evaluates perceived ease of use, usefulness and value of the mitigations under study.

Test Procedure

After providing informed consent and completing a video release form and payment voucher, participants completed the distraction driving survey and the planned behavior questionnaire. They then watched a PowerPoint presentation that described the simulator cab and the distraction tasks they would complete during their drives. In the informed consent document and the training presentation (Appendix T), the purpose of the study was described as evaluating feedback concepts related to distraction and driving. Participants in pilot testing and in the distraction detection data collection (see Chapter 4) rarely deferred tasks, and it was thought that the task and drive instructions may have influenced participants to prioritize task completion over safe driving. The briefing presentation and drive instructions were modified to deemphasize the urgency to complete tasks, and to balance safe driving with task completion. Participants were instructed to complete as many tasks as quickly as possible after the prompt to begin, while driving safely.

Next, participants entered the simulator, received additional training, and completed the study procedures. Figure 22 displays the timeline for study procedures in the simulator: blue boxes are study materials, including scales; green boxes are simulator drives; red boxes are mitigation demonstrations, if applicable; and the purple boxes are the post-drive feedback for participants in that condition. Real-time feedback is not shown in Figure 22, but it was presented during the training and mitigation drives. After a brief overview of the simulator cab, participants rehearsed the distraction tasks (in the cab). Participants then completed an 8-minute practice drive that familiarized them with the cab's controls, steering, accelerating, and braking, and provided practice on the distraction tasks while driving. Because of a high rate of simulator sickness in the distraction detection data collection discussed in Chapter 4, the practice drive was changed to eliminate turns at the beginning and end of the drive. Participants then completed a second task rehearsal without driving.

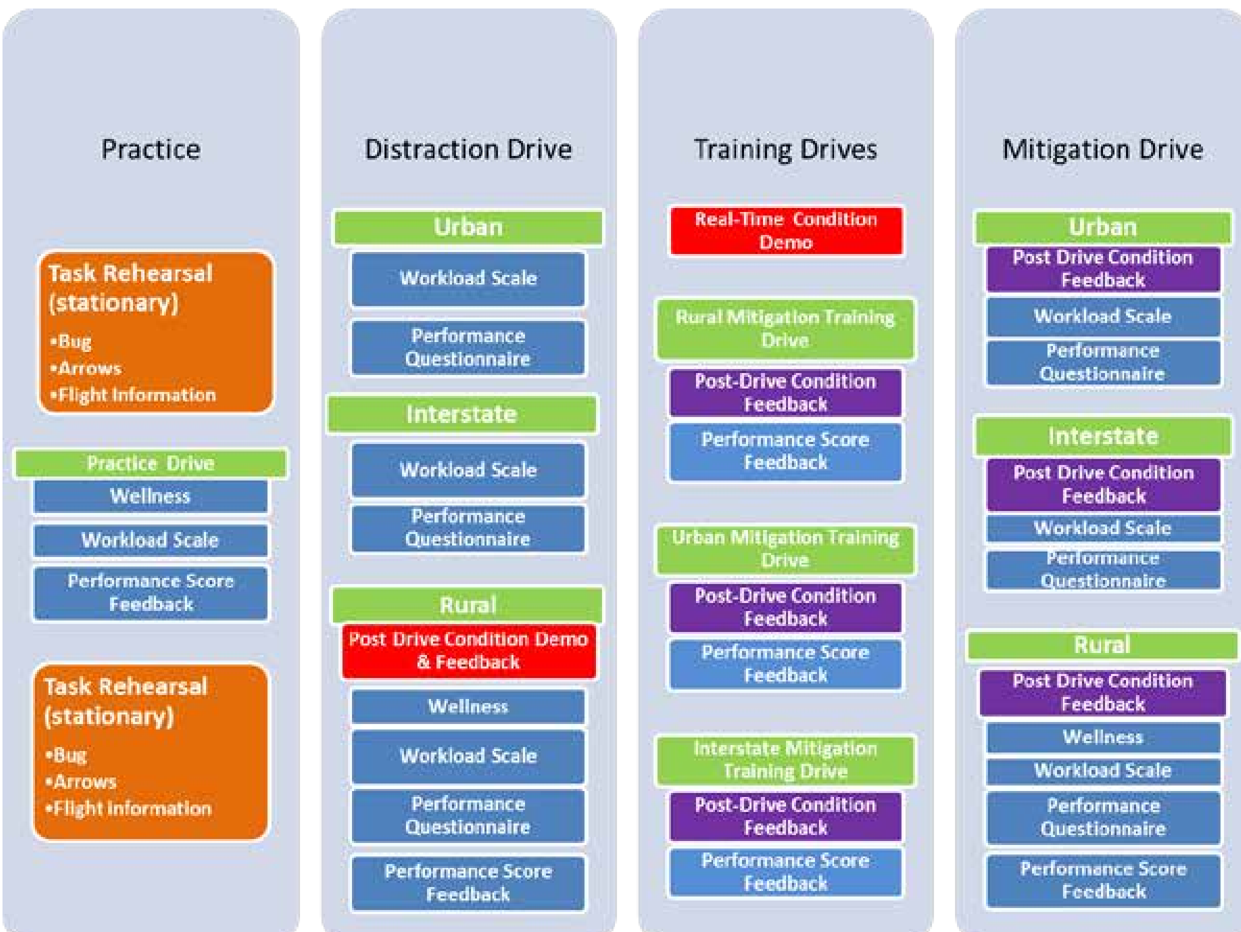


Figure 22 Study procedure timeline.

Then, following eye-tracking calibration, participants completed a series of drives: the distraction drive, an initial drive performing distracting tasks without mitigation; three short mitigation training drives; and a drive performing distracting tasks with mitigation. The three training drives provided distraction tasks and repeated exposure to the applicable mitigation. One third of the participants experienced post-drive feedback at the end of their distraction drive and following each segment of their mitigation drive, one third experienced real-time feedback during their mitigation drive, and the remaining third served as a control condition and received no feedback. The distraction and mitigation drives each lasted approximately 25 minutes, and the three training drives were approximately 5 minutes each. During the drives, a prompt (i.e., audio cue, bee buzzing) indicated when the driver should begin each of the three distracting tasks.

After each environment (urban, interstate, and rural), participants stopped driving. If tasks had been deferred to the end of the segment, the experimenter in the simulator cab would prompt the participant to complete tasks if they did not initiate the tasks in response to repeated prompts. The radio task could not be deferred. Drivers then used a visual, analog scale to assess their subjective workload and performance (lateral and longitudinal control) by distraction task during that environment. These stops coincided with natural stopping points (e.g., a stop sign) at the end of each driving environment. Wellness questionnaires were administered to participants after the practice drive, the distraction drive, and the mitigation drive. After the final drive the planned behavior questionnaire was administered again, and drivers completed the realism and post-drive questionnaires (see Table 21). They then received a debriefing statement (Appendix U) requesting that participants not discuss their participation with others until the end date for the data collection was provided so that they would not share any task performance strategies that they may have developed with other potential participants. The statement also disclosed the source of the peer data used in the post-drive feedback.

Incentive System

Drivers received task performance feedback at the end of their practice drives, distraction drives, training drives, and mitigation drives. The experimenter in the cab communicated scores verbally after the presentation of the post-drive feedback (if applicable) and completion of the workload and calibration surveys. As in the distraction detection protocol, participants were told prior to driving that their performance on the distraction tasks (except the radio task) would be evaluated based on:

- How long it took to initiate each distraction task;
- Task completion (slow or absent input once task begins); and
- Response accuracy

Tasks deferred to the end of each driving environment were counted as incomplete/never initiated. Task performance during the previous drive segment was evaluated on a 100-point scale.

Analysis

The data from this experiment were analyzed using a two-phase approach. First, a thorough review of the data was performed to remove any spurious data points. Any inconsistencies were resolved by analytic data visualization. Where distributions appeared to diverge substantially from normal, the data was transformed or statistical techniques appropriate to the underlying variable were used.

In the second phase, the effectiveness of each of the mitigation approaches was assessed relative to the baseline (no mitigation) drive. This phase used an analysis of variance within the SAS General Linear Models (GLM)[8] procedure to test the primary hypothesis that distraction feedback can improve driver performance. The statistical model included drive type (distraction or mitigation drive) and driving environment as within-subject variables and mitigation type as a between-subjects variable. The primary dependent measures consisted of lane departure, speed, percent road center, deferral duration, task score, and calibration. Where appropriate, post-hoc t-tests were used.

To assess calibration, the participants' subjective measures of performance were subtracted from their objective performance. A positive value would indicate that participants had a higher score on a specific metric than what they estimated. For example, with lane departures, a positive difference would indicate that the participant departed from their lane more than they estimated. With task score, a positive difference would mean that the participant scored higher than they estimated. Once the difference score was calculated, a mixed model analysis of variance was performed on each metric (lane departure, speed, percent road center, and task score) with mitigation condition as a between-subjects factor while drive and road were within-subjects factors.

In addition, following the work of Murphy (1988) and Stewart (1990), the skill score (SS) based on the mean square error (MSE) of the performance judgments was decomposed to evaluate the shape, magnitude, and scale components of the drivers' judgment errors. SS is calculated using the following Equation 4:

Equation 4

$$SS = 1 - \frac{MSE_y}{MSE_R}$$

where

$$MSE_y = \frac{1}{n}\sum_{i=1}^{n}(Y_i - O_i)^2$$

and

[8] This procedure was chosen because unequal cell sizes were expected.

$$MSE_R = \frac{1}{n}\sum_{i=1}^{n}(O_i - \bar{O})^2$$

In these equations, *n* is the number of measurements, Y_i is the human's judgment score, O_i is the true score, and $\bar{O}$ is the reference judgment, the average of the true scores.

The skill score evaluates the accuracy of the participants' subjective ratings relative to the average of the actual performance values. This skill score also permits the measurement of under- or overconfidence by comparing participants' estimates of their own performance to that of their peers. If participants consistently rated themselves as better than their peers, they would be considered overconfident. On the other hand, if participants consistently rated themselves as worse than their peers, they would be considered underconfident.

This analysis is sensitive to judgment correspondence (Horrey et al., 2006). The mean square error is a measure of the squared Euclidean distance between datasets representing the human's judgment (i.e., subjective performance) and the true environment state (i.e., objective performance) (Horrey et al., 2006).

After scaling judgment error to the reference value to obtain the skill score, Murphy's decomposition separates the correlation of the estimated and actual performance from errors due to magnitude and scale (Murphy, 1988) as shown in Equation 5.

Equation 5

$$SS = \left(r_{yo}\right)^2 - \left[r_{yo} - \left(\frac{s_y}{s_o}\right)\right]^2 - \left[\frac{\bar{Y} - \bar{O}}{s_o}\right]^2$$

where r_{yo} is the correlation coefficient, s_y is the standard deviation of the true judgments, s_o is the standard deviation of the human judgments, and $\bar{Y}$ is the average human judgment.

The first component on the right side of the equation represents the correlation measure, the second component is the scale error, and the third component is the magnitude error. In this way, the participants' overall judgment quality is decomposed to determine the individual contributions of shape or potential quality apart from any biases due to errors in scale and magnitude (Horrey et al., 2006).

APPLICATION OF THE DISTRACTION FEEDBACK EVALUATION PROTOCOL

Exemplar Mitigation Systems

Real-time and post-drive mitigations represent two general types of feedback that differ in both timescale and objective. Because many distraction mitigation systems use real-time, or concurrent, feedback to redirect drivers' attention to the roadway when distraction thresholds are exceeded, a real-time feedback system will be evaluated. The second, less common approach is post-drive, or retrospective, feedback, which aims to change driver behavior based on prior driving performance. Retrospective feedback therefore operates over a longer timescale than concurrent feedback (Donmez et al., 2009). A comparison of the two types of feedback may provide insight into the effects of the timescale and form of feedback on current and long-term driving performance and behavior. Real-time feedback may have an immediate impact on driving performance but may not affect drivers' willingness to engage in distracting tasks. Post-drive feedback displays patterns of behavior and performance that may better target attitudinal and "cultural" change regarding distracted driving.

Real-Time Feedback

The real-time feedback, a head-up display (HUD), was designed to recapture the driver's attention with visual and auditory alerts when visual or cognitive distraction was detected. Visual alerts were displayed on the windshield to the left, right, and in the center of the driver's field of view using 3 white LED lights. The three LEDs were installed on the dashboard, reflecting light off the windshield. The left light was located at 5 inches, the center at 27 inches, and the right at 44 inches from the left edge of the dash. The left and right LEDs were placed in these locations so that the lights would be in the driver's peripheral view (>20 degrees from the forward view) when the driver looked straight ahead.

The three positions of the visual display were intended to redirect attention during different types of distraction. The LEDs in the left and right positions were intended to mitigate distraction when the driver's gaze was concentrated on the center of the road during periods of cognitive distraction. There was no auditory component of this feedback. These LEDs flash in the same blink pattern but with an offset, creating a left-right alternating pattern (Blink pattern: 50 percent duty cycle, 1,000 ms period, 100 ms on/off).

The display in the center was used to redirect attention to the road during periods of visual distraction. There were two visual distraction alerts: (1) a long glance alert was initiated when drivers engaged in a single three-second glance away from the road center; and (2) a visual glance history alert was initiated when drivers were currently engaged in a glance outside of the road center and had not spent enough time looking at the road center within a certain window. The flash rate was the same for both visual alerts (1,000 ms warning period containing three sub-pulses of 100 ms), and was synchronized with the auditory tones. To differentiate the two visual distraction alerts, the long glance alert was issued in a lower frequency tone than the glance history alert.

Post-Drive Feedback

Post-drive feedback was designed to provide coaching that informed and motivated drivers to stay focused on the primary driving task. It is comprised of a report card with four screens that convey multilevel feedback about the driver's distraction level and related performance and behavior measures, as well as video of distracted driving from the completed trip. Screens were designed to be reviewed in 8 to 15 seconds.

The first screen provided feedback about the driver's distraction level over the course of the drive (Figure 23). It displayed a line graph of the maximum, normalized value of the three percent road center (PRC) outputs of the algorithm graphed across each driving environment (referred to as "Distraction Level") across three levels of distraction (low, medium, and high). The feedback also compared a participant's distraction level to his or her peer driver group. Peer data was generated from the distraction detection data collection (Chapter 4), weighting non-distracted drive data, and distracted drive data 10 to 1 in order to present credible peer distraction-level data that will be comparable to or better than the participant's data.

Participants also received a distracted driving letter score based on the comparison of the participant's distraction level to the peer distraction level, where:

A = Participant's average score is equal to or less than the peer score

B = Participant's average score is between 0 percent and 25 percent greater than peer score

C = Participant's average score is between 25 percent and 50 percent greater than peer score

F = Participant's average score is more than 50 percent greater than peer score

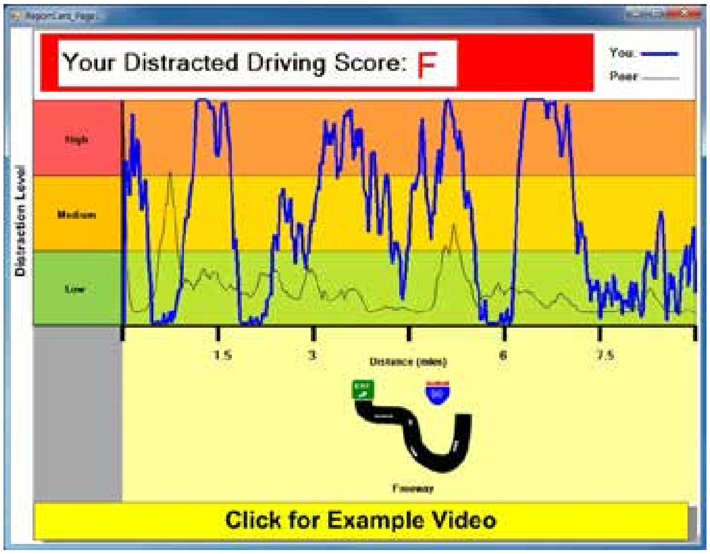


Figure 23 Post-drive report card distraction level screen.

From this screen, participants could click a button to view a video of their distracted driving during the previous trip. A second screen provided an explanation of the video, including the type of distraction ("too frequent glances off road," "tunnel vision," or "eyes off road") and the safety critical performance measure ("weaving," "lane departure," and "collision") (Figure 24).

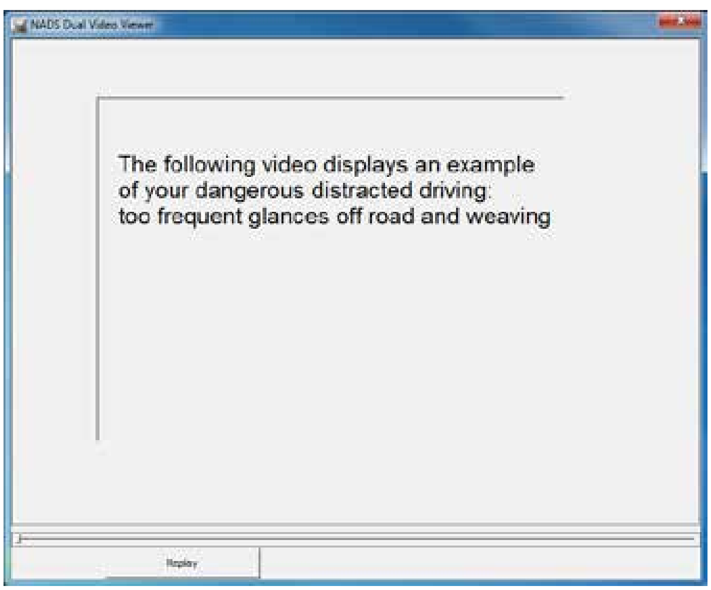


Figure 24 Post-drive report card video introduction screen.

The report card automatically forwarded to the video display (Figure 25). The 15-second video showed the participant's face and forward view for 5 seconds preceding a distraction output by the distraction detection algorithm and spanning the 10-second event window. The instance of distracted driving displayed for each driving environment was determined by a distraction event scoring algorithm. This algorithm rated the following measures over a 10-second window starting from the point at which one of the distraction algorithm outputs indicated distraction. In order of severity, the algorithm scoring components were: collision, lane departure to the left, lane departure to the right, maximum lateral acceleration. When possible, the highest scored measure was chosen to display. If no criteria were met for the video selection, the participant received a screen with the statement "No video clip available" (Figure 26). The report card moved forward automatically after 8 seconds from the completion of the video.

Figure 25 Post-drive report card video display.

Figure 26 Post-drive report card video clip not available screen.

The last screen presented detailed distraction and driving performance data (Figure 27). Bar graphs comparing the driver's performance and behavior to their peers were presented in two categories: driving errors and attention to driving. Graphs under "driving errors" represented distraction-related critical incidents (an algorithm was developed to rank and tally the number of distracted driving-related incidents) and number of drifts out of lane. Graphs under "attention to driving" represented number of unsafe glances (> 3 s) and percentage of time not looking at the roadway.

Figure 27 Post-drive report card detailed distracted driving data.

Method

Participants

Fifty-nine participants were enrolled in the study. Twenty-three were dropped from the study due to technical issues related to the post-drive video replay, incomplete data[9], noncompliance, or simulator sickness. Thirty-six participants completed the study protocol. The mean age was 34 (standard deviation (SD) 7.9 years), with a minimum age of 25 and a maximum age of 49. Participants were balanced for gender. Eighteen males had a mean age of 35 (SD 7.9 years), and 18 females had a mean age of 33 (SD 8.0 years). Three participants were Asian, 1 was black/African-American, 3 were Hispanic/Latino, and 29 were white/Caucasian.

Experimental Design and Independent Variables

A (3X4) X3 within-between subjects experimental design compared driving and task performance with four types of distraction task (bug task, arrows task, menu task, radio task) in each of three driving environments (urban, interstate, and rural). Three mitigation conditions (real-time versus post-drive versus no mitigation) were compared between subjects. Drivers performed the distraction tasks in their distraction and mitigation condition drives; the tasks were also presented in the three mitigation training drives. Because drivers might defer distraction tasks, the interactions with each task were treated as repeated measures across each of the driving environments (urban, interstate, rural) rather than considering them within each driving event. Performance data was aggregated to a single data point by taking the mean of the occurrences for each environment and task.

The order of events and distraction tasks was the same for all participants in the distraction and mitigation condition drives. The presentation of tasks to the drivers in the same context (e.g., road geometry) was intended to provide a clear, unconfounded indicator of task deferment. To minimize learning effects, the experiment used two random orders of distraction task trials that varied their content. The order of distraction task presentation was counterbalanced between road segments. However, the order of the initial distraction drive and mitigation drive, the driving environments, and the distraction tasks were not counterbalanced.

There were three independent variables in this set of analyses: driving environment, distraction type (task), and mitigation approach. A fourth independent variable, drive (distraction or mitigation) was included in the analysis of driver calibration. The preceding analyses instead used scores that indicate the difference in performance between the two drives. There were three driving environments in the baseline and distraction drives. The first was an urban driving environment that included driving events such as driving through a relatively dense urban area with numerous pedestrians, a yellow-light

---

[9] Included in this group were drivers who failed to reach the posted speed limit in the baseline drive by the second event and therefore did not engage the algorithm, which had an initiation speed threshold of 25 mph but remained active at 23 mph thereafter.

dilemma where the driver had to decide whether to stop or not, and an urban arterial event with gentle curves. The second was an interstate driving environment including straight roadway and curves, and that included interactions with heavy trucks travelling slower than the driver's vehicle. The third was a rural two-lane highway that included lighted and unlighted roadways and travel on a gravel road. Their order was the same for all drivers and for both drives.

Distraction type was represented by the task variable, which had four levels: a reaching task (bug), a visual/manual task (arrows), a cognitive task (menu), and an ongoing visual/manual (self-paced radio task).

The mitigation approach had three levels: two different mitigation approaches and a no-mitigation control condition. The two approaches are real-time feedback that provides alerts when driver distraction is detected, and post-drive feedback that provides information after the drive is complete.

The primary comparisons of interest were between metrics derived from each driving environment within the distraction and mitigation condition drives.

Dependent Variables

Table 22 lists dependent measures. Primary dependent measures assessing driver behavior include task engagement, lateral and longitudinal control, and eye movements.

Table 22 Dependent variables describing driver behavior

Dependent Measure	Description	Units
Deferral Time	The interval between task prompt and driver engaged in the task	Seconds
Task Score	The score earned by the driver on the distraction task	0-100
Initial Roadway Demand	Initial value of roadway demand (Hulse, Dingus, Fischer, and Wierwille, 1989)	0-400
Average Roadway Demand	Average value of roadway demand (Hulse, Dingus, Fischer, and Wierwille, 1989)	0-400
SDLP	Standard deviation of lane position	Feet
PRC	Percent of time that the forward gaze is centered on road	Percent
Average Glance Duration	Average length of time gaze is outside of road center	Seconds
95% Glance Time	The 95th percentile glance duration (gaze outside road center)	Seconds
Number Glances Over 2 Seconds	Number of glances longer than 2 seconds in duration	Count
Number of Large Left Lane Departures	Number of times the corner of the car crossed the left lane edge by more than 3.5 feet	Count
Number of Large Right Lane Departures	Number of times the corner of the car crossed the right lane edge by more than 3.5 feet	Count
Average Speed	Average speed	Miles per hour
Percent Speed Low	The percent of time that the driver was at least 10% below the speed limit	Percent
Percent Speed High	The percent of time that the driver was at least 10% above the speed limit	Percent

Apparatus

The experimental drives were conducted using the NADS-1, a high-fidelity, motion-based driving simulator. The simulator had a Chevrolet Malibu cab that is equipped with eye- and head-tracking hardware, active feel on steering, brake, and accelerator pedal, and a fully operational dashboard. The cab is mounted in a 24-foot dome. The motion system on which the dome is mounted provides 400 square meters of horizontal and longitudinal travel and ±330 degrees of rotation. Each of the three front projectors has a resolution of 1600 x 1200, the right and left projectors have a resolution of 1280 x 1024 pixels, and the three projectors in the back have a resolution of 1024 x 768 pixels. The edge blending between projectors is five degrees horizontal. Simulation graphics within the NADS-1 generally provides a 60 Hz frame rate. Driving data are collected at up to 240 Hz.

A research grade Seeing Machines faceLAB version 5.0.2 system with dash-mounted dual stereo head channels was used for eye tracking. A Seeing Machines Driver State Sensor (DSS) V3.4.260101, a single-camera system, was used for head tracking. Figure 28 shows the eye- and head-tracking cameras and the infrared pods for the eye tracker and DSS.

Figure 28 Eye and head tracking camera and infrared pod placement.

The driver's seat was equipped with 14 force-sensing registors (FSR) manufactured by Interlink Electronics (Camarillo, CA). Force-sensing registors provide gross dynamic measurements of force. Figure 29 shows the sensor placement: 8 sensors were located on the seat back, and 6 on the seat bottom.

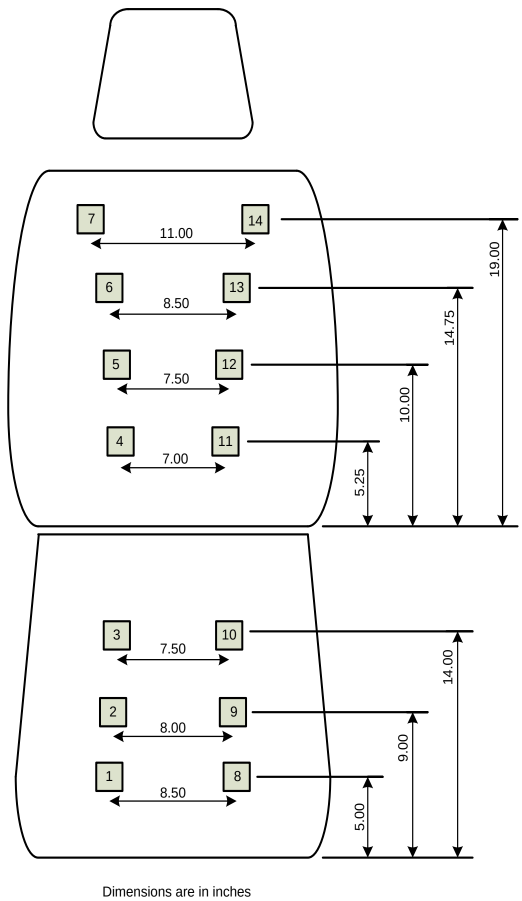


Figure 29 Seat sensor locations.

Results: Effects of Mitigations on Driving Performance and Behavior

Task Deferral

A distraction mitigation system might encourage safer behavior by encouraging drivers to delay engaging in a task, so that the demands of the task are less likely to compete with those of the roadway. To examine whether mitigations delay drivers' task engagement, the difference in the amount of time that drivers delayed each task in each environment was analyzed by subtracting the deferral time in the distraction drive from the deferral time in the mitigation drive. Mitigation ($F(2, 33) = 4.00$, $p = .028$), task ($F(3, 98) = 17.70$, $p < .001$), the interaction between mitigation and task ($F(6, 98) = 2.51$, $p = .027$), and the interaction between driving environment and task ($F(6, 180) = 2.83$, $p = .012$) influenced the change in task deferral time. The interaction between environment and task will not be discussed further as it does not relate to the effect of the mitigation.

Because the effect of mitigation depends on the type of distraction task, their interaction will first be considered. Means are provided in Table 23, with positive values indicating increased deferral. Considering differences between mitigation approaches by task, a simple effects test reveals that post-drive feedback increased bug task deferral time by 704 ms, which is significantly more than the 235 and 66 ms decreases in deferral with no mitigation and real-time mitigation, respectively. Additionally, the post-drive feedback provided a statistically smaller decrease in deferral time for the radio task compared to the no mitigation condition. It should also be noted that the mitigation systems did not significantly affect the deferral of the menu or arrows tasks.

Table 23 Interaction between mitigation and task for change in deferral time measured in milliseconds. For each task, changes in deferral times highlighted in different shades of gray are statistically different from each other.

	No Mitigation	Real-time	Post-drive
Arrows	-86	98	88
Bug	-235	-66	704
Menu	-322	-108	-287
Radio	-3010	-1930	-780

As difference data can sometimes mask underlying effects, the data was plotted to illustrate the changes in deferral time from the distraction to the mitigation drive (Figure 30). The effectiveness of the post-drive mitigation for the bug task is evident in the difference in slope relative to the no mitigation and real-time feedback conditions. The effect of the mitigations on the radio task is more complex. The participants in the post-drive mitigation condition deferred this task on average 5 seconds longer than the other two groups during the distraction drive when no mitigation was presented, indicating a pre-existing difference in how this group performed the radio task. There is no way with current data to precisely define how the post-drive feedback would have performed on the radio task had the group been more similar to the other two groups. This confound complicates the interpretation of the statistically smaller decrease in radio task deferral time that was found for post-drive feedback compared to the no mitigation condition.

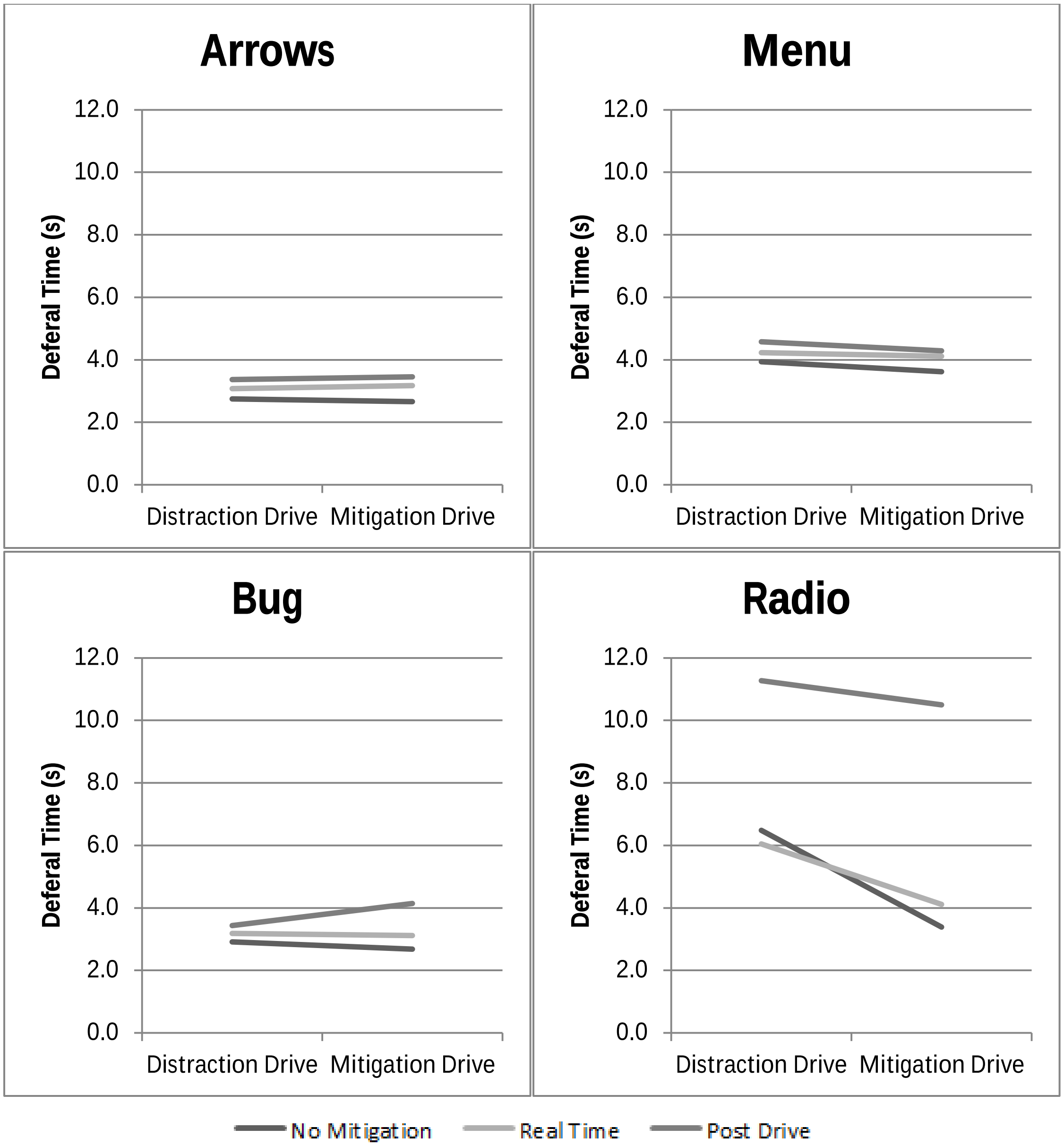


Figure 30 Changes in deferral times between distraction and mitigation drives.

Overall, there are several important points to note concerning the results:

- Post-drive feedback had a small positive effect on driver deferral of the most demanding task, the bug task. An improvement of 892 milliseconds in average deferral time was found for drivers with post-drive feedback compared to the no mitigation group.
- The mitigation effects we are seeing are small, particularly in comparison to some of the differences among the groups during their initial distraction drive.

- The largest effect was associated with the post-drive feedback condition, which had a substantially different response to deferral of the radio task. The differences associated with those drivers had a substantial effect on deferral and merit investigation because the effect was much larger than the experimental manipulations.

Task Engagement during High Demand Situations

The small increase in bug task deferral time might have resulted in a large effect on drivers' abilities to attend to the roadway demand if the deferments were associated with periods of high roadway demand. Two measures were examined: change in roadway demand at task initiation and change in average demand during the task from the distraction to the mitigation drive.

Neither measure showed a statistically significant effect of any of the independent variables: mitigation did not affect the change in level of demand at which the driver chose to engage in the tasks between the distraction and mitigation drives ($F(2, 33) = 0.19$, $p = .83$), nor did it affect the overall average level of demand during the tasks ($F(2, 33) = 1.81$, $p = .18$). No other statistically significant differences related to driving environment or tasks were found for either measure. The study provides no evidence that mitigations cause drivers to defer tasks strategically by delaying them during periods of particularly high roadway demand.

Lane Keeping

Changes in the standard deviation of lane position (SDLP) from the distraction drive to the mitigation drive were analyzed for each task and environment to examine the effect of mitigation on drivers' lane keeping. For each task, SDLP in the distraction drive was subtracted from SDLP in the mitigation drive so that smaller difference scores indicate that the mitigation was more effective. There was no main effect of mitigation on the difference scores ($F(2, 33) = 0.21$, $p =.812$); however, mitigation did significantly interact with driving environment ($F(4, 66) = 5.36$, $p < .001$) and task ($F(6, 99) = 2.27$, $p =.04$). There was also a main effect of driving environment ($F(2, 66) = 3.77$, $p = .028$) that will be discussed in light of its interaction with mitigation.

The performance of the feedback systems was assessed to determine their effect on SDLP during the performance of each task (see Figure 31). For the arrows task, post-drive feedback produced an improvement in lane keeping, whereas real-time feedback degraded lane keeping. Overall, there was no difference in lane keeping between the distraction and mitigation drives for participants in the no mitigation condition. During bug task performance, both post-drive and real-time feedback improved lane keeping relative to the no mitigation condition although post-drive feedback improved it more. For the menu task, participants without mitigation feedback showed an improvement in lane keeping between drives; however, both the real-time and post-drive feedback degraded lane keeping with significantly greater degradation from post-drive feedback. Post-drive feedback also produced a small degradation in lane keeping performance during radio task performance.

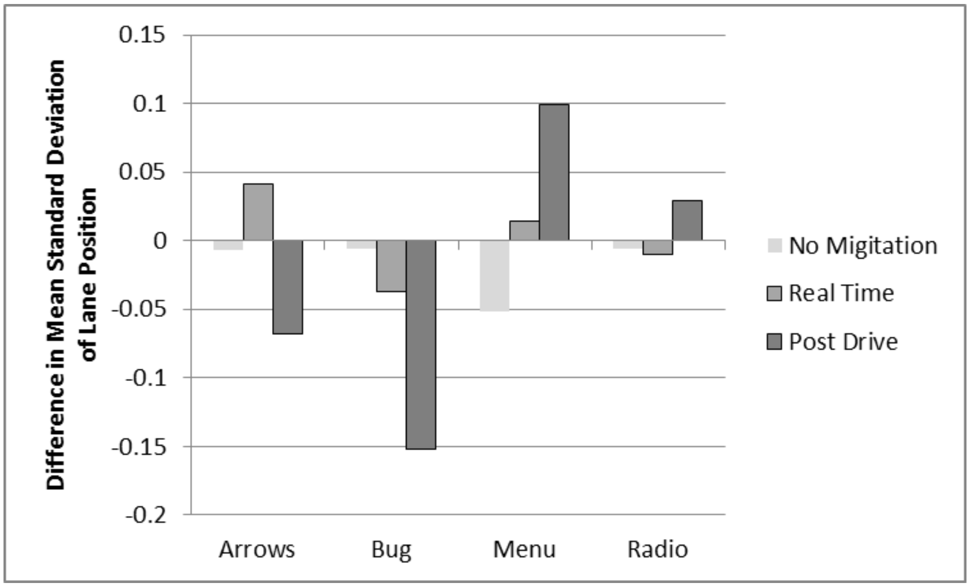


Figure 31 Difference between the distraction and mitigation drives in mean standard deviation of lane position by mitigation and task. Note: negative values indicate reduced SDLP in the mitigation drive.

We will next consider how performance of the mitigation systems varies across driving environments (see Figure 32). The effect of the system is focused in the interstate and rural driving environments. On the interstate, participants without mitigation showed improved lane keeping, whereas participants with post-drive feedback exhibited degradation. However, in the rural environment, participants with post-drive feedback showed improved lane keeping, but both the no mitigation and real-time feedback conditions showed decreased performance. All three mitigation conditions showed improvement in the urban environment although the improvement was slightly greater for participants in the post-drive condition.

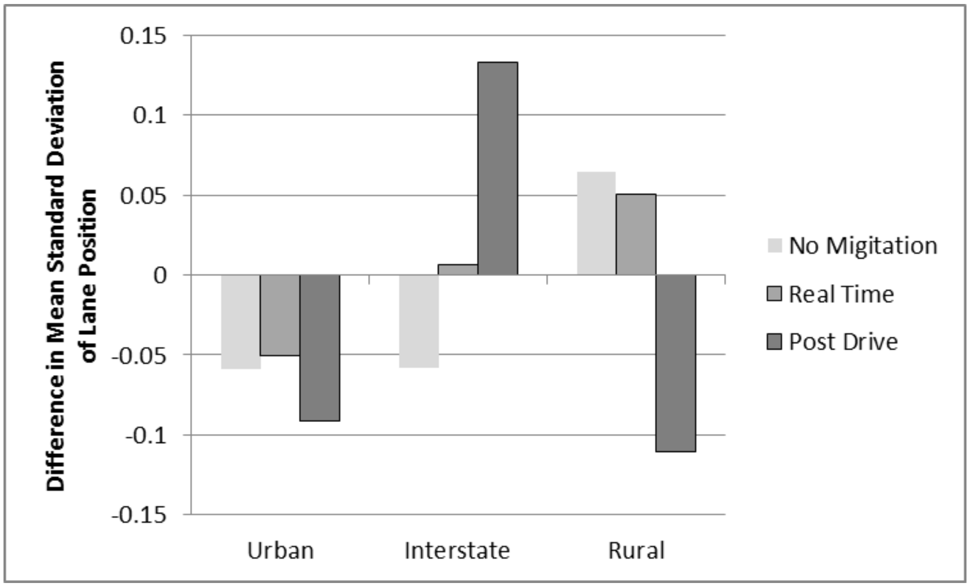


Figure 32 Difference between the distraction and mitigation drives in mean standard deviation of lane position by mitigation and driving environment.

Gaze Concentration

Changes from the distraction drive to the mitigation drive in the percent of time drivers' gaze focused forward were analyzed for each task and environment to examine the effect of mitigation on drivers' gaze concentration. For each task, the gaze concentration in the distraction drive was subtracted from the gaze concentration in the mitigation drive to provide a measure of effectiveness. Positive values indicate an improvement for visual distractions whereas negative values are an improvement for the cognitive menu task. Data from the tasks with a visual component (arrows and bug) were considered separate from the menu task.

The visual tasks showed a significant effect of mitigation ($F(2, 33) = 4.61$, $p = .017$). No other factors were significant. The post-drive feedback increased focus on the forward roadway for these tasks whereas the real-time feedback decreased it (see Figure 33). Participants with no mitigation showed a small increase in percent road center (PRC) that did not differ statistically from either mitigation condition.

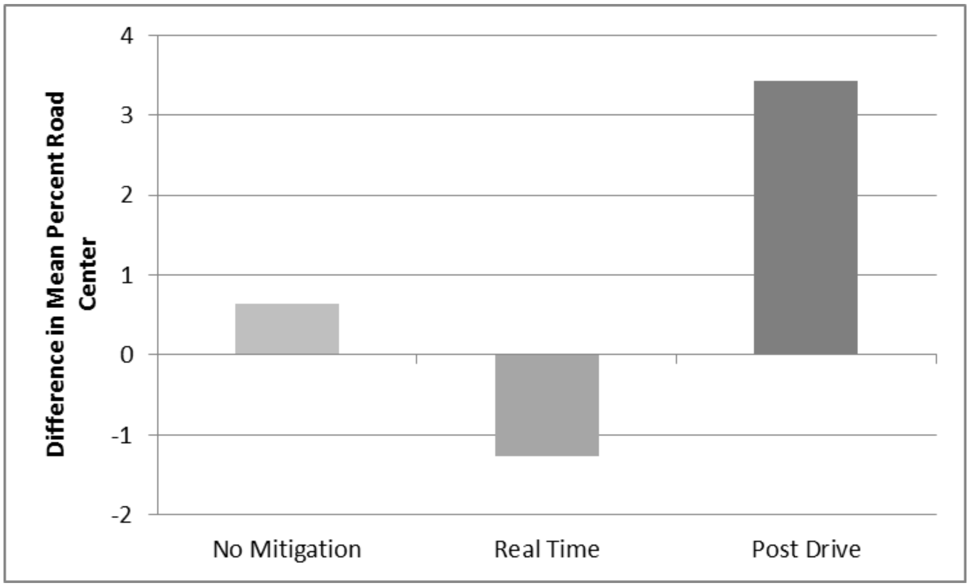


Figure 33 Difference in mean percent road center by mitigation for visual tasks. Note: positive values indicate higher PRC in the mitigation drive.

The analysis of the cognitive (menu) task found no effect of mitigation ($F(2, 33) = 0.21$, p=.81) on PRC. There was a statistically significant difference among driving environments ($F(2, 65)$=4.89, p=.010) that does not relate to mitigation and is therefore not discussed.

To further explore changes in glance behavior, three additional variables were analyzed: differences in 95th percentile glance duration, differences in mean glance duration, and differences in number of glances over 2 seconds. Differences in 95th percentile glance duration between the distraction and mitigation drives showed no significant main effect for mitigation ($F(2, 33)$=1.54, p=.23); however a significant interaction with driving environment was found ($F(4, 66)$=2.56, p=.047). Differences in mean glance duration again showed no main effect for mitigation ($F(2,33)$=0.12, p=.88), nor were any interactive effects significant. No main effect of mitigation was found for differences in number of glances over 2 seconds between the distraction and mitigation drives, ($F(2,33)$=1.26, p=.30), nor were there any interactive effects. Statistically significant effects were found for task ($F(3, 99)$=3.54, p=.018), driving environment ($F(2, 66)$=3.64, p=.031), and their interaction ($F(6, 198)$=4.77, p<.001). Effects not related to mitigation are not discussed further in this section.

For the 95th percentile glance duration, the differences in performance of the mitigation varied across the driving environments, as is illustrated in Figure 34. In the urban environment, real-time feedback increased the 95th percentile glance duration, whereas no mitigation and post-drive feedback resulted in a decrease during the mitigation drive. For the interstate, both the real-time and post-drive mitigation reduced the 95th percentile glance duration, whereas the participants in the no-mitigation group had no change. The effect in the rural environment was much smaller, with small increases for the real-time feedback and no-mitigation conditions and a small decrease for the post-drive feedback.

Overall, the post-drive feedback provided an improvement across driving environments, whereas the real-time system only provided a benefit in the interstate environment.

It is important to note that these differences in the extreme of the glance distribution (95th percentile glance duration), did not translate into differences in the central tendency of the glances as measured by mean glance duration.

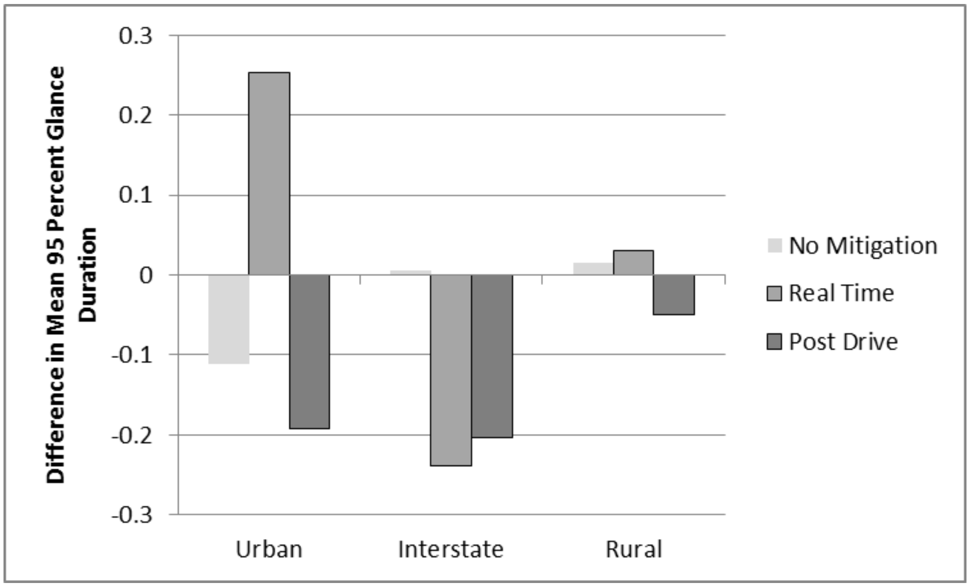


Figure 34 Difference in mean 95th percentile glance duration by mitigation and road type.

Results: The Influence of Mitigations on Driver Attitudes

Planned Behavior

Participants' attitudes towards distracting activities were measured with subjective ratings before the study began and after the study was completed. These attitudes toward engaging in distracting activities, needing to engage in these activities, and assessment of peers' likelihood of engaging in distractions provide an indicator of how distraction mitigation systems might influence driver behavior. The ratings were interpreted in terms of the theory of planned behavior, by examining the difference between the two measurements (Figure 35). Distraction mitigation might influence behavior by changing drivers' attitudes towards distractions and increasing their intention to defer distractions. Across the five questions aimed at determining whether participants would engage less in distracting activities while driving, the majority of participants said they would change the radio less ($F(1,33)$ = 11.15, p = *.0021)*. However, the desire to change the radio differed across the mitigations ($F(2,33)$ = 5.81, p = 0.007) with the drivers receiving the real-time mitigation (M = -0.08) stating that they would engage in this activity more, while those receiving either no mitigation (M = 0.33) or post-drive mitigation (M = 0.75) said they would do it less. All participants also said they would eat or drink less in the car (*F*(1, 33) = 5.97, *p = .020)*.

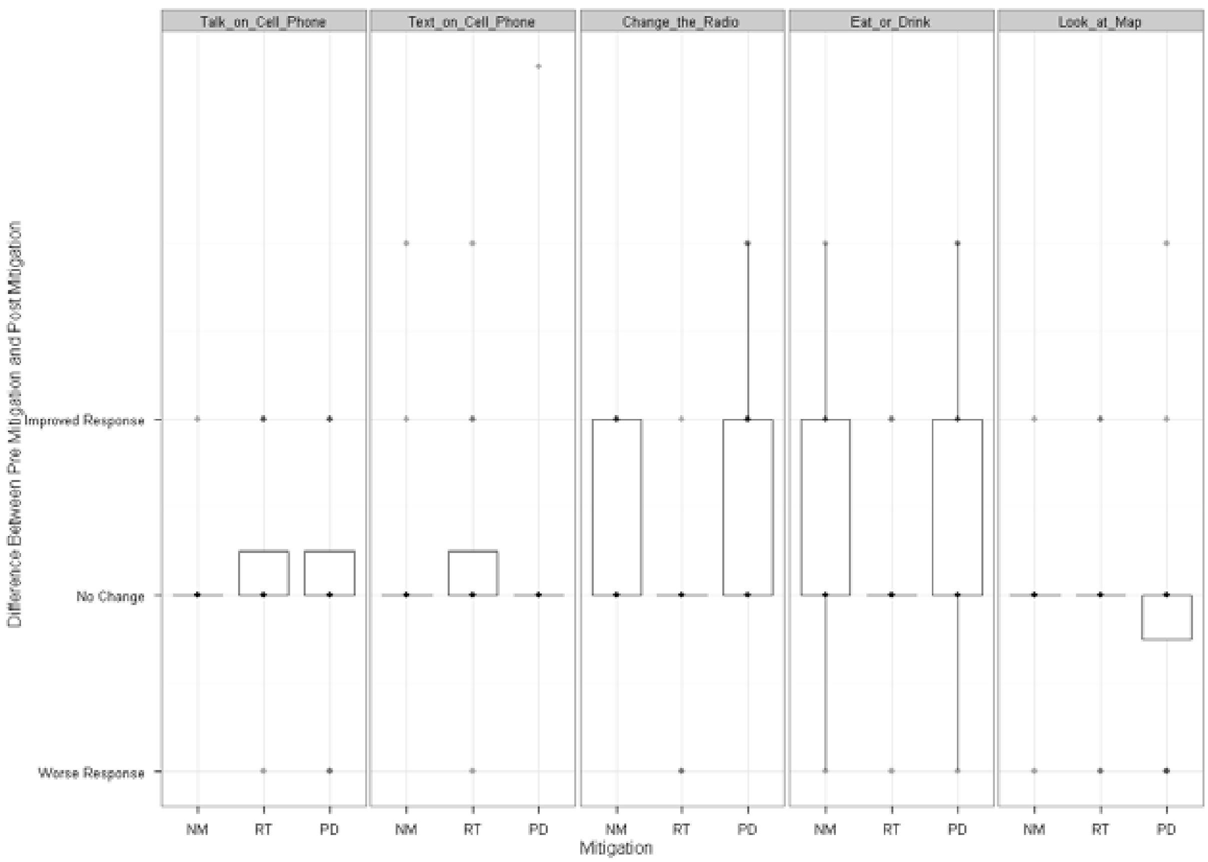


Figure 35 The difference between survey responses before and after the study for questions about the frequency of engagement in distracting activities while driving. Note: NM = No mitigation, RT = Real Time mitigation, and PD = post-drive mitigation.

In terms of intent to defer tasks (Figure 36), all participants said they would put off tasks until they pulled over (*F*(1, 33) = 22.34, *p* < .001) and that they would turn off distracting devices more (*F*(1, 33) = 12.12, *p* = .001). The mitigation had no effect on drivers' intentions to defer tasks.

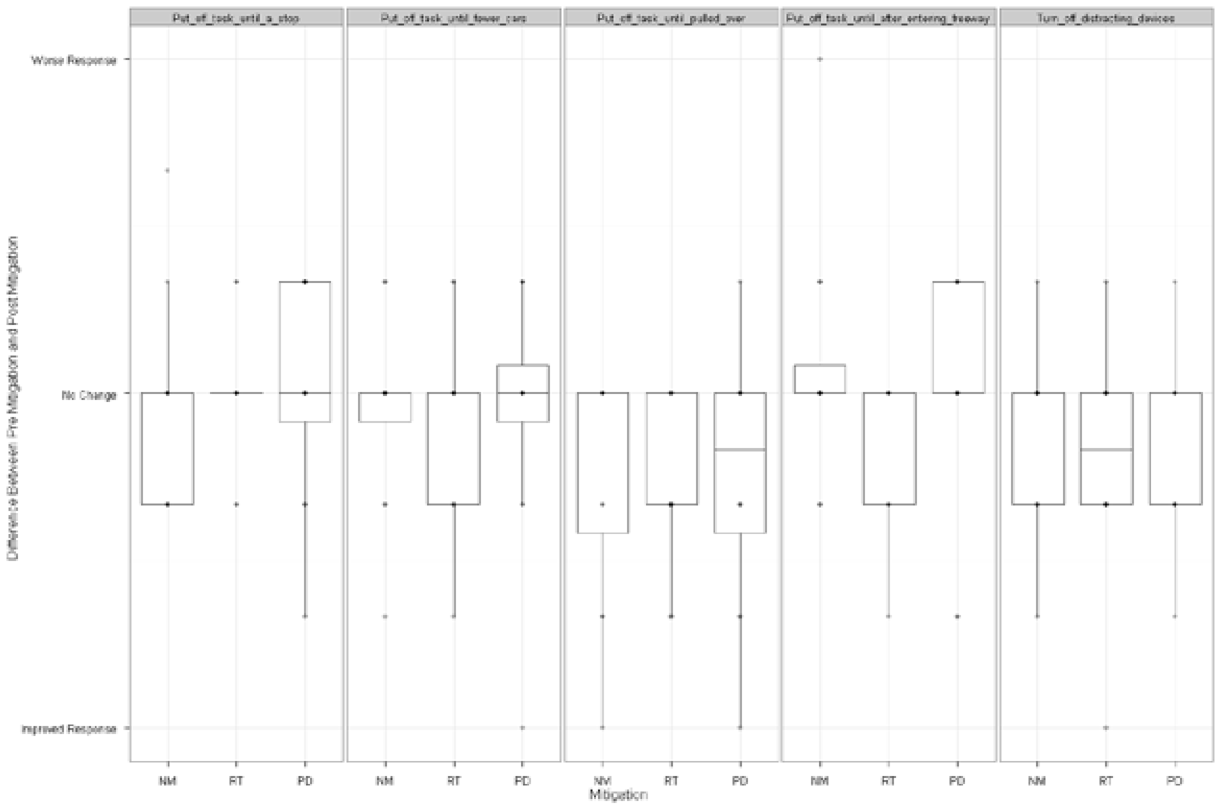


Figure 36 The difference between survey responses before and after the study for questions about the frequency of engagement in self-mitigation strategies while driving.

Besides changing the radio, there was not a strong difference between the mitigation groups as to how often they plan to engage in distracting activities or self-mitigation strategies (Figure 37). This suggests that while driving performance differed among the mitigation groups, their intention to engage in activities was not heavily influenced by the mitigation.

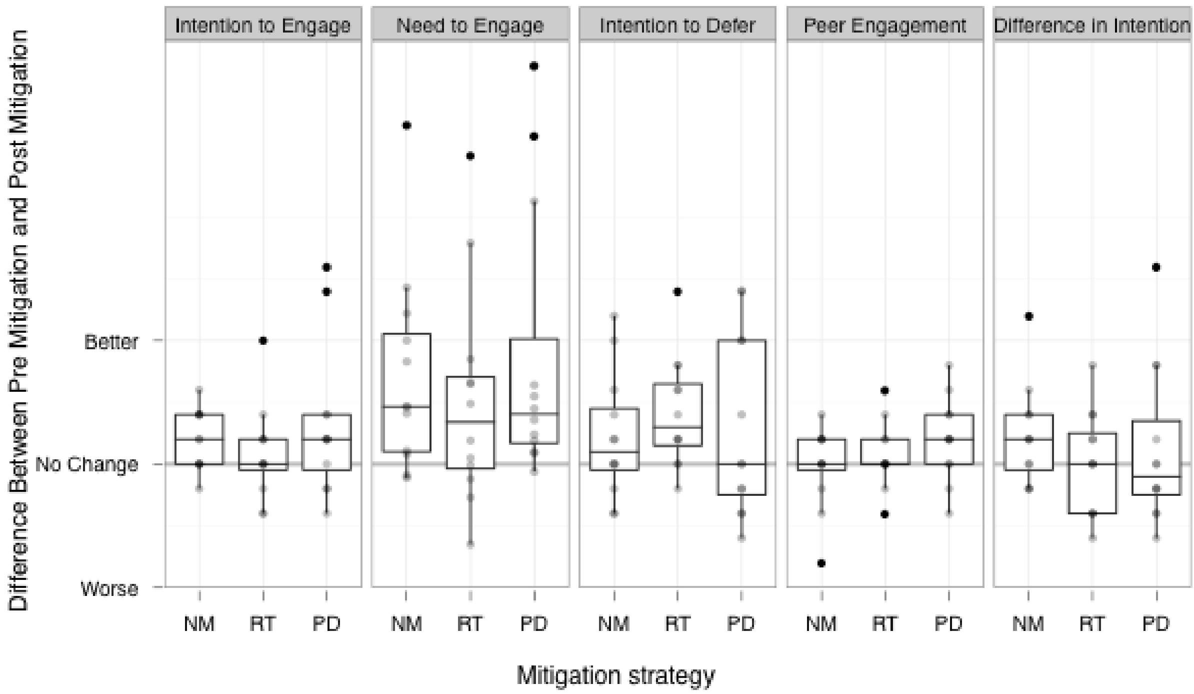


Figure 37 Responses associated with: Intention to engage, Need to engage, Intention to defer, Peer engagement, Difference between driver and peer intention to engage. None reach statistical significance.

Performance-Perception Calibration and Confidence

Comparing a participant's estimates of his or her performance to his or her actual driving performance determines whether the participant is calibrated. The difference between estimated and actual performance was calculated and analyzed to assess participants' awareness of their driving performance (Figure 38). To ensure that all measures could be compared to each other, lane departures were normalized to a 100-point scale using Equation 6.

Equation 6

$$100 * \frac{\textit{Number of lane departures} - \textit{Minimum number of lane departures}}{\textit{Max. number of lane departures} - \textit{Minimum number of lane departures}}$$

The differences for lane departures and speed were negated so that across all measures, a positive difference indicates that drivers performed better than their estimates and a negative difference indicates that drivers performed worse than their estimates. Post-drive feedback led to better calibration in almost all comparisons.

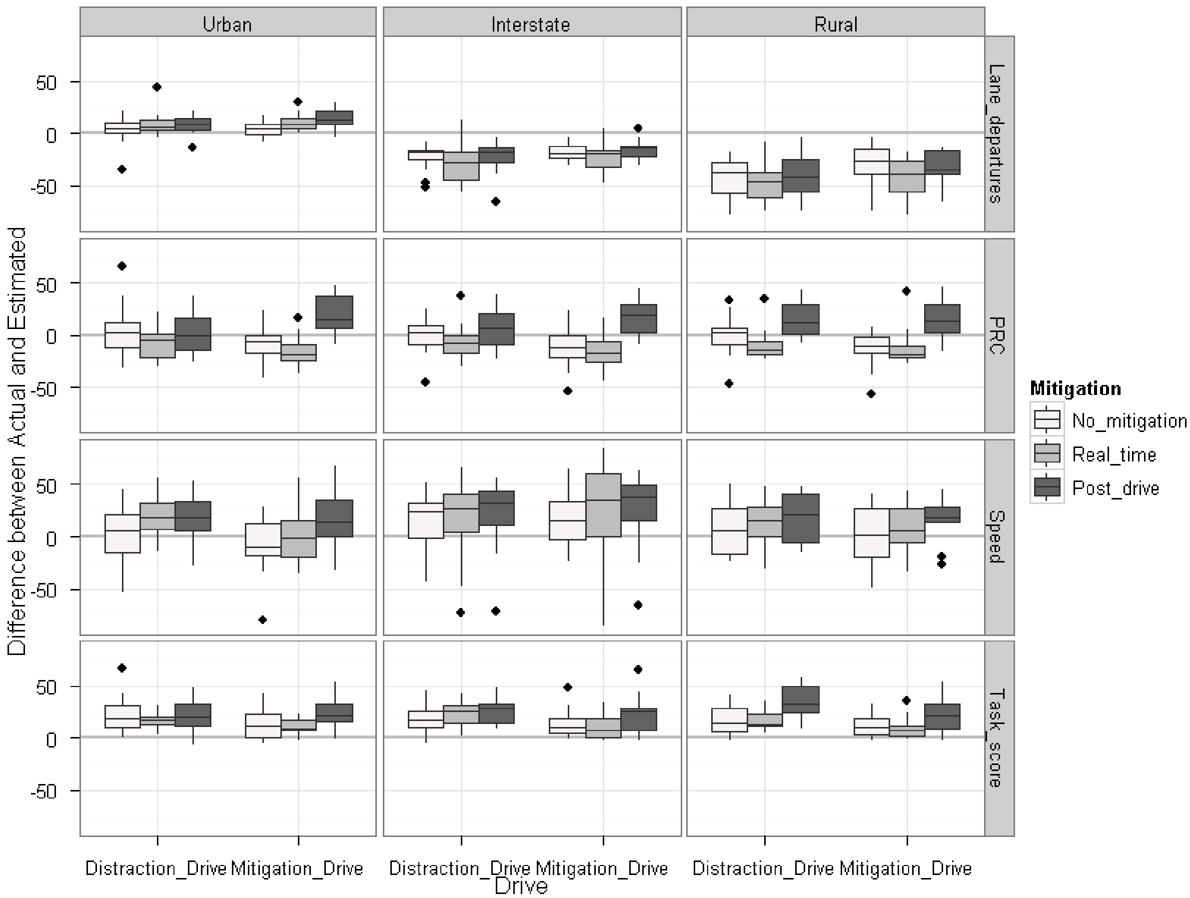


Figure 38 The difference between actual and estimated performance. Note: higher values correspond to greater underestimations of performance.

Lane Departures

Participants were asked to estimate the number of times they departed from their lane, where a lane departure occurred when one tire crossed the lane marking. This number was then compared to their actual number of lane departures, which were estimated during the simulation by monitoring the position of the left and right front corners of the car in relation to the lane lines. During the mitigation drive, drivers overestimated their performance less than during the distraction drive (Distraction: M = -20.45; Mitigation: M = -14.73) (F(1,33) = 10.73, p = 0.002). There was no significant effect of mitigation (F(2,33) = 0.68, p = 0.51) indicating that across all mitigation conditions, drivers overestimated their performance to roughly the same extent. Road environment had a significant effect (F(2,66) = 193.58, p <0.001), as drivers believed they departed from their lanes less than they actually did during the interstate (M = -21.74) and rural (M = -39.37) segments, while they believed they departed from their lanes more during the urban segment (M = 8.33). Over time, all drivers become better calibrated in terms of estimating lane departures. However, even with practice, calibration was not perfect, as most drivers continued to believe they departed their lanes less than in actuality. This overestimation of performance occurred during interstate and rural road segments, but not during urban road segments.

Task Score

In addition to lane departures, participants were asked to give an estimate of their average task score as indicated by accuracy, continuous attention, and promptness of their responses to all arrows, bug, and menu tasks. This number was then compared to the average of their actual task scores. During the mitigation drive, drivers underestimated their performance (i.e., believed they performed worse than they actually did) less than during the distraction drive (Distraction: M = 22.71; Mitigation: M = 15.88), $F(1,33) = 19.69$, $p < .001$. Similar to lane departures, with time and practice, drivers became better calibrated, but their calibration was still not perfect.

Speed

Participants were asked to estimate the percentage of the drive segment in which they drifted above or below the speed limit. This number was then compared to the number of times that drivers deviated 2.5 percent away from the posted speed limit. Mitigation did not have a significant effect, $F(2,33) = 1.12$, $p = .34$, as drivers overestimated their performance (i.e., believed they performed better than they actually did) to roughly the same extent. However, there was a significant interaction between the effects of drive (distraction or mitigation) and road environment, $F(2,66) = 3.33$, $p = .042$, with the means shown in Table 24. Post hoc analysis showed that the difference between actual and estimated speed was only significant during the urban section, indicating that with time, drivers overestimate their speed performance more.

Table 24 Means of the difference between actual and estimated performance of speed separated by drive and road type

	Urban	Interstate	Rural
Distraction	13.3	16.6	12.4
Mitigation	30.2	20.8	9.2

Percent Road Center

Participants indicated the percentage of time that they gazed at the forward roadway, i.e., through the windshield, on the performance questionnaire following each drive segment. This number was compared to the percent of time that they gazed at the road center, a circle of 10° centered on the most frequently gazed area of the road. Mitigation had a significant effect ($F(2,33) = 6.58$, $p = 0.004$) as those in both the no mitigation (M = -4.97) and real-time (M = -10.72) groups overestimated their performance (i.e., believed they performed better than they actually did), while the post-drive group underestimated their performance (M = 12.68). Mitigation and drive also interacted ($F(2,33) = 11.74$, $p < .001$), with the means shown in Table 25. Post hoc analysis shows that all mitigation groups changed significantly between drives. The no-mitigation group changed from underestimating performance to overestimating performance. The real-time group overestimated more during the mitigation drive. On the other hand, the post-drive group underestimated more during the mitigation drive. Note that although the interaction is interesting, the differences among the mitigation groups before any mitigation was presented (i.e., during the distraction drive) would make any conclusions drawn less reliable as there seem to have been strong group differences before the experiment began.

Table 25 Means of the difference between actual and estimated percent of gaze on road center by drive and mitigation condition

	No mitigation	Real-time	Post-drive
Distraction	1.5	-7.5	8.3
Mitigation	-11.5	-14.0	17.1

Overall, the results from the calibration study indicate that:

- Across time, all drivers became better estimators of both lane departures and task scores, but their calibration was still not perfect.
- The post-drive feedback caused participants to underestimate their performance in both task score and percent road center, indicating that they believed they performed worse than they actually did.

Skill Score

The skill score was used to evaluate participants' judgment of performance (Figure 39). It extends the calibration analysis and provides a more comprehensive measure that evaluates the human judgment in comparison to the actual state. An ideal skill score of one indicates that the human's judgment of performance is exactly the same as the actual state. A negative skill score indicates that one estimates their performance to be the opposite of the true environmental state. A skill score was calculated for each participant and was produced by aggregating responses across all measures (i.e., lane departures, task score, speed, and percent road center) and across all roads (i.e., urban, interstate, and rural).

Although the effect of mitigation was not significant ($F(2,33) = 0.25$, $p = .78$), participants' skill scores did improve after the mitigation drive ($F(1,33) = 9.53$, $p = .004$), indicating that with time, participants' judgment of their performance improved. The skill score was not decomposed because no significant main or interaction effects of mitigation were found.

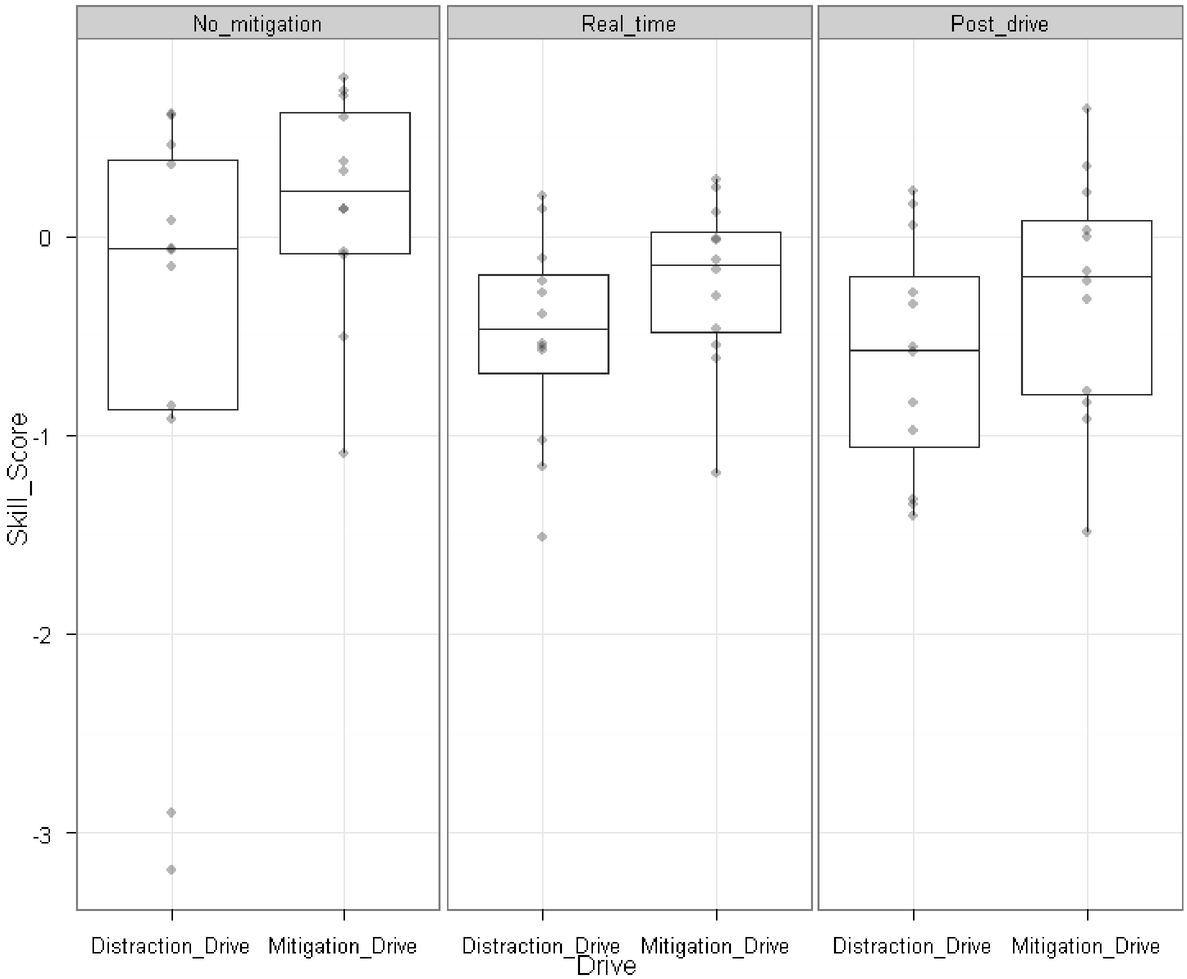


Figure 39 The skill score by drive and mitigation condition. Note: a higher score indicated more accurate judgments.

Driver Acceptance

Driver acceptance was assessed using the technology acceptance framework (Davis, 1989, 1993), which considers acceptance in terms of perceived ease of use, usefulness and value. Seven items defined ease of use and five items defined usefulness. Items defining ease of use included “would be easy to learn” and “would be distracting.” Items defining usefulness include “would make it easier to drive” and “would reduce distractions.” In addition to these measures of acceptance, three items assessed perceived value, for example, “At the actual price of $300, how likely would you be to consider purchasing a distraction warning system like the one you used during your study drive?” These items were measured on a Likert scale that ranged from “1 Strongly agree” to “7 Strongly disagree.” Some items were recoded so that lower values were associated with positive responses. The measures of acceptance and value were combined in an equally weighted average.

Figure 40 shows the relationship between ease of use, usefulness, and value. The clustering of small dots associated with the post-drive mitigation in the lower left shows it to be more acceptable and valuable to drivers. The numbers of points above and below the diagonal show that the post-drive mitigation was relatively easier to use than it was useful. The reverse was shown for the real-time system.

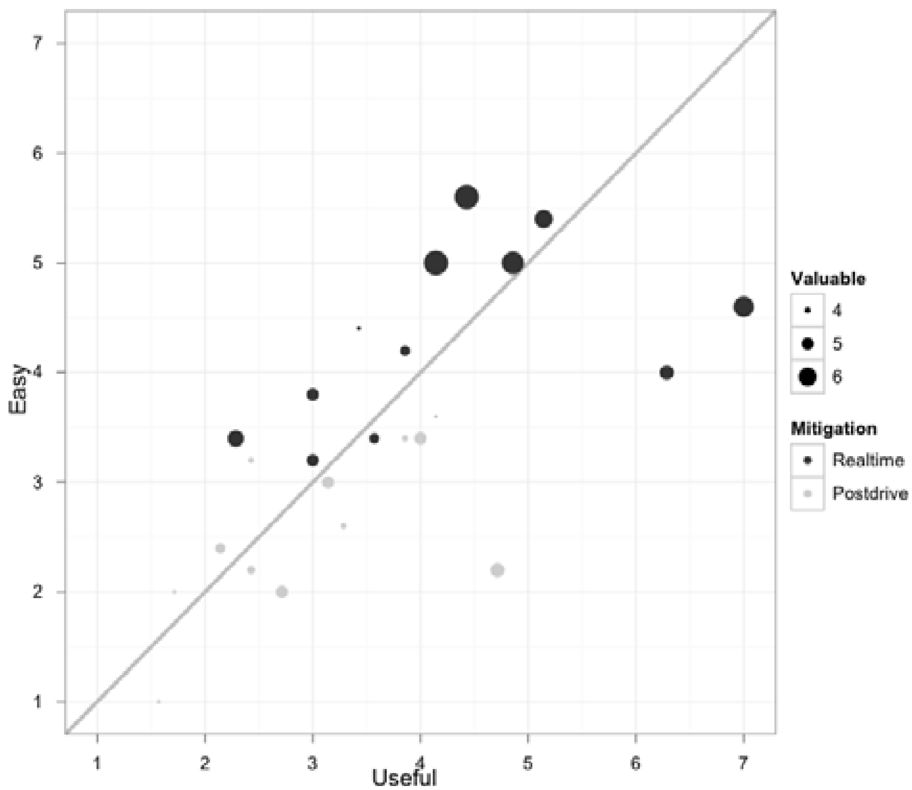


Figure 40 The relationship between ease of use, usefulness, and value. Note: Lower numbers indicate higher acceptance.

Figure 41 compares the mean acceptance ratings more directly. Drivers who experienced the post-drive system rated it as more useful than drivers' experiencing the real-time system ($F(1, 22) = 6.23$, $p = .020$). The ratings for ease of use and perceived value followed a similar pattern ($F(1, 22) = 29.76$, $p<.001$ and $F(1, 22) = 12.22$, $p = .002$). The effect size (eta squared) was greatest for ease of use (0.574), followed by value (0.221), and usefulness (0.220). To the extent the scales can be compared, drivers agreed that the post-drive feedback was fairly easy to use (M = 2.583) and useful (M = _3.012), and the real-time system was low in value (M = 5.583) when compared with the neutral rating of 4. Drivers tended to disagree with the statement "At the actual price of $300, how likely would you be to consider purchasing a distraction warning system like the one you used during your study drive?" (M =

2.000 with 1 being definitely would not consider and 5 being definitely would consider).

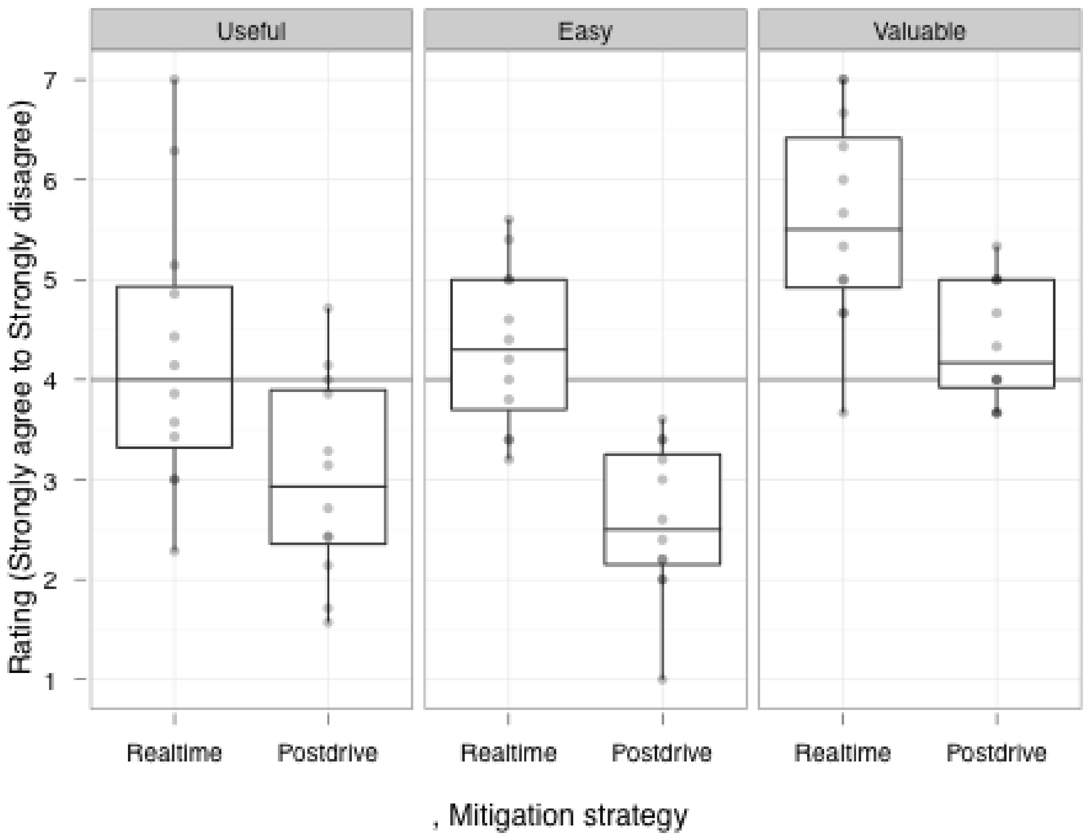


Figure 41 Ratings of acceptance. Note: Lower levels of acceptance are shown towards the top of the graph.

A linear regression model was developed to predict perceived value as a function of usefulness and ease of use. For these systems, ease of use dominates the perception of value, accounting for 42.9 percent of the perceived value variance ($t(22) = 3.26$, $p = 0.003$). Neither perceived usefulness nor the interaction between usefulness and ease of use reached statistical significance.

Discussion

These results inform both the development and evaluation of systems to mitigate distraction. They reveal the relative efficacy of two strategies to reduce distraction and suggest ways to improve such systems. The results also indicate how future systems might be evaluated.

Mitigation Effectiveness

The two mitigations tested in this study provide examples of how systems might try to affect driver interaction with distracting tasks. The real-time feedback system focused on changing driver behavior as it happened by alerting the driver when distraction was detected. The post-drive feedback system tracked driver distraction over the course of the drive and reported them in summary after each segment of the drive, along with information about unsafe driver behaviors that were related to the driver distraction. These two mitigation approaches resulted in subtle but distinct differences in driver response.

It was thought that mitigation might cause drivers to defer task engagement to less demanding parts of the drive. This did not occur with either mitigation tested. Although the post-drive feedback did result in deferral of the most intensive task, i.e., the bug task, that deferral did not result in lower roadway demand when the task was later begun. Real-time feedback produced no differences in deferral compared to no feedback, and no change in roadway demand occurred as a result of task deferral. Deferral was really only found with post-drive feedback given during the bug task, and its magnitude was small at 704 ms. When looking at the underlying data, even the long deferral times observed for the bug task were not substantial (see Table 26). Although the 95th percentile change in deferral time for the post-drive feedback was 6.5 times as great as when no mitigation is present, the 4.1 seconds provides only a limited opportunity for changes in the driving environment to reduce the risk of engaging in the task. The changes in deferral observed with the mitigation systems are off by at least an order of magnitude for the types of deferral that could reasonably provide the most benefit.

Table 26 Long deferral times for bug task

	Change In Deferral Time	
	75th Percentile	95th Percentile
No Mitigation	0.16	0.63
Real-time Feedback	0.51	1.64
Post-drive Feedback	1.47	4.10

Although drivers did not delay engagement in the task, there were some more compelling results related to the effect on visual scanning. Overall, drivers who received the post-drive feedback increased the amount of time they spent looking towards the roadway when engaged in visual distraction. Additionally, these drivers showed an improvement in reducing the 95th percentile glance duration across driving environments. This indicates that the post-drive feedback is often effective at improving driver attention to the road while decreasing unsafe glances away from the road.

Surprisingly, real-time feedback was intended to return drivers' attention to the roadway; however, it actually decreased drivers' focus on the roadway and only reduced the 95th percentile glance duration on the interstate. The data provides no clear explanation of this negative impact, but there are many potential explanations that should be considered: the drivers might have used the warnings to indicate when it was necessary to look back to the roadway, the complexity of the underlying warning algorithm and multiple warnings might have left them unsure how to respond, or due to the frequency of warnings, the drivers may not have trusted their accuracy and so they may have ignored the alerts. It is also not clear why the real-time feedback was effective at reducing 95th percentile glance duration on the interstate but not in the two higher demand driving environments. If this is indeed the case, it would suggest that real-time algorithms may benefit if they take into account visual sampling demands that vary by environment.

Looking at the net impact on driving performance and the ability of the driver to maintain lane keeping, the results provide a mixed message for both types of mitigation. The post-drive feedback provided improved lane keeping for the two most visually challenging distraction tasks, but degraded performance during the other two tasks. It also produced improved performance in the two high-demand driving environments (urban and rural), but decreased performance on the interstate. The real-time feedback, on the other hand, improved lane keeping with one of the four tasks (bug) and in one of three driving environments (urban). Overall, the feedback systems provide a benefit in some cases (higher demand tasks and environments) but decrease performance in others. Additional work is needed to understand why this is the case.

In terms of the intention to engage in distracting activities, mitigation did not have a significant effect on any of the secondary tasks mentioned, except changing the radio. The frequency of engagement in this activity differed across groups as both the no-mitigation and post-drive groups stated that they would engage in the activity less, while the real-time group stated that they would engage in this activity more. The cause for this difference is not clear, especially as the no-mitigation group received no feedback. Mitigation seemed to have had mixed results in terms of how it affected calibration. All drivers' lane departure and task score estimates became better calibrated during the mitigation drive. However, their calibration was nowhere near perfect, as drivers still overestimated their lane departure performance and underestimated their task score. Estimating the amount of time drivers spend drifting above or below the speed limit proved to be difficult, as almost all participants overestimated their speed control performance. In addition, there was no significant effect of mitigation, indicating that no evidence was found for an effect of feedback on speed control judgments. Participants' estimates of the amount of time they gazed at the road center indicated that the post-drive feedback caused drivers to underestimate their performance, making them under-confident in this skill. On the other hand, the no-mitigation and real-time groups overestimated their performance. However, this effect was confounded by the evidence, indicating that there were strong differences between the experimental groups before the experiment began.

Evaluation Protocol

The results of the study provide an important context for discussing ways in which the protocol was successful and ways in which it may need to be improved. The protocol was able to show changes in engagement with the distraction tasks over a very limited exposure, particularly for the post-drive feedback. The protocol was also able to detect complex relationships between the mitigation systems, the tasks, and driving environments. Although the protocol showed success in meeting the aims of this project, the results point to some changes that could make a distraction feedback protocol more effective.

The participants showed very limited deferral when they were asked to engage in tasks. The current protocol prescribes a fixed number of engagements with the forced-pace tasks that resulted in little down time between tasks particularly in the urban and rural environments. When the self-paced radio task is added to this, engagement was almost continuous across the drive for most participants (see Figure 11 in Chapter 4). The density of these tasks left little room for drivers to defer tasks to lower demand stretches of the road. While this may be appropriate to study systems that hope to delay engagement long enough for drivers to make an assessment of the driving environment, it would not be sufficient for evaluating systems that are designed to provide a gross change in behavior, such as deferring until a less demanding road type is reached or until stopped.

Building upon this, the limited deferrals produced even more limited opportunities for changes in driving demand to be observed, as was seen by the lack of results related to demand. This is exacerbated by the relative uniformity of demand in the urban and interstate environments. There was little opportunity for the driver to systematically shift task engagement to lower demand areas. The stops between the discrete changes in demand, which are represented by the driving environments to collect subjective assessment data, also limit the participants' ability to shift, for example, from the higher demand urban environment to the lower demand interstate environment.

The data about participant visual scanning also provides some interesting insights into the protocol. One of the primary measures of visual scanning was percent of time focused on the forward roadway; however, *increases* in this measure were an improvement when considering visual distractions, but *decreases* were an improvement for cognitive distraction. This asymmetry points to the importance of considering the types of changes in behavior that the protocol will assess relative to the mitigations being evaluated.

The results relating to 95th percentile glance duration and mean glance duration have important implications for an evaluation protocol. Relying solely on the mean glance duration, we would not have seen the differences between the mitigation approaches. When studying unsafe behavior, which often lies at the extremes of the distribution of data, we must avoid the tendency to focus on the central tendency and assess differences in the means. Instead, metrics of distraction and distraction mitigation should be developed that most closely represent the patterns of scanning behavior associated with increased crash risk.

Another important issue that became evident from the analysis of the data was the importance of the statistical approach. Two major approaches were available: one in which differences between performance in the two drives were explicitly considered and the primary effect of interest is mitigation, and another in which data from both drives is included in the statistical model and the primary effect of interest is the interaction between mitigation and drive. There are advantages and disadvantages for each approach. For the differences approach, the key advantages were in ease of understanding of the effects of the statistical model and of the variable being studied. The ability to focus on an analysis where a main effect is of primary interest has the inherent value of increased power. Differences have a certain inherent simplicity that makes them attractive for explaining the effect of the system on driver performance, as they provide a direct measure of the change we seek to examine. The risk is illustrated by the results of deferral for the bug task. When looking only at the differences, there is a clear message that the post-drive feedback produces a benefit relative to the no mitigation condition; however, when we looked at the data for the distraction and mitigation drives (see Figure 30), we saw that there were systematic differences in the groups that would have been missed if we had not looked beyond the differences. For the interaction approach, the key advantage is that all data can be considered in its raw form in the analysis, making it less likely that underlying differences between the groups would be missed. The key disadvantages are the greater difficulty in finding interactive effects, the greater difficulty of explaining them, and the lack of meaningfulness of the main effect for mitigation. They may be of greater concern in widespread application of an assessment protocol than in the university research environment. The present results point to the dangers of looking at differences without making sure that the underlying data do not exhibit systematic preexisting differences between the groups being studied.

An overall lesson from this is that care must be taken in the design of the protocol for assessing mitigation. The current protocol is a modification of a protocol designed to evaluate the effectiveness of distraction detection algorithms. While the protocol was effective in that regard, as discussed in Chapter 4, starting from that protocol as a basis for evaluating distraction mitigation created limitations on what the protocol could examine. Although a distraction detection evaluation needs to contain many opportunities to interact with a variety of distracting tasks, this may be counterproductive for examining the effectiveness of the mitigation in changing behavior. Additionally, forcing the driver to engage in distracting tasks in a variety of environments is needed to assess how detection systems work in a variety of situations. However, to assess the mitigation, it is necessary to see how drivers shift engagement in the tasks both within and across driving environments, actions which often represents systematic shifts in demand. Ultimately, a protocol to assess distraction detection may not be appropriate for assessing the effectiveness of a mitigation approach due to competing constraints.

Recommendations for Future Protocol Development

The development of a protocol for assessing distraction mitigation systems is a complex task to which this study provides great insight. Determining the outcome measures on which to determine system effectiveness is a topic open to debate, with arguments for driving performance, driver behavior, and driver attitudes. Even though this data is not conclusive, the results indicate that it is important to evaluate the effects of these systems on driving performance, behavior that leads to decrements in driving performance, and attitudes (calibration and acceptance). That is, we need to measure not only risky behavior and risky outcomes, but also how these systems affect driver understanding of risky behavior. Using this paradigm for evaluation, refinements of the protocol, additional development needs, and additional areas of research are recommended to finalize an effective overall assessment of distraction mitigation systems.

Driving Scenario and Distraction Tasks

Protocol to assess detection may not be appropriate for assessing effect of distraction mitigation.   The protocol included a variety of tasks that were very effective in terms of detecting the effects of distraction; however, the implantation of these tasks included the radio task as a filler task that would be available across the drive that the driver could engage with as needed. The drivers treated this task as they did the others, and most drivers, with few exceptions, engaged the task almost continuously across the drive. Although effective for a protocol for detecting distraction, this task did not fully meet the needs of a protocol for mitigating distraction.

Less time spent engaged in tasks. In terms of protocol development, the protocol used in this evaluation provided limited opportunities for deferral, although the extent to which drivers would delay engagement with the secondary tasks was a primary metric. Although this protocol was effective for distraction detection, drivers did not themselves delay these tasks for extensive durations despite the fact that they were not prohibited from doing so. To better assess the effectiveness of mitigation systems in causing increased deferral, a reduction is recommended in the ratio of time the driver is engaged in completing secondary tasks to the total time available to no greater than 0.5.

Strategic placement of driving demand levels across drives. For mitigation assessment, more so than for distraction detection, high and low demand roadway segments need to be strategically placed so that the effect of mitigation on deferral from high to low demand environments can be assessed systematically.

Self-paced tasks are important, but not for filler tasks. The use of the filler task seems to complicate driver choice of deferment. With greater latitude in when to engage in the task, it will be possible to better assess when and how drivers choose to defer.

A variety of distracting tasks is needed, including self-terminating tasks. This protocol included several tasks that differed in terms of attributes including demand and pace; however, all tasks were system-terminated versus self-terminating. In many situations drivers can choose to apply additional resources to finish a task more quickly rather than applying fewer resources over a longer period of time. However, in the current protocol, task duration was fixed. If the driver instead had a goal to accomplish that would terminate the task, drivers might utilize different strategies than currently observed. It is important to understand how mitigation systems might affect these types of engagements.

Consider self-scheduling of tasks. In the current protocol, drivers were forced to complete the tasks in a predetermined order; however, in a non-experimental setting drivers are free to choose to do the tasks as they wish. If tasks were freed from a sequential requirement, it would be possible to assess if a mitigation system causes drivers to differentially engage in the tasks across demand level of the drive.

Short repeated drives are preferable. The current drive database may not provide the best approach to mitigation evaluation. Shorter drives would allow for repeated exposure to the mitigation, environments, and tasks, and would facilitate better calibration data. The drives need not be identical, but it would be important to have drives of sufficient length for each to have two discreet levels of demand.

Measurement Development and Analysis

Analysis of demand is contingent on deferral. It became clear from the results that the analysis of the demand only made sense if the drivers did indeed defer. The recommendations would be to make the analysis of demand a conditional analysis, and to provide more opportunities for deferral under varied demand.

Safety criticality of outcome measures needs to be evaluated. Due to the lack of crash events in this protocol, SDLP was evaluated as the primary indicator of safety. Although this measure has shown great value in studying impaired driving, examination of the SHRP 2 data to insure that it is the most effective at conveying safety relevance in distracted driving is warranted.

Consider changes in the extremes, not just changes in the mean. As was evident from looking at the glance duration data, changes in mean performance do not always provide an accurate indication of the actual effect of the mitigation. For this reason, it is critical to look at the tails of the distribution rather than just the central tendency.

Careful consideration to implications of changes in dependent measures is critical. As was evidenced by PRC, some dependent measures behave differently with regard to different types of distraction, and the effect of the mitigation. Although increased PRC indicates improved attention to the driving environment in general, that is not the case with cognitive distraction where the driver is not attending robustly enough to the sides of the driving environment. In that case, a decrease in PRC is desirable. The cautionary tale is that directionality of an eye metric like PRC is not always obvious in terms of what's better/worse for a particular distraction, and care should be taken in interpreting the effects of systems.

Looking only at differences in performance is risky. Although looking at differences in performance between the mitigation and distraction drive is compelling from a protocol standpoint due to the simplification of the analysis, the results from this study indicate that there is significant risk in this approach. Differences do not tell the whole story and may mask potentially important differences associated with sampling, as was evident with deferral of the radio task.

More robust scale needed for assessing planned behavior. One of the challenges is accurately assessing how long-term driver behavior might change as a result of the mitigation. Using the data available before the study, an approach was developed to assess how drivers' planned behavior and willingness to engage in tasks while driving changed through feedback; however, the instrument developed was not able to capture the differences as well as anticipated. A more robust scale that is grounded in a sample of reported deferment strategies is crucial for effectively gathering this type of data.

Measures of driving performance used for comparison to questionnaire data must correspond to what the driver can perceive of their performance. There are challenges when attempting to assess how mitigation systems affect driver perception of their performance on the driving task due to the fact that what the driver perceives as their performance may not match what is reported by the data. An example of this would be lane departures: at what point is a lane departure a lane departure? In the simulator data, this occurs as soon as a single tire leaves the lane, even if by a fraction of an inch; however, the driver may not be able to perceive a lane departure until the vehicle is several inches beyond the lane boundary. The fact that there may be many of these departures throughout the drive that the driver does not or cannot perceive creates challenges for understanding the effect of the system by adding noise to the comparison. Care needs to be taken to ensure that the comparison is relative to those situations that the participant could perceive.

Keeping surveys up to date with changes in technology and driver behavior is a challenge. With the constant change in the types of technology that can be used in a vehicle, the surveys may need to be constantly updated to be consistent with the distraction behaviors of drivers.

Keep in mind constraints of sample size and number of data points when choosing instruments. The skill score analysis for this study provides an important example of the necessity to match the data collected with the availability of data. Some instruments that could be useful require much more data than might actually be available in this type of assessment. As appealing as certain types of data might be, they should only be implemented if the instrument can provide useful data with the sample size available.

Driver acceptance can be measured reliably. The measures of driver acceptance that were used in this study behave in a sensitive and reasonable manner and were able to differentiate between the two systems. This approach was grounded in previous research, and the fact that it was diagnostic in this study indicates that it should be used in the protocol moving forward.

CHAPTER 6. ASSESSING THE BENEFITS OF DISTRACTION DETECTION AND MITIGATION SYSTEMS

Distracted driving exacts a substantial societal cost. Of the 33,808 total fatalities in 2009, 5,474 fatalities were associated with distraction. Distractions accounted for 16 percent of the total traffic fatalities, up from 10 percent in 2005. In addition, distracted driving accounted for 448,000 injuries in 2009, 20 percent of the total (NHTSA, 2010). A variety of vehicle technologies promise to counter the effects of distraction, improve driver safety, and mitigate its overall costs.

The degree technology designed to mitigate distraction achieves its promise depends on whether it reduces crashes and associated deaths, injuries, and property damage. Directly measuring such benefits is quite difficult. Even where an intervention, such as handheld cell phone and texting bans, is implemented on a specific date with a clear intent, a wide array of confounding factors make its benefits difficult to quantify (IIHS, 2010; McCartt et al., 2010). These difficulties make indirect estimation methods necessary.

The purpose of this section is to describe an indirect method to estimate the degree to which distraction countermeasures reduce crash risk and the associated costs, where costs include property damage, injuries, and deaths. This method could also apply to costs from congestion-related delays associated with distracted drivers.

Several indirect estimation methods have been developed and applied to a range of vehicle technologies—from collision warning systems to intelligent infrastructure (Burgett et al., 1998; Misener et al., 2001; Najm et al., 2005). The application of these methods to estimate the safety benefit of collision warning systems is perhaps most similar to the challenge of estimating the safety benefit of distraction mitigation systems. One such benefit analysis estimated that drivers could avoid 60 percent of rear-end crashes if they had 0.5 seconds more time to respond, and a second to respond would prevent 90 percent of all rear-end crashes (Ankrum, 1992). Driver response to the technology has a powerful influence on its benefits, such as benefits acquired through the influence of interface characteristics on driver response (Scott & Gray, 2008). A central challenge to benefit estimation is estimating the effectiveness of the system in preventing prototypical crashes. For this reason, the benefits estimation methods of Najm and his colleagues provide a promising starting point for estimating the benefits of distraction mitigation systems (Najm, 2003).

EXTENDING BENEFIT ESTIMATION TO DISTRACTION DETECTION AND MITIGATION SYSTEMS

Benefit estimation techniques developed for assessing collision warning systems face important challenges when applied to distraction mitigation systems. These challenges stem from two primary sources: diversity of distraction mitigation systems and diversity of mechanisms underlying distraction-related crashes. Chapter 3 of this report describes the diversity of distraction mitigation systems, which can range from alerting drivers to long glances, adjusting collision warning thresholds based on distraction, and to post-drive feedback. In comparison, the functionality of collision warning systems is much more homogeneous.

The diversity of distraction mitigation systems reflects the diversity of mechanisms underlying distraction. For collision warning systems, the underlying mechanism governing the benefit is a reduction in drivers' reaction time. In contrast, distraction mitigation systems might affect driver reaction time to imminent crashes, discourage dangerously long glances that lead to crash situations, or even discourage drivers from engaging in certain distractions. The diversity of distraction mitigation systems and diversity of distraction mechanisms underlying crashes require modification of the benefit estimation techniques previously applied to collision warning systems.

Benefit estimation typically considers historical crash data to establish a baseline cost associated with various crash types that the system aims to reduce. The benefit is defined by how effectively the technology reduces the likelihood and severity of each crash type summed across all crash types. Crash types typically reflect a kinematic description of the collision, such as rear-end or roadway departure crashes. Relating crash types to benefit estimation follows the same general approach used in other risk analyses, where a "risk triplet" defines risk in terms of the situation, consequence and likelihood (Kaplin & Garrick, 1981). Benefits accrue by reducing crash frequency or severity. Equation 7 defines this relationship more formally.

Equation 7

$$Benefit = CurrentCost - \sum_{i}^{Crashtypes} \sum_{j}^{Severitylevels} Cost_{i,j} \cdot Frequency_{i,j} \cdot \frac{P(Crash_{i,j} \mid System_{i,j})}{P(Crash_{i,j})}$$

where the current cost is defined by Equation 8.

Equation 8

$$CurrentCost = \sum_{i}^{Crashtypes} \sum_{j}^{Severitylevels} Cost_{i,j} \cdot Frequency_{i,j}$$

Considering crash types and levels of crash severity separately is particularly important in estimating benefits of distraction mitigation systems. Findings from naturalistic driving studies show that distraction and inattention differentially affect crash types and severity. Crashes and near-crashes are over-represented by inattentive drivers—Nearly 80 percent of crashes occurred with some form of inattention compared to 65 percent of near-crashes, and less than 30 percent of incidents. Similarly, some crash types are particularly common for inattentive drivers— 93 percent of rear-end collisions involved inattention as a contributing factor compared to 78 percent for all crashes (Klauer et al., 2006). Consequently, Equation 7 includes crash types and severity levels separately to estimate benefits more precisely.

Calculating the benefit using Equation 7 and Equation 8 requires historical data regarding the frequency and cost of distraction-related crashes, and also on how effectively the system prevents crashes. Existing historical crash data can define the frequency and severity of various crash types (Wang et al., 1999). System effectiveness is considered in the last portion of Equation 7—the conditional probability of a given crash type and severity if the vehicle had been equipped with a distraction mitigation system. This parameter can be estimated from a combination of Monte Carlo models and driver responses observed in simulators or naturalistic driving situations (Najm, 2003). The diversity of mitigation systems and mechanisms of distraction make estimating system effectiveness particularly challenging.

Figure 42 shows a framework for estimating the effectiveness of distraction mitigation systems. The top of the figure shows the distraction-related crashes and the associated distractions. Mitigation systems are represented in the center of the figure. The benefits of mitigation systems can accrue by discouraging drivers from enabling distracting devices, engaging in distracting activities, and persisting in distracting activities that put them into crash-imminent situations. System effectiveness is the proportion of crashes prevented by a combination of all three mitigation mechanisms. Each mitigation mechanism can be viewed as a layer of protection eliminating a certain proportion of crashes. The dotted lines represent potential crashes and the fewer lines after each layer indicate the cumulative effectiveness of the system. Assuming the mitigation system acts on each mechanism independently its overall effectiveness is the product of its effects on each distraction mechanism.

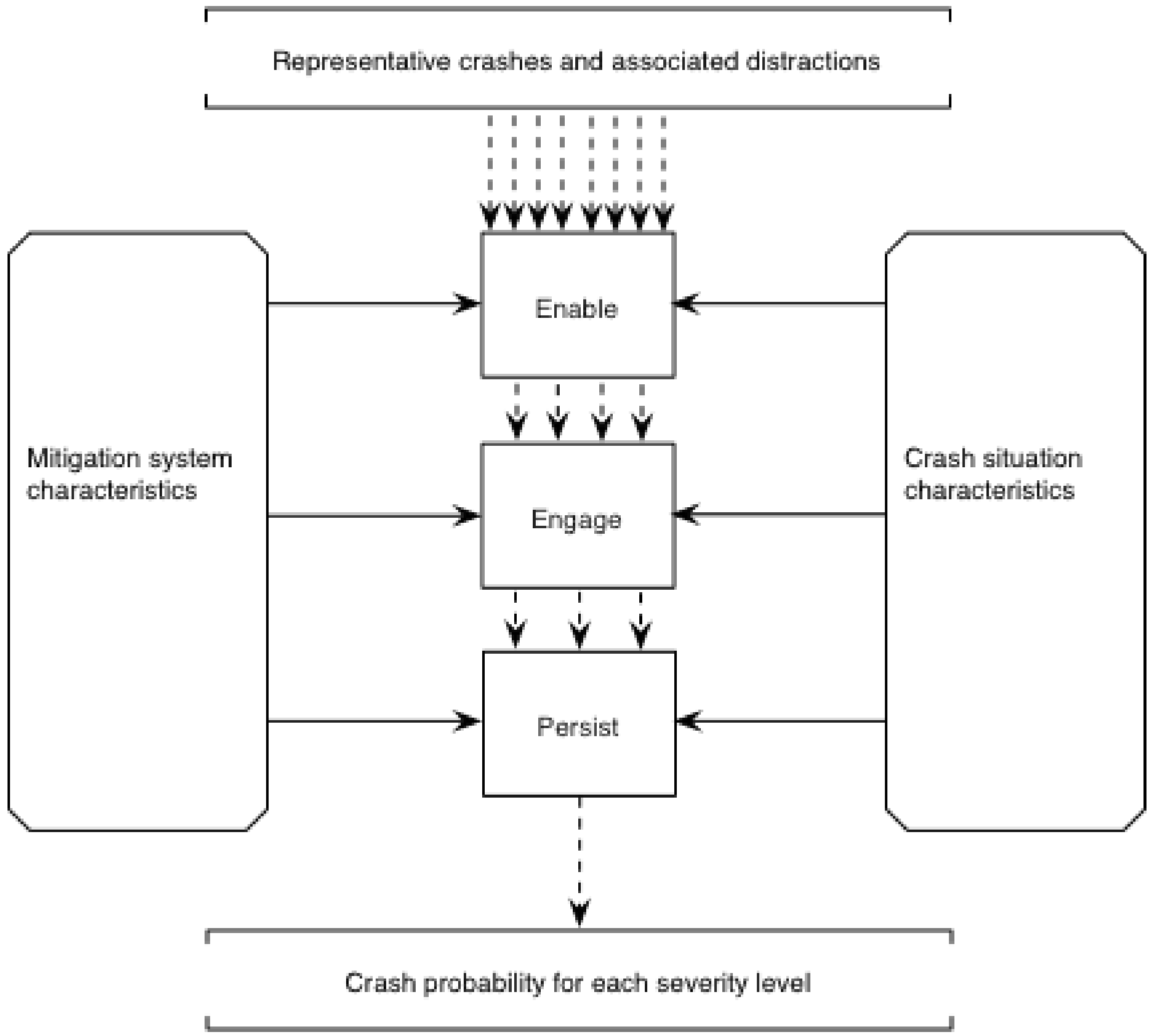


Figure 42 System effectiveness estimation framework for distraction mitigation systems.

The effectiveness of each layer of distraction protection depends on the particular characteristics of the mitigation system and the crash situation. The left side of Figure 42 shows that some systems have little influence on drivers' tendency to engage in distractions, such as those that modulate collision warning thresholds, whereas others address engagement directly, such as post-drive feedback. Additionally, some systems are sensitive to some types of distraction and not others, such as those that use head pose data to estimate when drivers' eyes are off the road, compared to systems that use eye movements to estimate cognitive distraction. Systems that do not detect cognitive distraction might not change the associated probability of crashes.

The right side of Figure 42 shows the influence of the particular characteristics of the crash situation on system effectiveness, such as the time available to respond. For example, the effectiveness of a distraction mitigation system that warns drivers of long glances away from the road depends on the time available for the driver to respond to the crash situation after receiving the warning. If dangerously long glances are detected only after the crash, then the system would be ineffective. This consideration is similar to the kinematic constraints used in estimating collision warning effectiveness (Burgett et al., 1998).

The specific steps associated with the calculations in Equation 9 begin with selecting representative distraction-related crashes and concludes by estimating driver response. These steps are briefly described.

Equation 9

$$SystemEffectiveness_{i,j} = \frac{P(Crash_{i,j} \mid System_{i,j})}{P(Crash_{i,j})} = \frac{\sum_{k}^{RepresentativeCrashes} \prod_{l}^{Mechanisms} MitigationMechanism_l}{RepresentativeCrashes}$$

1. Identify a representative sample of distraction-related crashes

The analysis requires a representative sample of crashes or near crashes for each crash type identified in Equation 7 that is large. Describe the range of pre-crash situations including the type of distraction associated with the crash as a critical element. The crash events can be estimated from ongoing driving studies, such as SHRP2 and the 100-car naturalistic driving study (Klauer et al., 2006). These crashes could be clustered into prototypical crash scenarios and weighted by their frequency (Najm, 2003).

2. Define the mitigation system

The mitigation system needs to be defined in sufficient detail to support an estimate of its ability to prevent the crashes described in Step 1. This includes the types of distraction it can detect, such as the distinctions in this report between distractions associated with visual, extreme visual-manual (reaching), and cognitive distraction. The system should also be described in terms of the proportion of drivers it would cover. This is particularly relevant for systems that use eye-tracking systems that might not operate reliably for all eye types and driver anthropometrics.

3. Define the crash configuration

The crash configuration needs to be defined in terms that describe the timing of the events that would trigger the distraction mitigation system, such as long glances. This description also needs to describe the crash kinematics to define the time available to respond—the window between when an alert would occur and when the crash would, given the pre-crash situation.

4. Estimate driver response

Figure 43 shows the major components of a simple model to estimate driver response. Brown et al. (2001) developed this model to compare different algorithms and parameter choices for a wide range of roadway conditions. The model takes as input the state of the following and the lead vehicles at the point where the imminent collision situation begins to evolve, which is the moment the lead vehicle begins to decelerate. This input includes the velocity and position of the following vehicle as well as the position, velocity, and deceleration of the lead vehicle. The input also includes the warning algorithm and its parameters. The output describes the vehicle states over time, the driver's response process, and the internal state of the driver. It uses the acceleration, velocity, and position of the two vehicles as the collision situation evolves to characterize the state of the vehicles. The model in Figure 43 focuses on driver attention in the seconds preceding a crash. Entirely different models are needed to describe drivers' more strategic decisions regarding whether to enable or disable a distracting system and engage in a distracting activity.

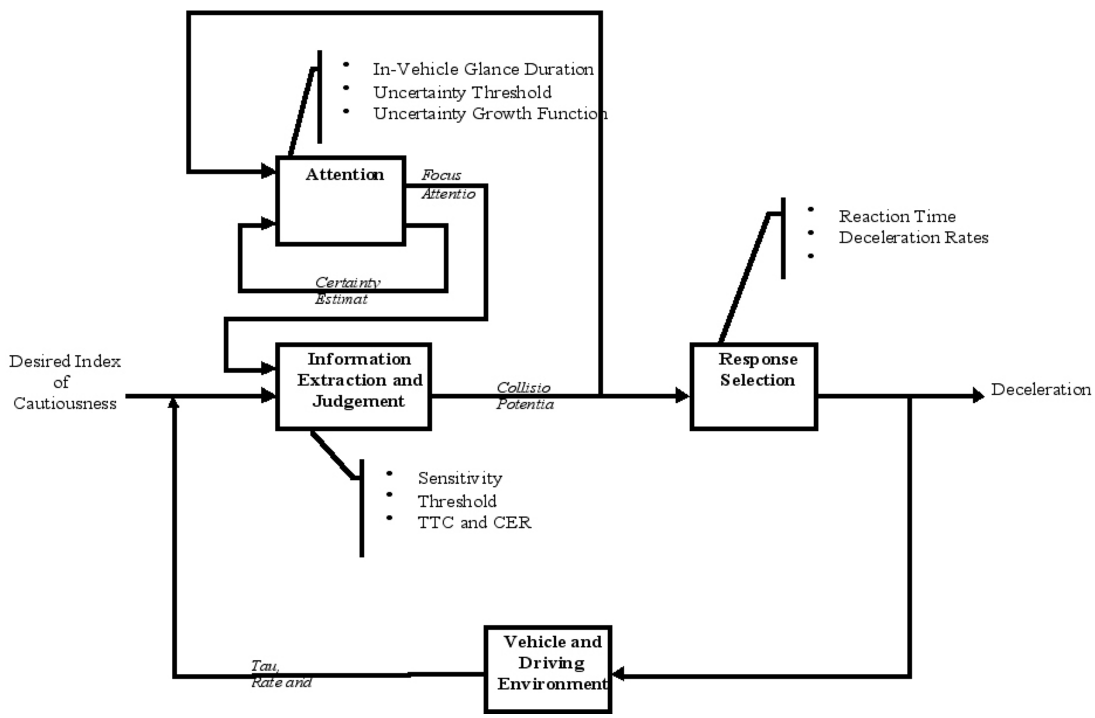


Figure 43 Model of driver response for assessing collision warning effectiveness that might be adapted to estimate distraction mitigation effectiveness (Brown et al., 2001).

Figure 44 shows the output of the model with probability of collision displayed as a function of the assumed reaction time and deceleration of the driver. Similar estimates of crash probability would estimate the contribution of driver response to overall system effectiveness.

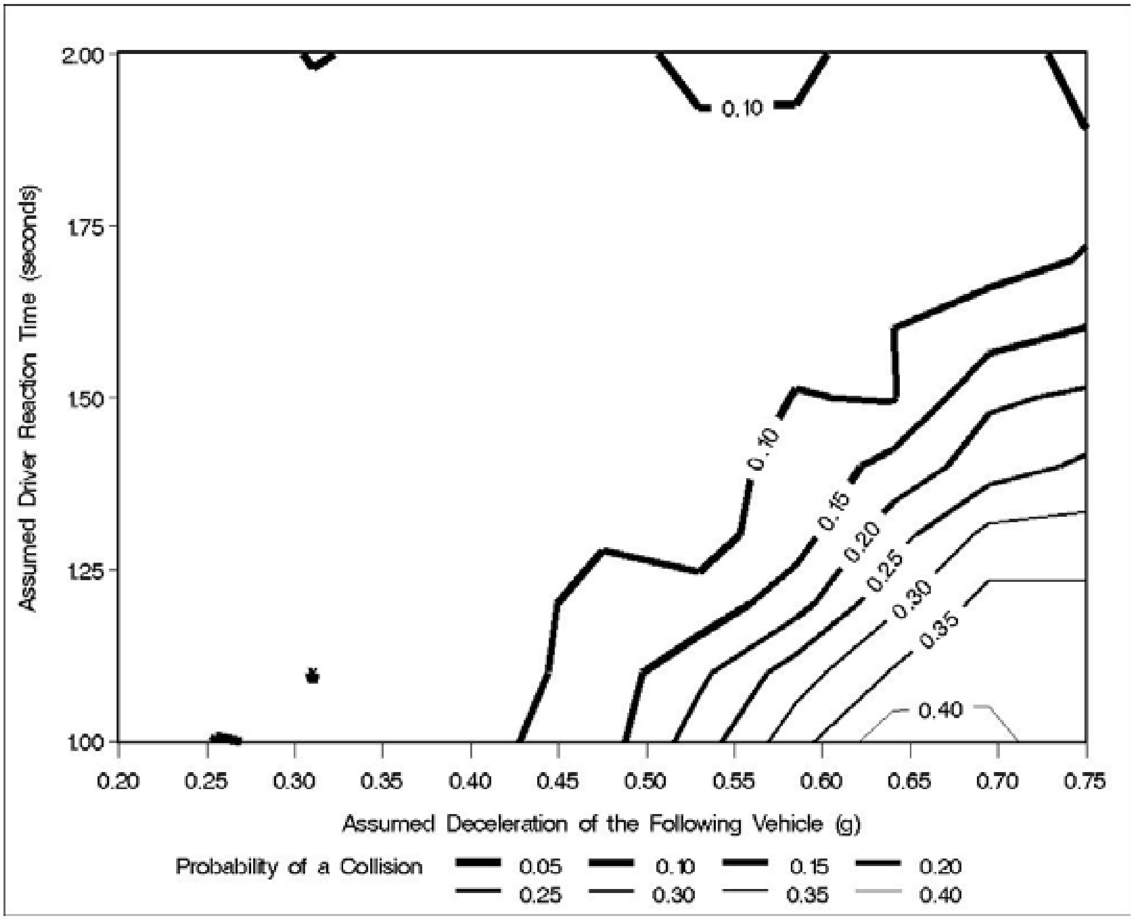


Figure 44 Model output indicating estimated system effectiveness as a function of model parameters (Brown et al., 2001).

SIMPLIFYING ASSUMPTIONS AND EXTENSIONS TO THE PROPOSED METHOD

This benefit analysis makes several notable simplifying assumptions. Most critically, several assumptions of independence might not be justified. One such assumption concerns the relationship between vehicles equipped with the systems and the drivers of those vehicles. Similar to other advanced safety systems, they are likely to be more prevalent in high-cost vehicles where the rates of distraction related crashes and their consequences might be different than the typical vehicle. Most notably such vehicles are less likely to be driven by young drivers who might be most likely to engage in distractions (Braitman & McCartt, 2010). As a consequence, this analysis might overestimate the benefit of distraction mitigation systems. The analysis also assumes the mitigation mechanisms influence crash outcomes independently; however, it is likely that they are interdependent and generate a larger benefit than one might expect if they were to act independently (Donmez et al., 2008). Careful assessment of these assumptions is needed as this benefits analysis is developed.

A critical challenge to estimating benefits associated with distraction mitigation systems concerns how to model the full breadth of their influence. These systems have a relatively predictable influence on distraction in the moment (as depicted by Figure 43). They are also likely to change driver behavior over time and even shift societal norms.

The greatest benefits are likely to accrue from any change in safety norms they might induce, but such changes are difficult to model. More concretely, compliance, reliance, and acceptance are more directly related to system features. Initial models of these phenomena suggest it might be possible to include them in benefits analysis (Gao & Lee, 2006; Misener et al., 2001; Taylor & Todd, 1995).

REFERENCES

Aasman, J., Mulder, G., & Mulder, L. J. M. (1987). Operator effort and the measurement of heart-rate variability *Human Factors, 29*(2), 161-170.

Akamatsu, M. (2009). Japanese Approach to Principles, Codes, Guidelines, and Checklists for In-Vehicle HMI. In M. A. Regan, J. D. Lee & K. L. Young (Eds.), *Driver Distraction: Theory, Effects, and Mitigation* (pp. 425-443). Boca Raton, FL: CRC Press.

Al-Shihabi, T., & Mourant, R. R. (2003). Toward more realistic driving behavior models for autonomous vehicles in driving simulators *Human Performance, Simulation, User Information Systems, and Older Person Safety and Mobility* (pp. 41-49).

Alexander, G., & Lunenfeld, H. (1990). *A user's guide to positive guidance.* (Report No. FHWA-SA-90-017). Washington, DC: Federal Highway Administration, Office of Traffic Operations.

Angell, L. S., Auflick, J., Austria, P. A., Kochar, D., Tijerina, L., Biever, W. J., et al. (2006). *Driver Workload Metrics Project: Task 2 Final Report*. (Report No. DOT HS 810 635). Washington, DC: National Highway Traffic Safety Administration. Available at http://www.nhtsa.gov/DOT/NHTSA/NRD/Multimedia/PDFs/Crash%20Avoidance/Driver%20Distraction/Driver%20Workload%20Metrics%20Final%20Report.pdf

Ankrum, D. (1992). Smart vehicles, smart roads. *Traffic Safety, 92*(3), 6-9.Bensinger, K. (Producer). (2008) Chrysler will offer wireless Internet access in 2009 models. *Los Angeles Times*.

Bergasa, L. M., Nuevo, J., Sotelo, M. A., Barea, R., & Lopez, M. E. (2006). Real-time system for monitoring driver vigilance. *IEEE Transactions on Intelligent Transportation Systems, 7*(1), 63-77.

Berka, C., Levendowski, J. D., Cvetinovic, M. M., Petrovic, M. M., Davis, G., Lumicao, N. M., et al. (2004). Real-time analysis of EEG indexes of alertness, cognition, and memory acquired with a wireless EEG indexes of alertness. *International Journal of Human Interaction, 17*(2), 151-170.

Boer, E. R. (2000). *Behavioral Entropy as an Index of Workload*. Paper presented at the 44th Annual Meeting of the Human Factors and Ergonomics Society (HFES2000).

Boer, E. R., Rakauskas, M. E., Ward, N. J., & Goodrich, M. A. (2005). *Steering entropy revisited*. Paper presented at the Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, Rockport, MN.

Brennan, S. E. (1998). The Grounding Problem in Conversations With and Through Computers. In R. J. Kruez (Ed.), *Social and Cognitive Approaches to Interpersonal Communication* (pp. 201-225). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Braitman, K. A., & McCartt, A. T. (2010). National reported patterns of driver cell phone use in the United States. *Traffic Injury Prevention, 11*(6), 543-548. doi: 10.1080/15389588.2010.504247

Brookhuis, K. A., de Vries, G., & de Waard, D. (1991). The effects of mobile telephoning on driving performance. *Accident Analysis and Prevention, 23*(4).

Brouwer, R. F. T., & Hoedemaker, D. M. (2006). *Driver support and information systems: experiments on learning, appropriation, and effects of adaptiveness.* (No. AIDE IST-1-507674-IP Deliverable D1.2.3).

Brown, T. L., Lee, J. D., & McGehee, D. V. (2001). Human performance models and rear-end collision avoidance algorithms. *Human Factors, 43*(3), 462-482.

Burgett, A. L., Carter, A., Miller, R. J., Najm, W. G., & Smith, D. L. (1998). A collision warning algorithm for rear-end collisions. Washington, DC: National Highway Traffic Safety Administration.

Byrne, E. A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological Psychology, 42*(3), 249-268.

Caird, J., Willness, C., Steel, P., & Scialfa, C. (2008). A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis & Prevention, 40*(4), 1282-1293.

Carsten, O., Merat, N., Janssen, W. H., Johansson, E., Fowkes, M., & Brookhuis, K. A. (2005). *HASTE Final report*: E.U.European Commission.

Chisholm, S. L., Caird, J. K., & Lockhart, J. (2007). The effects of practice with MP3 players on driving performance. *Accident Analysis & Prevention, 40*, 704-713.

Clark, H. H., & Brennan, S. (1991). Grounding in communication. In L. Resnick, J. Levine & S. Teasley (Eds.), *Socially Shared Cognition* (pp. 127-149). Washington, DC: American Psychological Association.

Clark, H. H., & Wilkes-Gibbs, D. (1990). Referring as a collaborative process. In Cohen, Morgan & Pollack (Eds.), *Intentions in Communication* (pp. 463-493). Cambridge, MA: MIT Press.

Cohen, P. R., & Levesque, H. J. (1994). Preliminaries to a collaborative model of dialogue. *Speech Communication, 15*, 265-274.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience, 3*(3), 201-215.

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions: Biological Sciences, 353*, 1245-1255.

Dingus, T. A., Antin, J. F., Hulse, M. C., & Wierwille, W. W. (1989). Attention demand requirements of an automobile moving-map navigation system. *Transportation Research, 23A*(4), 301-315.

Dingus, T. A., Hardee, H. L., & Wierwille, W. W. (1987). Development of models for on-board detection of driver impairment. *Accident Analysis & Prevention, 19*(4), 271-283.

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., et al. (2006). *The 100-Car Naturalistic Driving Study, Phase II – Results of the 100-Car Field Experiment* (No. DOT HS 810 593). Blacksburg, Virginia: Virginia Tech Transportation Institute.

Donmez, B., Boyle, L., & Lee, J. D. (2008a). Accounting for the covariate effects in driving simulator studies. *Theoretical Issues in Ergonomics Science, 9*(3), 189-199.

Donmez, B., Boyle, L. N., & Lee, J. D. (2006). Drivers' attitudes toward imperfect distraction mitigation strategies. *Transportation Research Part F: Traffic Psychology, 9*(6), 387-398.

Donmez, B., Boyle, L. N., & Lee, J. D. (2008b). Designing feedback to mitigate distraction. In M. A. Regan, J. D. Lee & K. L. Young (Eds.), *Driver Distraction: Theory, Effects, and Mitigation* (pp. 519-531): CRC Press.

Donmez, B., Boyle, L. N., & Lee, J. D. (2008c). Mitigating driver distraction with retrospective and concurrent feedback *Accident Analysis & Prevention, 40*, 776–786

Donmez, B., Boyle, L. N., Lee, J. D., & Scott, G. (2007). Safety implications of providing real-time feedback to distracted drivers. *Accident Analysis and Prevention, 39*(3), 581-590.

Emberger, M. (1993). The contribution of PROMETHEUS to European traffic safety. In A. M. Parkes & S. Franzen (Eds.), *Driving Future Vehicles*. Bristol, PA: Taylor & Francis Ltd.

Engström, J., Arfwidsson, J., Amditis, A., Andreone, L., Bengler, K., Cacciabue, P. C., et al. (2006). *Towards the automotive HMI of the future: Mid-term results of the AIDE project*. Berlin: Springer.

Engstrom, J., Johansson, E., & Ostlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F-Traffic Psychology and Behaviour, 8*(2), 97-120.

Engström, J., & Mårdh, S. (2007). *SafeTE final report*: Vägverket.

Evans, L. (2004). *Traffic Safety*. Bloomfield Hills, Michigan: Science Serving Society.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861-874.

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing*. Oxford: Oxford University Press.

Fisher, D. L., Pollatsek, A. P., & Pradhan, A. (2006). Can novice drivers be trained to scan for information that will reduce their likelihood of a crash? *Injury Prevention, 12*, 25-29.

Fletcher, L. (2007). *An automated co-driver for advanced driver assistance systems: the next step in road safety.* Unpublished PhD Thesis, Australian National University.

Fletcher, L., & Zelinsky, A. (2009). Driver Inattention Detection based on Eye Gaze-Road Event Correlation. *International Journal of Robotics Research, 28*(6), 774-801.

Gao, J., & Lee, J. D. (2006). Extending the decision field theory to model operators' reliance on automation in supervisory control situations. *IEEE Systems, Man, and Cybernetics, 36*(5), 943-959.

Genno, H., Ishikawa, K., Kanbara, O., Kikumoto, M., Fujiwara, Y., Suzuki, R., et al. (1997). Using facial skin temperature to objectively evaluate sensations. *International Journal of Industrial Ergonomics, 19*, 161-171.

Gerhardt, D. (1993). Advanced transport telematics systems for the European road infrastructure and human reliablity: Contributions from the DRIVE I programme. In A. M. Parkes & S. Franzen (Eds.), *Driving Future Vehicles* (pp. 7-16). Bristol, PA: Taylor & Francis Ltd.

Gibson, J. J., & Crooks, L. E. (1938). A theoretical field-analysis of automobile driving. *American Journal of Psychology, 51*, 453-471.

Goodrich, M. A., Stirling, W. C., & Boer, E. R. (2000). Satisficing revisited. *Minds and Machines, 10*(1), 79-110.

Goodwin, C. (1986). Between and within: Alternative sequential treatments of continuers and assessments. *Human Studies, 9*, 205-217.

Green, M. (2000). "How long does it take to stop?" Methodological analysis of driver perception-response times. *Transportation Human Factors, 2*(3), 195-216.

Green, P. (2002). Where do drivers look while driving (and for how long?). In R. E. Dewar & P. L. Olson (Eds.), *Human Factors in Traffic Safety* (pp. 77- 110). Tucson, AZ: Lawyers & Judges.

Green, P. E., Wada, T., Oberholzer, J., P.A., G., Schweitzer, J., & Eoh, H. J. (2008). A Final Report of SAfety VEhicles using adaptive Interface Technology (Phase II: Task 3C): Driving Performance

(How Do Distracted and Normal Driving Differ: An Analysis of the ACAS Naturalistic Driving Data). Retrieved March, 2009

Green, P., Levison, W., Paelke, G., & Serafin, C. (1994). *Preliminary Human Factors Guidelines for Driver Information Systems* (Technical Report UMTRI-93-21), FHWA-RD-94-087, McLean, VA: Federal Highway Administration.

Greenberg, J., Tijerina, L., Curry, R., Artz, B., Cathey, L., Grant, P., et al. (2003, January 12-16). *Evaluation of driver distraction using an event detection paradigm.* Paper presented at the Transportation Research Board 82nd Annual Meeting, Washington, D.C.

Hall, M. A., & Holmes, G. (2003). Benchmarking Attribute Selection Techniques for Discrete Class Data Mining. *IEEE Transactions on Knowledge and Data Engineering, 15*(6), 1437-1447.

Hansen, P. (2006). Driver Distraction offers huge safety opportunity: Volvo and Toyota lead the way. *The Hansen report on Automotive Electronics, 19*(7).

Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transaction on Intelligent Transportation Systems, 6*(2), 156-166.

Hoc, J. H. (2001). Towards a cognitive approach to human-machine cooperation in dynamic situations. *International Journal of Human-Computer Studies, 54*(4), 509-540.

Horrey, W. J., Lesch, M. F., & Garabet, A. (2008). Assessing the awareness of performance decrements in distracted drivers. *Accident Analysis & Prevention, 40*(2), 675-682.

Horrey, W. J., & Simons, D. J. (2007). Examining cognitive interference and adaptive safety behaviours in tactical vehicle control. *Ergonomics, 50*(8), 1340-1350.

Horrey, W. J., & Wickens, C. D. (2004). Driving and side task performance: The effects of display clutter, separation, and modality. *Human Factors, 46*(4), 611-624.

Horrey, W. J., & Wickens, C. D. (2006). Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human Factors, 48*(1), 196-205.

Horrey, W. J., Wickens, C. D., & Consalus, K. P. (2006). Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. *Journal of Experimental Psychology-Applied, 12*(2), 67-78.

IIHS. (2010). Texting Laws and Collision Claim Frequencies. Highway Loss Data Institute Bulletin, 27(11).

Itti, L., & Baldi, P. (2005). The salience map: A local surprise detector? *Journal of Cognitive Neuroscience*, 5-5.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2*(3), 194-203.

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

Jahn, G., Oehme, A., Krems, J. F., & Gelau, C. (2005). Peripheral detection as a workload measure in driving: Effects of traffic complexity and route guidance system use in a driving study. *Transportation Research Part F-Traffic Psychology and Behaviour, 8*(3), 255-275.

JAMA (Japan Automobile Manufacturers Association) *Guideline for In-Vehicle Display Systems*, version 2.1. February 2000.

James, W. (1890). *The principles of psychology*. New York: Holt.

Jamson, A. H., & Merat, N. (2005). Surrogate in-vehicle information systems and driver behaviour: Effects of visual and cognitive load in simulated rural driving. *Transportation Research Part F-Traffic Psychology and Behaviour, 8*(2), 79-96.

Jamson, S., Batley, R., Portouli, V., Papakostopoulos, V., Tapani, A., Lundgren, J., et al. (2004). Obtaining the functions describing the relations between behaviour and risk. Retrieved March, 2009

Johansson, E., Engström, J., Cherri, C., Nodari, E., Toffetti, A., & Schindhelm, R. (2004). *Review of existing techniques and metrics for IVIS and ADAS assessment*. Paper presented at the IEEE Transactions on Intelligent Transportation Systems.

Jurafsky, D., Bates, R., Coccaro, N., Martin, R., Meteer, M., Ries, K., et al. (1997). Automatic detection of discourse structure for speech recognition and understanding *Proceedings IEEE Workshop on Speech Recognition and Understanding* (pp. 88–95). Santa Barbara, CA.

Kaplin, S., & Garrick, G. J. (1981). On the quantitative definition of risk. Risk Analysis, 1(1), 11-27.

Karlsson, R. (2004). *Evaluating driver distraction countermeasures.* Linköping University, Sweden.

Kincses, W. (2006). Building augmented cognition systems for automotive applications. In C. Forsythe, M. L. Bernard, & T.E. Goldsmith (Eds.), Cognitive Systems: Human Cognitive Models in Systems Design pp. 237-255. Mahwah, NJ: Lawrence Erlbaum Associates.

Kircher, K., Ahlstrom, C., & Kircher, A. (2009). *Comparison of two eye-gaze based real-time driver distraction detection algorithms in a small-scale field operational test*. Paper presented at the Driving Assessment 2009: 5th International Driving Symposium on Human Factors in Driving Assessment, Training and Vehicle Design, Big Sky, MT.

Kircher, K., Kircher, A., & Ahlstrom, C. (2009). Results of a field study on a driver distraction warning system. Linköping, Sweden: VTI, Swedish National Road and Transport Research Institute.

Kircher, K., Kircher, A., & Claezon, F. (2009). Distraction and drowsiness- a field study. Linköping, Sweden: VTI, Swedish National Road and Transport Research Institute.

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). *The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data* (Report No. DOT HS 810 594). Washington, DC: National Highway Traffic Safety Administration.

Kruse, A. A., & Schmorrow, D. D. (2005). *Session overview: Foundations of Augmented Cognition*. Paper presented at the Presented at the 1st Annual Augmented Cognition International Conference.

Lamble, D., Kauranen, T., Laakso, M., & Summala, H. (1999). Cognitive load and detection thresholds in car following situations: safety implications for using mobile (cellular) telephones while driving. *Accident Analysis and Prevention, 31*(6), 617-623.

Lamble, D., Laakso, M., & Summala, H. (1999). Detection thresholds in car following situations and peripheral vision: implications for positioning of visually demanding in- car displays. *Ergonomics, 42*(6), 807-815.

Land, M. F., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society of London, Series B* 352: 1231-1239.

Lee, J.D., et al. (2010, August). *Assessing the Feasibility of Vehicle-Based Sensors to Detect Alcohol Impairment*. (Report No. DOT HS 811 358). National Highway Traffic Safety Administration: Washington DC. p. 98. Available at http://www.nhtsa.gov/DOT/NHTSA/NVS/Crash%20Avoidance/Technical%20Publications/2010/811358.pdf.

Lee, J. D. (2010). Calibrating demand and capacity in multitask environments: Interactions of micro and macro cognition. In E. S. Patterson & J. Miller (Eds.), *Macrocognition Metrics and Scenarios: Design and Evaluation for Real-World Teams*. New York: Ashgate.

Lee, Y.-C., Lee, J. D., & Ng Boyle, L. (2007). Visual Attention in Driving: The Effects of Cognitive Load and Visual Disruption. Human Factors: The Journal of the Human Factors and Ergonomics Society, 49(4), 721-733.Lee, J. D., McGehee, D. V., Brown, T. L., & Reyes, M. L. (2002a). Collision warning timing, driver distraction, and driver response to imminent rear-end collisions in a high-fidelity driving simulator. *Human Factors, 44*(2), 314-334.

Lee, J. D., McGehee, D. V., Brown, T. L., & Reyes, M. L. (2002b, March). *Driver Distraction, Warning Algorithm Parameters, and Driver Response to Imminent Rear-end Collisions in a High-Fidelity Driving Simulator*. (Report No. DOT HS 809 488). Washington, DC: National Highway Traffic Safety Administration. Available at http://www.nhtsa.gov/DOT/NHTSA/NRD/Multimedia/PDFs/Human%20Factors/Driver%20Assistance/DOT%20HS%20809%20448.pdf.

Lee, J. D., Regan, M. A., & Young, K. L. (2008a). Defining driver distraction. In M. A. Regan, J. D. Lee & K. L. Young (Eds.), *Driver Distraction: Theory, Effects, and Mitigation* (pp. 31-40). Boca Raton, FL: CRC Press.

Lee, J. D., Regan, M. A., & Young, K. L. (2008b). What drives distraction? Distraction as a breakdown of multilevel control. In M. A. Regan, J. D. Lee & K. L. Young (Eds.), *Driver Distraction: Theory, Effects, and Mitigation* (pp. 41-56). Boca Raton, FL: CRC Press.

Lee, J. D., & Strayer, D. L. (2004). Preface to a special section on driver distraction. *Human Factors, 46*, 583-586.

Lee, Y. C., Lee, J. D., & Boyle, L. N. (2007). Visual attention in driving: The effects of cognitive load and visual disruption. *Human Factors*.

Lenneman, J. K., & Backs, R. W. (2009). Cardiac autonomic control during simulated driving with a concurrent verbal working memory task. *Human Factors, 51* (3), 404-418.

Lenneman, J. K., Shelly, J. R., & Backs, R. W. (2005). *Deciphering psychological-physiological mappings while driving and performing a secondary memory task*. Paper presented at the Proceedings of the Third International Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.

Levy, J., Pashler, H. E., & Boer, E. R. (2006). Is there any stopping the psychological refractory period? *Psychological Science, 17*(3), 228-234.

Liang, Y. (2009). *Detecting driver distraction.* University of Iowa, Iowa City.

Liang, Y., & Lee, J. D. (in press). Combining cognitive and visual distraction: Less than the sum of its parts. *Accident Analysis and Prevention*.

Liang, Y., Reyes, M. L., & Lee, J. D. (2007a). Non-intrusive detection of driver cognitive distraction in real-time using Bayesian networks. *Transportation Research Board*(2018), 1-8.

Liang, Y., Reyes, M. L., & Lee, J. D. (2007b). Real-time detection of driver cognitive distraction using Support Vector Machines. *IEEE Intelligent Transportation Systems, 8*(2), 340-350.

Mack, A., Tang, B., Tuma, R., Kahn, S., & Rock, I. (1992). Perceptual Organization and Attention. *Cognitive Psychology, 24*(4), 475-501.

Markkula, G., & Engström, J. (2006). A steering wheel reversal rate metric for assessing effects of visual and cognitive secondary task load. *Proceedings of the ITS World Congress, London*.

Marple-Horvat, D. E., Chattington, M., Anglesea, M., Ashford, D. G., Wilson, M., & Keil, D. (2005). Prevention of coordinated eye movements and steering impairs driving performance. *Experimental Brain Research, 163*, 411–420.

Marple-Horvat, D. E., Cooper, H. L., Gilbey, S. L., Watson, J. C., Mehta, N., Kaur-Mann, D., et al. (2008). Alcohol badly affects eye movements linked to steering, providing for automatic in-car detection of drink driving. *Neuropsychopharmacology, 33*, 849 – 858.

Mattes, S., & Hallén, A. (2009). Surrogate distraction measurement techniques: The lane change test. In M. A. Regan, J. D. Lee & K. L. Young (Eds.), *Driver Distraction: Theory, Effects and Mitigation*. Boca Raton, Florida: CRC Press.

McCartt, A. T., Hellinga, L. A., Strouse, L. M., & Farmer, C. M. (2010). Long-term effects of handheld cell phone laws on driver handheld cell phone use. Traffic Injury Prevention, 11(2), 133-141. doi: 10.1080/15389580903515427

McEvoy, S. P., Stevenson, M. R., & Woodward, M. (2006). The impact of driver distraction on road safety: results from a representative survey in two Australian states. *Injury Prevention, 12*(4), 242-247.

McGehee, D. V., Raby, M., Carney, C. H., Reyes, M. L., & Lee, J. D. (2007). Extending parental mentoring using an event-triggered video intervention in rural teen drivers. *Journal of Safety Research, 38*, 15-227.

Michon, J. A. (1985). A critical view of driver behavior models: What do we know, what should we do? In L. Evans & R. C. Schwing (Eds.), *Human Behavior and Traffic Safety* (pp. 485-520). New York: Plenum Press.

Misener, J. A., Thorpe, C., Ferlis, R., Hearne, R., Siegal, M., Perkowski, J., & Trb. (2001). Sensor-friendly vehicle and roadway cooperative safety systems - Benefits estimation Highway Safety: Modeling, Analysis, Management, Statistical Methods, and Crash Location - Safety and Human Performance (pp. 22-29).

Nakayama, O., Futami, T., Nakamura, T., & Boer, E. R. (1999). SAE Technical Paper Series: Development of a steering entropy method for evaluating driver workload. *Human Factors in Audio Interior Systems, Driving, and Vehicle Seating, SP-1426*.

Najm, W. G. (2003). Alternative Methods for Safety Benefits Estimation. Cambridge, MA: Volpe National Transportation Systems Center

Najm, W. G., Stearns, M. D., Howarth, H., Koopmann, J., & Hitz, J. (2005). Evaluation of an Automotive Rear-End Collision Avoidance System. Washington, DC: Research and Innovative Technology Administration, Advanced Safety Technology Division.

NHTSA. (2010, September). Distracted driving 2009 Traffic safety facts. DOT HS 811 379. Washington, DC: Author. Available at http://www.distraction.gov/research/PDF-Files/Distracted-Driving-2009.pdf.

Opperud, A., Jeftic, Z., & Jarlengrip, J. (2005, November, 14-16). *Real-time adaptation of advanced driver assistance system to driver state.* Paper presented at the Proceedings of the 2005 International Truck and Bus Safety and Security Symposiom, Washington, Alexandria, Virginia.

Östlund, J., Nillson, L., Torson, J., & Forsman, A. (2006). *Effects of cognitive and visual load in real and simulated driving*. Linköping, Sweden: VTI.

Östlund, J., L. Nilsson, O. Carsten, N. Merat, H. Jamson, S. Jamson, S. Mouta, J. Carvalhais, J. Santos, V. Anttila, H. Sandberg, J. Luoma, D. de Waard, K. Brookhuis, E. Johansson, J. Engström, T. W. Victor, J. Harbluk, W. Janssen and R. Brouwer. (2004). "HASTE Deliverable 2. HMI and Safety-Related Driver Performance."

Östlund, J., B. Peters, B. Thorslund, J. Engström, G. Markkula, A. Keinath, D. Horst, S. Juch, S. Mattes and U. Foehl. (2005). "Deliverable 2.2.5 - Driving performance assessment - methods and metrics." No. IST-1-507674-IP. from http://www.aide-eu.org/pdf/sp2_deliv_new/aide_d2_2_5.pdf.

Patten, C. J. D., Kircher, A., Ostlund, J., & Nilsson, L. (2004). Using mobile telephones: cognitive workload and attention resource allocation. *Accident Analysis and Prevention, 36*(3), 341-350.

Pickrell, T. M., & Ye, T. J. (2009, September). *Driver Electronic Device Use in 2008*. (Report No. DOT HS 811 184). Washington, DC: National Highway Traffic Safety Administration. Available at http://www-nrd.nhtsa.dot.gov/pubs/811184.pdf.

Pradhan, A. K., Hammel, K. R., DeRamus, R., Pollatsek, A., Noyce, D. A., & Fisher, D. L. (2005). Using eye movements to evaluate effects of driver age on risk perception in a driving simulator. *Human Factors, 47*(4), 840-852.

Rakauskas, M. E., & Ward, N. J. (2007). Driving performance during 511 Information Retrieval: Cell Phone 2 (p. 93). St. Paul, MN: Minnesota Department of Transportation.

Rakauskas, M. E., Ward, N. J., Boer, E., Bernat, E. M., Cadwallader, M., & Patrick, C. J. (2008). Combined effects of alcohol and distraction on driving performance. *Accident Analysis & Prevention, 40*, 1742–1749.

Ranney, T. A. (1994). Models of driving behavior: A review of their evolution. *Accident Analysis and Prevention, 26*(6), 733-750.

Recarte, M. A., & Nunes, L. M. (2003). Mental workload while driving: Effects on visual search, discrimination, and decision making. *Journal of Experimental Psychology-Applied, 9*(2), 119-137.

Recarte, M. A., Perez, E., Conchillo, A., & Nunes, L. M. (2008). Mental workload and visual impairment: differences between pupil, blink, and subjective rating. *The Spanish Journal of Psychology, 11*(2), 374-385.

Regan, M. A., Lee, J. D., & Young, K. L. (2008). *Driver Distraction: Theory, Effects and Mitigation*. Boca Raton, Florida: CRC Press.

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology, 53*, 245-277.

Reyes, M. L., & Lee, J. D. (2008). Effects of cognitive load presence and duration on driver eye movements and event detection performance. *Transportation Research Part F: Traffic Psychology*, 391-402.

Rockwell, T. H. (1988). Spare visual capacity in driving - revisited. In A. G. Gale, M. H. Freeman, C. M. Hasleman, P. Smith & S. P. Taylor (Eds.), *Vision in Vehicles II*. North Holland: Elsevier Science Publishers.

Rouse, W. B. (1988). Adaptive aiding for human/computer control. *Human Factors, 30*, 431-443.

Rumar, K. (1990). Driver Requirements and Road Traffic Informatics. *Transportation, 17*(3), 215-229.

Salvucci, D. D. (2007). iPod distraction: Effects of portable music-player use on driver performance. *CHI 2007 Proceedings*, 243-250.

Salvucci, D. D., Boer, E. R., & Liu, A. (2001). Toward an integrated model of driver behavior in a cognitive architecture. *Transportation Research Record*.

Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of 'Uh Huh' and other things that come between sentences. In D. Tannen (Ed.), *Georgetown University Roundtable on Languages and Linguistics*. Washington, D.C: Georgetown University Press.

Schmorrow, D. D., & McBride, D. (2004). Special issue: Augmented cognition - Introduction. *International Journal of Human-Computer Interaction, 17*(2), 127-130.

Schmorrow, D. D., & Stanley, K. M. (2008). *Augmented Cognition: A Practitioner's Guide.* Santa Monica, CA: Human Factors and Ergonomics Society.

Scott, J. J., & Gray, R. (2008). A comparison of tactile, visual, and auditory warnings for rear-end collision prevention in simulated driving. Human Factors, 50(2), 264-275.

Senders, J. W., Kristofferson, A. B., Levison, W. H., Dietrich, C. W., & Ward, J. L. (1967). The attentional demand of automobile driving. *Highway Research Record, 195*, 15-33.

Sharon, T., Selker, T., Wagner, L., & Frank, A. J. (2005). CarCoach: a generalized layered architecture for educational car systems. *Proceedings of IEEE International Conference on Software - Science, Technology and Engineering* (pp. 13-22).

Sheridan, T. B. (1970). Big brother as driver: New demands and problems for the man at the wheel. *Human Factors, 12*(1), 95-101.

Sheridan, T. B. (2004). Driver distraction from a control theoretic perspective. *Human Factors, 46*(4), 587-599.

St. John, M., Kobus, D. A., Morrison, J. G., & Schmorrow, D. D. (2004). Overview of the DARPA augmented cognition technical integration experiment. *Journal International Journal of Human-Computer Interaction, 17*(2), 131-149.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137-149.

Strayer, D. L., Drews, F. A., & Johnston, W. A. (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of Experimental Psychology-Applied, 9*(1), 23-32.

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

Stevens, A., Quimby, A., Board, A., Kersloot, T., & Burns, P. (2004). *Design Guidelines for Safety of In-Vehicle Information Systems.* TRL Limited.

Taylor, S., & Todd, P. A. (1995). Understanding Information Technology Usage - a Test of Competing Models. Information Systems Research, 6(2), 144-176.

Tijerina, L., Parmer, E. B., & Goodman, M. J. (1999). *Individual differences and in-vehicle distraction while driving: A test track study and psychometric evaluation.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting.

Utter, D. (2001). *Research Note: Passenger Vehicle Driver Cellular Phone Use*. (Report No. DOT HS 809 293). Washington, DC: National Highway Traffic Safety Administration. Available at http://www-nrd.nhtsa.dot.gov/Pubs/809-293.PDF.

Veltman, J. A., & Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics, 41*(5), 656-669.

Victor, T. W. (1999). *Driver support by recognition of visual behavior, VISREC project description, Program Board for Swedish Automotive Research,* .

Victor, T. W. (2005). *System and method for monitoring and managing driver attention loads. United States Patent No.6974414.*

Victor, T. W. (2006). *Keeping Eye and Mind on the Road.* Uppsala University, Uppsala, Sweden.

Victor, T. (2010). *The Victor and Larsson (2010) Distraction Detection Algorithm and Warning Strategy*. Volvo Technology.

Victor, T. W., Harbluk, J. L., & Engstrom, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F, 8*, 167-190.

Victor, T. W., & Larsson, P. (2004). International application published under the patent cooperation treaty: I. B. W. I. P. organization.

von Hofsten, C. (1995). Planning and perceiving what is going to happen next. In M. Haith, J. Benson, R. Roberts & B. Pennington (Eds.), *The development of future oriented processes*. Chicago: University of Chicago Press.

Wang, J. A., Knipling, R. R., & Blincoe, L. J. (1999). The dimensions of motor vehicle crash risk. Journal of Transportation and Statistics, 2(1), 19-44.

White, M. P., Eiser, J. R., & Harris, P. R. (2004). Risk perceptions of mobile phone use while driving. *Risk Analysis, 24*(2), 323 - 334.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*(2), 159-177.

Wierwille, W. W. (1979). Physiological measures of aircrew mental workload. *Human Factors, 21*(5), 575-593.

Wierwille, W. W. (1993). Visual and manual demands of in-car controls and displays. In B. Peacock & W. Karwowski (Eds.), *Automotive Ergonomics* (pp. 299-320). London: Taylor and Francis.

Wierwille, W. W., & Tijerina, L. (1998). Modeling the relationship between driver in-vehicle visual demands and accident occurrence. In A. G. Gale, I. Brown & C. Haselgrave (Eds.), *Vision in vehicles* (pp. 233-243). Amsterdam: Elsevier.

Wierwille, W. W., Tijerina, L., Kiger, S., Rockwell, T., Lauber, E., & Bittner, A. (1996, October). *Heavy Vehicle Driver Workload Assessment -- Task 4: Review of Workload and Related Research. Final Report Supplement.* (Report No. DOT HS 808 467 [4]). Washington, DC: National Highway Traffic Safety Administration.

Wikman, A. S., Nieminen, T., & Summala, H. (1998). Driving experience and time-sharing during in-car tasks on roads of different width. *Ergonomics, 41*(3), 358-372.

Wilson, G. F., & Russell, C. A. (2003). Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. *Human Factors, 45*(4), 635-643.

Wilson, M., Stephenson, S., Chattington, M., & Marple-Horvat, D. E. (2007). Eye movements coordinated with steering benefit performance even when vision is denied. *Experimental Brain Research, 176*, 397–412.

Wilson, M., M. Chattington and D. E. Marple-Horvat (2008). "Eye movements drive steering: reduced eye movement distribution impairs steering and driving performance." Journal of Motor Behavior 40(3): 190–202.

Witt, G. J. (2003). SAfety VEhicle(s) using adaptive Interface Technology (SAVE-IT) Program: A Proposal Submitted to U.S. Department of Transportation RSPA/Volpe National Transportation Systems Center.

Yekhshatyan, L. (2010) Detecting distraction and degraded driver performance with visual behavior metrics. The University of Iowa, Iowa City.

Zhang, H., Smith, M. R. H., & Witt, G. J. (2006). Identification of real-time diagnostic measures of visual distraction with an automatic eye-tracking system *Human Factors, 48*(4), 805-821.

Zhang, H., & Smith., M. (2004). A Final Report of SAfety Vehicles Using Adaptive Interface Technology (Phase I: Task 7): Visual Distraction Research. Retrieved January, 2005

Zhang, Y., Owechko, Y., & Zhang., J. (2004). Driver cognitive workload estimation: a data-driven perspective *IEEE Intelligent Transportation Systems Conference*. Washington, D.C.: IEEE.

Zylstra, B., Tsimhoni, O., Green, P., & Mayer, K. (2004, December 2004). SAfety VEhicles using adaptive Interface Technology (Task 3B): Driving performance for dialing, radio tuning, and destination entry while driving straight roads.   Retrieved March, 2009

DOT HS 811 547A
May 2013

U.S. Department
of Transportation

National Highway
Traffic Safety
Administration

NHTSA
www.nhtsa.gov

8339A-043013-v4