Integrating the Saliency Map with Distract-R to Assess Driver Distraction of Vehicle Displays

By

Joonbum Lee

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Industrial and Systems Engineering)

at the

UNIVERSITY OF WISCONSIN-MADISON

2014

Date of final oral examination: 1/22/2014

The dissertation is approved by the following members of the Final Oral Committee:

John D. Lee (Chair), Professor, Industrial and Systems Engineering

Gregg C. Vanderheiden, Professor, Industrial and Systems Engineering

Douglas A. Wiegmann, Associate Professor, Industrial and Systems Engineering

David A. Noyce, Professor, Civil and Environmental Engineering

Bilge Mutlu, Assistant Professor, Computer Sciences

To my parents, Sungjae Lee and Boksook Lee, for their love and sacrifice

Abstract

There are a growing number of potential distractions in vehicles today, such as navigation, collision warning, and entertainment systems. These systems promise substantial benefits for driving comfort, efficiency and safety, but they might also distract drivers. This dissertation develops computational cognitive models of driver behavior to assess the distraction potential of vehicle displays.

One of the main goals of this dissertation is to integrate a saliency-based model, a saliency map, into Distract-R to build a computational model that can account for top-down and bottom-up attentional influences. The saliency-based model quantifies exogenous influences (e.g., visual features of a display) of visual attention while Distract-R quantifies endogenous influences (e.g., drivers' goals and expectations) of visual attention with respect to secondary tasks and vehicle displays.

Two experiments were conducted to guide model development and to validate model predications. The experiments showed that design features of vehicle displays affected driving performance and glance duration to the secondary task, and both top-down and bottom-up attentional processes were engaged when drivers interacted with driver-vehicle interfaces.

To integrate Distract-R and the saliency map, activation fields that describe the interaction between top-down and bottom-up attentional process were used to determine glance duration to the display. The integrated model was validated with empirical data, showing that the model could predict drivers' pattern of glance durations to a level comparable to between-subject variability—the theoretical limit of prediction. This dissertation contributes to modeling driver distraction by integrating two models to account both top-down and bottom-up influence on visual attention, and by building a tool for assessing the potential distraction of vehicle displays.

Acknowledgements

There were so many people who helped me. I would like to acknowledge to three of them:

- To John Lee – John guided me as a researcher, and showed me how to approach to research problems. We had countless meetings and conversations, and he always inspired and challenged me. This dissertation would not be completed without him. I was blessed to meet him.
- To Dario Salvucci – This dissertation augmented Dario's Distract-R system, and he helped me to develop the integrated model in many ways. I really enjoyed collaborating with him.
- To Jaesik Lee – He invited me to join his lab as an undergraduate intern, and that opportunity made me to have a bigger dream. I always appreciate him.

Table of Contents

List of Tables

List of Figures

1. Introduction

1.1. Problem Statement and Research Overview

Driving is a complex task that requires scanning the environment to maintain control and identify hazards. In terms of drivers' visual attention, the driving environment is cluttered with information from inside the vehicle (e.g., interacting with in-vehicle devices) and outside the vehicle (e.g., advertisement panels or signs). This information has the potential to distract. In 2011, over 3,000 people were killed, and 421,000 people were injured in crashes involved driver distraction (National Highway Traffic Safety Administration, 2013).

Regan, Young, and Lee (2009) defined driver distraction as "the diversion of attention away from activities critical for safe driving toward a competing activity" (p. 7). Distraction represents a substantial driving safety concern. Visual distraction, in which a driver looks away from the forward roadway, represents a particularly important threat.

Sources of distraction are various. The sources of distraction can be classified into six broad categories—"things brought into vehicle," "vehicle systems", "vehicle occupants," "moving object or animal in vehicle," "internalized activity," "external objects, events of activities," and "other source of distraction" (Regan, Young, Lee, & Gordon, 2009, p. 253). Among these sources, this dissertation focuses on interaction with the "vehicle systems" such as displays built into the vehicle.

These vehicle systems represent a growing number of potential distractions, including in-vehicle navigation, collision warning, traveler information, and entertainment systems. These systems promise substantial benefits for driving comfort, efficiency and safety, but they might also distract drivers. Regan and colleagues (2009) also mentioned that "the rapid proliferation within vehicle cockpits of factory-fitted entertainment systems (e.g., DVD players), vehicle information and communication systems (VICS; e.g., traveler information, Internet), and advanced driver-assistance systems (ADVS; e.g., adaptive cruise control, in-vehicle navigation, collision warning) may create new sources of distraction if the systems, and the functions they support, are poorly designed and located, or used inappropriately" (p. 276). That is, as more in-vehicle systems are developed, the greater potential for distraction.

There are a variety of methods to assess driver distraction. Lee, Lee, & Salvucci (2012) identified three general methods to assess driver distraction: guidelines, collecting behavioral data, and computational modeling. Guidelines define distraction and specify how designers and manufacturers can develop less distracting devices (e.g., Alliance of Automobile Manufacturers, 2003). Some of these guidelines suggest very specific interface features. The Federal Highway Administration (FHWA)'s Human Factors Guidelines state that highly saturated blue should be avoided (Campbell, Carney, & Kantowitz, 1998). Additionally, the Adaptive Integrated Driver-Vehicle Interface (AIDE) report states that the use of too many colors should be avoided and maximum of five easy-to-distinguish colors is recommended (Schindhelm et al., 2004). Such guidelines have the benefit of providing direct and precise answers to specific design questions. However, guidelines often take a reductionist approach that may not capture problems that emerge from an interaction of features, and may offer little to no guidance for new systems that deviate in significant ways from the original concepts for which the guidelines were developed.

A second method to understand and alleviate distraction involves evaluating designs by collecting behavioral data from drivers. Data collection can be performed in controlled laboratory experiments (e.g., Horberry, Anderson, Regan, Triggs, & Brown, 2006; Horrey & Lesch, 2009) and in on-road field studies (e.g., Harbluk, Noy, Trbovich, & Eizenman, 2007; Stutts et al., 2005), and may use a variety of test environments, ranging from desktop simulators to large-scale simulators (e.g., Strayer & Drews, 2004) to instrumented vehicles (e.g., Horrey & Lesch, 2009). Behavioral driver data provide a holistic description of the phenomena being tested, and so may offer a more comprehensive assessment of design features than is possible with guidelines. However, there are typically tradeoffs between the realism of the behaviors observed and the costs of collecting the data. For instance, a large-scale field study may provide the most comprehensive data, but may be prohibitively expensive and time-consuming (not to mention exposing drivers to unacceptable safety risks), whereas a small-scale laboratory study can be done quickly, but the resulting data may be less representative of actual driving situations. Especially for novel systems, a rigorous behavioral study often cannot include all the possible design alternatives that need to be addressed.

Computational models of driver behavior present a third way to evaluate the distraction potential of system designs. This approach relies on the development of cognitive models that represent the cognitive processes and physical actions of a driver in various simulated situations. For example, the ACT-R driver model (Salvucci, 2006) was developed to simulate driver behavior in a highway-driving context, and was later extended to account for aspects of distraction from phone dialing (e.g., Salvucci & Taatgen, 2008) and cognitive tasks such as memory rehearsal (e.g., Salvucci & Beltowska, 2008). Cognitive models have the benefit that a designer can perform systematic tradeoffs among a large variety of design features to predict distraction much more quickly and safely than with tests involving human drivers.

Designers may have an intuition for the potential distraction of interfaces that they designed. This intuition might reflect the qualitative influence of color, contrast, or layout in a design. Computational models can confirm these intuitions, but can also indicate the precise quantitative effect of the design on distraction. In particular, repeated application of the models (e.g., the Monte-Carlo simulation) can provide indication of the relative effect of each design parameter, their interactions, and the likely variability associated with human performance with the design. In this way, computational models can confirm and extend designers' intuitions. However, the development of such models is time-consuming, knowledge-intensive, and ultimately, limited to our understanding of human behavior as can be expressed in a computational framework.

This dissertation focuses on the third approach to assess distraction potential of driver-vehicle interface (DVI) system—using computational cognitive models of driver behavior. Previous models used for assessing driver behavior were reviewed to identify requirements for building a model to evaluate distraction potential. Among the previous models, ACT-R, a cognitive architecture, has the potential to satisfy the requirements. Furthermore, one application of the ACT-R model, Distract-R (Salvucci, 2009)—a rapid prototyping and evaluation tool—has demonstrated substantial promise as a tool to test the distraction potential of a DVI. However, ACT-R and Distract-R have a limited ability to represent visual attention and design features that might lead to long glances away from the road. Moreover, there is a significant gap between drivers' glance patterns and a model of driver's glance pattern. For example, the model might always complete a secondary task with a single glance, whereas drivers might actually complete the task in a series of cycles of glancing to a secondary task and glancing back to the roadway. This dissertation aims to narrow the gap between predictions made by existing cognitive models of driver behavior and actual driver behavior.

The first two chapters of this dissertation describe how current models of visual attention and salience can be used to extend the capabilities of Distract-R as a tool to address the design challenges of DVIs. The extension will maintain Distract-R's ability to account for low-level visual features, while transforming the tool into an integrated model that can account for top-down and bottom-up attentional processes. The following two chapters investigate driver distraction by collecting empirical data, and augment the integrated model's capability by implementing a function for glance switching.

Visual attention depends on both top-down, goal-driven (i.e., endogenous) and bottom-up, stimulus-driven (i.e., exogenous) processes. Expectation, beliefs, and general knowledge of tasks can influence visual attention (Wright & Ward, 2008), and drivers also may have goal-driven input (e.g., spatial expectation for specific items). Therefore, an integrated model that accounts for both top-down and bottom-up mechanisms could provide a powerful way to describe the distribution of drivers' visual attention and predict potential distraction, because it could account for symbolic representations of drivers' task goals and expectations along with a continuous field of activation associated with the salience of visual stimuli.

One of the main goals of this dissertation is to integrate a saliency-based model, a saliency map, into Distract-R. The saliency-based model quantifies exogenous influences (e.g., visual features of a display) of visual attention and Distract-R quantifies endogenous influences (e.g., drivers' goals and expectations) of visual attention with respect to secondary tasks and driver distraction.

For example, designers may use salient colors for more important information such as warnings and emergency information. This could promote faster search and reaction times, because the salient features attract attention, leading drivers to see the important information. However, if the visual salience of the display is not aligned with information needs of the driver, the driver might need more time to find needed information because the salient features pull drivers' attention to irrelevant information. Such misplaced salience could distract drivers. In other situations, designers might spatially group information by their sources or types, and drivers could develop spatial expectations for specific target information. Appropriate spatial expectations could reduce distraction. The proposed model will integrate Distract-R and the saliency map to examine the effect of coordinating salient visual features and drivers' expectations to assess the distraction potential of vehicle displays. The proposed model will facilitate display design by quantifying the effect of aligning salience and expectations with drivers' tasks in display design, so that attention quickly moves to the most informative parts of the display.

Moreover, the proposed model is expected to mimic how drivers switch between the primary and secondary task. Task switching is a key aspect in multitasking while driving. When the secondary task requires a long-duration glance to the in-vehicle device, drivers use multiple short glances to complete the secondary task. However, Distract-R lacks a function to switch back to the primary task (i.e., driving) while performing the secondary task. Implementing a function to determine maximum glance duration on the secondary task and to pause the secondary task whenever the single glance duration exceeds the time threshold will improve the integrated model's ability to simulate distracted drivers.

1.2. Objectives

This dissertation proposes the integrated model comprised of Distract-R and the saliency map to account for both top-down and bottom-up influences on visual attention, with the ultimate goal of building a robust model that can simulate distracted drivers. The current version of Distract-R does not account for low-level visual features (e.g., colors and contrast) and does not model drivers' task switching to the primary task during the secondary task. Therefore, implementing the saliency map will significantly extend the application of Distract-R by enabling designers to identify display design flaws associated with misplaced salience, and implementing a function to determine timing for switching back to the primary task will enhance the integrated model's ability to predict driver distraction. These broad goals will be achieved through the following specific aims:

1) Identify requirements for the integrated model to evaluate distraction potential (Chapter 2)
2) Validate the saliency map (Chapter 3)
3) Integrate the saliency map into Distract-R (Chapter 4)
4) Investigate driver distraction with empirical data (Chapter 5)
5) Validate the integrated model (Chapter 6)

1.3. Overview

The following chapters describe studies to investigate driver distraction and the development of a computational model to assess distraction potential of vehicle systems. Chapter 2 describes previous efforts to build a computational model for driver distraction, and defines requirements for the model based on previous models and the challenges of applying design guidelines or empirical evaluation. Distract-R system satisfies most of the requirements, but lacks an ability to account for bottom-up visual process. The saliency map has an ability to complement the limitation, so the integration of two models is proposed.

Chapter 3 describes the implementation of the saliency map to predict search time of items on a vehicle display. The saliency map has not been validated for its ability to predict search time, and it does not account variability of human performance. The Monte-Carlo technique was applied to the saliency map augmented with a stochastic process, and the augmented saliency map's predictive capability was compared with empirical data. The result showed that the augmented saliency map could be applied to predict search time to a level comparable to between-subject variability.

Chapter 4 describes the integration of the saliency map with Distract-R. The model integration brings an ability that can assess influence of top-down and bottom-up processes. To integrate two models, conceptual background was required. The activation field that describes how the top-down and bottom-up attentional processes complement each other was used to integrate the saliency map into Distract-R. Integration of two models also required technical modifications for both the saliency map and Distract-R. Details of the modification were described in this chapter.

In Chapter 5, an experiment was conducted to investigate how drivers manage the secondary task while driving, and what leads long glances away from the road. The purpose of the experiment was to assess how the integrated model, equipped with an ability to determine total glance duration to in-vehicle devices based on visual features, predicts search time and glance patterns. The result showed that (1) drivers glanced to the vehicle display multiple times when the secondary task required long completion time, and (2) both design features of displays and vehicle stability affect the duration of glances away from the road.

Based on the result from the Chapter 5, the integrated model was modified to model how drivers switch attention between the road and the secondary task. The integrated model's prediction was compared with the empirical data and the original Distract-R's prediction. The result showed that the integrated model outperformed the original Distract-R. The integrated model showed a similar pattern of total glance time across the experimental conditions. Moreover, the integrated model could break a long-duration glance into multiple glances to complete the secondary task in a many similar to drivers.

1.4. Theoretical and Practical Contributions

Theoretical contributions of this study include (1) the integration of two models to develop a computational model that predicts the top-down and bottom-up influence on visual attention, and (2) the modification of the integrated model to enhance its ability to model drivers' task switching behavior, which is one of the key features in driver distraction. A practical contribution of this study will be the development of a tool for assessing potential distraction of in-vehicle displays by enhancing Distract-R's visual module and the addition of a task switching function.

2. Background

2.1. General Objectives

The main objectives of this chapter are (1) to review existing computational models of driver performance for identifying requirements for the proposed model, and (2) describe the role of bottom-up visual attention as explained by literature concerning attention shift and guided search. The first section of this chapter reviews previous computational models and includes a review of Distract-R, which is a main part of the proposed model. The second section of this chapter reviews theories of visual attention. This section aims to provide a foundation for understanding the saliency map, which is another part of the proposed model. The last section introduces the saliency map and describes how the model can contribute to understanding driver distraction, and how the model can complement Distract-R's visual module.

2.2. Models of Driver Behavior and Performance

Over the past four decades, there have been numerous efforts to represent driver behavior with mathematical (or computational) models. A number of these models focused on control-theoretic representations of behavior (e.g., Brackstone & McDonald, 1999; Donges, 1978; Godthelp, 1986). These models, however, are less applicable to driver distraction because they assume continuous attention to perceptual inputs that may not be directly accessible to the driver, such as speed and distance (Boer, 1999). Other work has focused on the basic perceptual aspects of driving (e.g., Fajen & Warren, 2003; Wilkie & Wann, 2003), but again with a representation that resembles continuous control rather than the intermittent attention common in distraction scenarios. An early exception to this tendency demonstrated that drivers sample the forward roadway more intensely as uncertainty in the road situation increases (Senders, Kristofferson, Levison, Dietrich, & Ward, 1967). Subsequent sections review the efforts that relate most closely to this dissertation, with a focus on predicting distraction from Driver-Vehicle Interfaces (DVIs).

2.2.1. The Driver Performance Model

The Driver Performance Model (DPM) is a computational model of a driver's perception, cognition, and control processes to generate steering, braking, and acceleration inputs (Bittner, Simsek, Levison, & Campbell, 2002). The model is a part of the Interactive Highway Safety Design Model (IHSDM) which is a set of modules (e.g., geometrics, vehicle dynamic module, and driver model) to evaluate highway geometric design (Levison, 1998). The DPM simulates interactions of driver and vehicle with roadway geometry, and provides prediction of the effects of highway geometry on driver behavior (Levison, Simsek, Bittner, & Hunn, 2001). It consists of six computational functions (perception, speed decision, path decision, speed control, path control, and attention) and generates control actions (speed control and path control) in response to the situation (Bittner et al., 2002) (Figure 1).

The perception component receives the physical description of the situation from the input data (e.g., geometric component), and then translates the description to estimates of vehicle state and other relevant variables for the decision and control components (Levison, 1998). The speed/path decision components compute desired speed/path profiles that reflect geometric features, and then the attention component modifies the information-processing components to reflect driver's mental capacity and attentional demand that are controlled by roadway geometry. The speed control module generates accelerator and brake pedal action to maintain the speed that is commanded by the speed decision module, and path control module generates steering wheel movements to track the path that is commanded by the path decision module (Levison, 1998). The DPM can be used to evaluate impact of highway geometry, and simulate interaction among drivers, a vehicle, and roadway characteristics (Bittner et al., 2002), but it does not deeply model human cognition, particularly as it relates to how drivers share attention between the road and DVIs. Therefore, it has a limited capacity to evaluate the distraction potential of DVI displays.

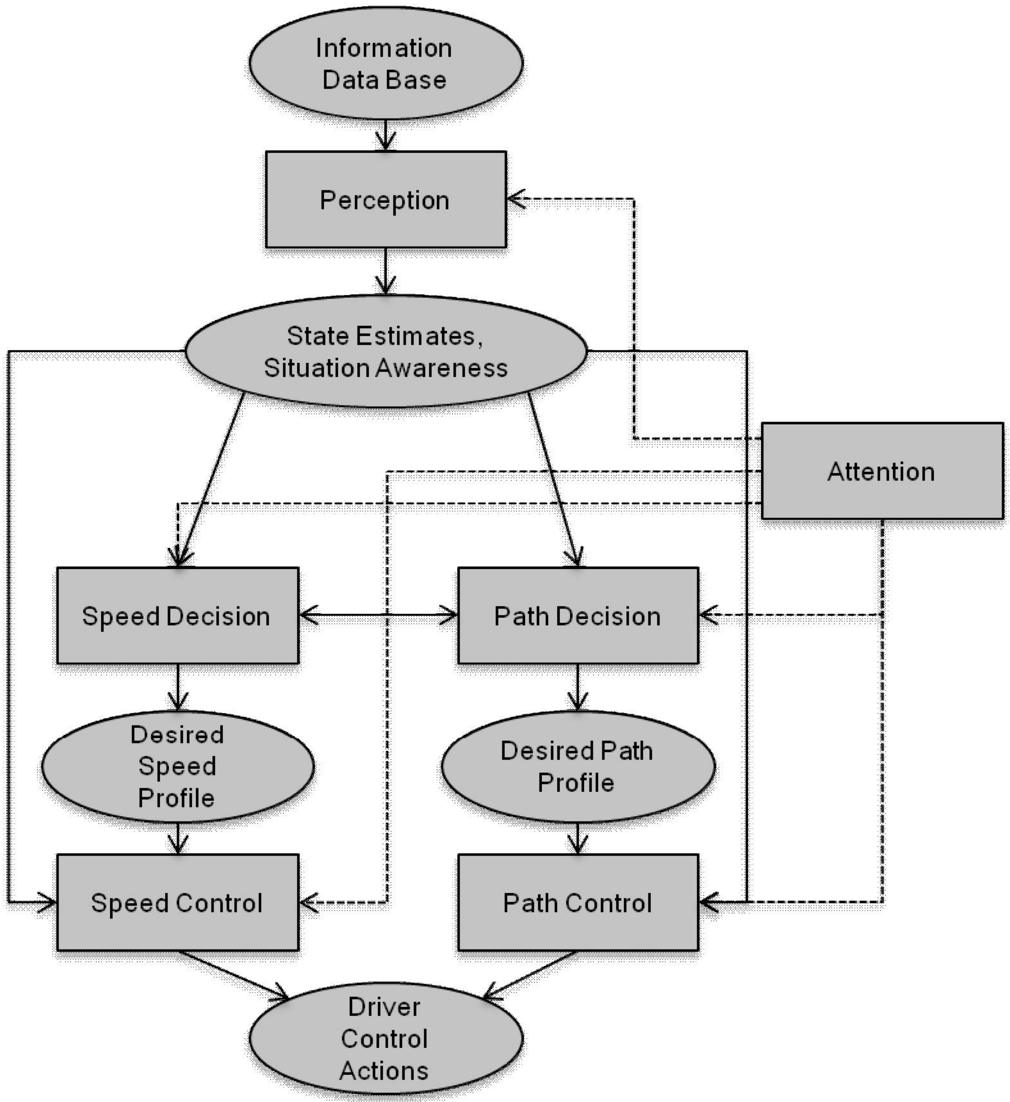


Figure 1. The Driver Performance Model (modified from Levison, Humm, Bittner, & Simsek, 2001).

2.2.2. The SEEV Model

Wickens and colleagues (2001) reviewed models of visual information acquisition and identified four factors that affect visual attention. The four factors are physical "salience" of the scene (or objects), "effort" needed to move attention from a previously fixated location, "expectancy" of the signal, and objective "value" of processing information. Based on these four factors, Wickens et al. (2001) developed a descriptive model, known as the SEEV (Salience, Effort, Expectancy, and Value) model, to predict scanning and distribution of visual attention. In this model, the expectancy and value components account for top-down processing, whereas the salience and effort components account for bottom-up processing. The equation below shows that the four elements of this model combine in a linear fashion to predict probability of attending to an area (Wickens & McCarley, 2008). The variables in the equation are described (Wickens & McCarley, 2008) as "S reflects the strength of salience, EF reflects the inhibitory strength of effort, EX, or expectancy, reflects the collective forces of bandwidth (i.e., event frequency along a channel) and contextual cueing (i.e., event frequency given that an information cue has occurred), and V refers to the value" (p. 253).

$$P(A) = s\boldsymbol{S} - ef\boldsymbol{EF} + ex\boldsymbol{EX} + v\boldsymbol{V}$$

(Note: Uppercase letters reflect the strength of each factor and lowercase letters represent coefficients for each factor in the equation.)

The SEEV model has been applied to study effects of in-vehicle tasks on drivers' performance. For example, Horrey and his colleagues (2006) applied the SEEV model to drivers' visual attention while interacting with in-vehicle technologies (IVT). In this model, salience was defined as a property of specific events in a given area of interests rather than physical properties of area of interests. The experiment examined visual attention allocation to locations with varying task bandwidth (i.e., "expectancy" in the model) and task priority (i.e., "value" in the model). Task priority was manipulated by offering incentives that prioritized the driving task, the IVT task, or both. Bandwidth of the driving task was manipulated by varying the frequency of wind turbulence, and bandwidth of the IVT task was manipulated by varying the rate of IVT presentation. The results showed a greater proportion of scanning to the outside world when driving was prioritized, whereas the proportion of scans to the outside world decreased as IVT bandwidth increased. Although the experiment measured driving performance (e.g., lane position), but the SEEV model did not predict driving performance.

Recently, Steelman et al. (2011) modified the SEEV model by integrating Walther and Koch's (2006) salience model. However, there are limitations in applying this model to support design and to evaluate in-vehicle displays. For example, the modeler has to assign a "pertinence score" to each factor based on the usefulness for a given task condition/phase (e.g., in the study, two of the authors independently assigned pertinence scores for each condition, and the area of interests' value and expectancy were estimated by a subject matter expert). Another limit of the model is that it cannot predict switching between tasks, whereas a critical aspect of distraction is the potential of a highly salient object to attract long glances away from the road.

2.2.3. IVIS DEMAnD

The In-Vehicle Information System (IVIS) Design Evaluation and Model of Attentional Demand (DEMAnD) was developed to help designers of IVISs to evaluate attentional resources demanded by interactions with IVISs and to estimate the impact of these interactions on driving performance (Hankey, Dingus, Hanowski, & Wierwille, 2000). DEMAnD is not a computational model of driving per se, but rather a set of lookup tables and regression equations that combine task demands. The underlying theoretical assumption of DEMAnD is that drivers act with limited attentional resources and that resources demanded by a secondary task (e.g., interacting with IVISs) draw away from resources available for the primary task (e.g., driving) and lead to decreased driving performance (Hankey et al., 2000). The behavioral model consists of five demands on drivers' cognitive resources: visual, auditory, supplemental information processing, manual, and speech (Figure 2). The model uses variables to measure demands, such as estimated single glance time to a display (in the visual resource), estimated number of message presentations (in the auditory resource), hand-at-task time (in the manual resource), and estimated number of command attempts (in the speech resource).

Several field studies were conducted to develop DEMAnD. Data from these studies and preexisting real-world data on driver-task measurement were fit with regression equations to produce a task library that includes 198 tasks such as "adjust temperature" (Hankey et al., 2000). Users can specify characteristics of tasks (e.g., frequency of use, roadway complexity, etc.), driver population, driving environment, and other modifiers relevant to designs (e.g., display density, character height, etc.), and DEMAnD provides a hypothetical workload level and a summary of driving task performance (Hankey et al., 2000). One limitation of DEMAnD is that it does not provide specific measurements of performance such as lateral position variability or speed deviation. Moreover, the model does not incorporate a detailed process model of driver behavior that would enable designers to assess a wide range of tasks and interface features: analyses are limited to the predefined catalog of tasks.

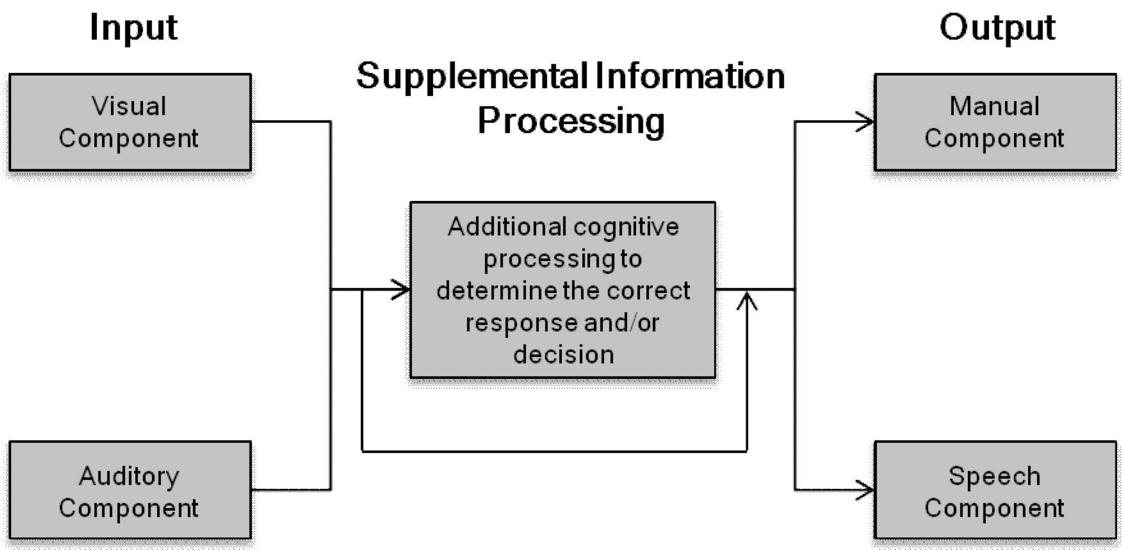


Figure 2. The behavioral model in DEMAnD (modified from Hankey, Dingus, Hanowski, & Wierwille, 2000).

2.2.4. ACT-R and Distract-R

ACT-R (Adaptive Control of Thought-Rational) (Anderson, 2007) is a cognitive architecture that simulates human cognition by incorporating both the abilities and limitations inherent in human problem solving and response selection, including those involving memory, procedural skills, learning, and perceptual-motor behavior. Figure 3 shows a general structure of ACT-R. The model assumes that human knowledge can be divided into two: “declarative memory” (e.g., fact-based knowledge) and “procedural memory” (e.g., skill-based knowledge). The model consists of several computational modules. Perceptual-motor modules (e.g., visual module and motor module) interact with the real world (e.g., environment) and memory modules (e.g., declarative memory and procedural memory) retrieve information from memory aid and cognitive procedure. Buffers link modules and enable communication between modules (Anderson, 2007).

Using ACT-R as a base, the ACT-R driver model (Salvucci, 2006) was developed to model driver behavior. The driver model has been validated with respect to curve negotiation and lane-changing behavior in a highway context, as well as for steering and braking performance under distraction (see Salvucci & Taatgen, 2008; Salvucci, 2005).

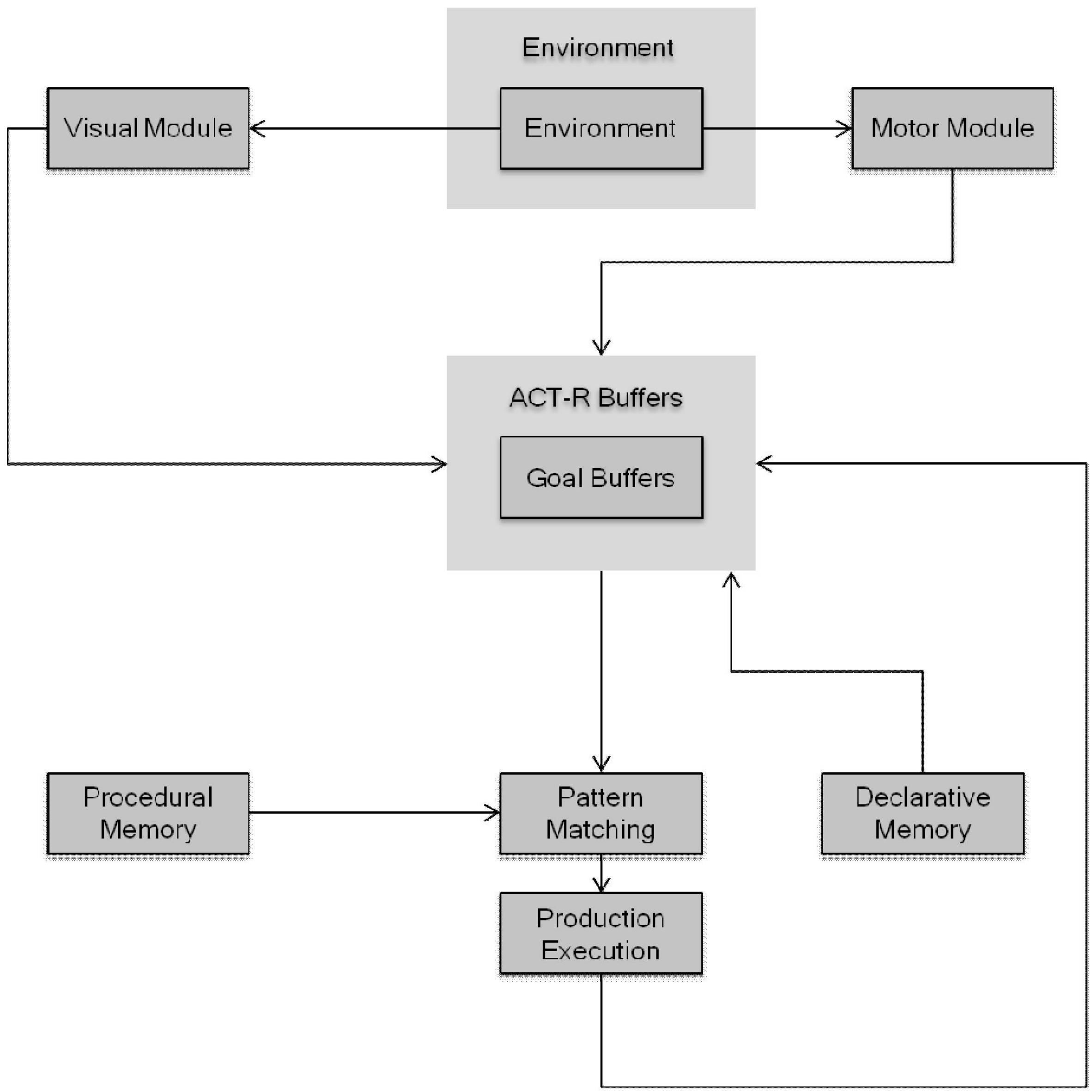


Figure 3: The general structure of ACT-R.

Salvucci and Taatgen (2008) developed threaded cognition theory to enable ACT-R to model multitasking situations. The core idea of threaded cognition can be summarized as "multitasking behavior can be represented as the execution of multiple task threads, coordinated by a serial cognitive processor and distributed across multiple processing resources" (Salvucci & Taatgen, 2008, p.103). Threaded cognition simulates task switching between driving and interacting with DVIs. For example of "driving and dialing", a driving thread performs checking lane position, adjusting steering angle, accelerating and decelerating, and checking the stability of the vehicle. Switching to a telephone dialing thread depends on the last process of the driving thread, checking the stability: "If the vehicle is not sufficiently stable, the driving thread maintains control and iterates until stable; if the vehicle is stable, the dialing thread intercedes" (Salvucci & Taatgen, 2008, p.122). Validating the model with data from participants who dialed while driving showed that the model could successfully predict the total time needed to complete the dialing task (R^2 = .99) and the vehicle's lateral velocity (R^2 = .96) (Salvucci & Taatgen, 2008).

Although the ACT-R driver model can predict driver distraction, the formal computational model requires experience with the ACT-R modeling language and programming of the driving environment. To address this difficulty, Distract-R (Salvucci, 2009) was developed to provide a rapid prototyping and evaluation environment based on the ACT-R driver model; in essence, a designer can benefit from the ACT-R architecture, but the model itself is largely "under the hood," freeing the user to specify design alternatives and scenarios pertinent to their interests. Distract-R has been shown to generate reasonable predictions of distracted behavior for a few DVI designs (Salvucci, 2009). However, Distract-R and ACT-R are currently limited with respect to their account of visual attention: Although they can account for the timing of visual attention shifts and eye movements between visual objects (Salvucci, 2001), they rely on the human modeler to explicitly state how attention is directed from one place to the next. The model has an ability to determine timing to initiate the secondary task, but it has not an ability to pause the secondary task to switch back to the primary task. It would be more useful if the driver model itself determined its path of visual attention, as driven by cues such as visual salience, and if the driver model switched between the primary and the secondary task when the secondary task takes a long time to complete.

2.3. Visual Attention and Visual Distraction

Attention has been studied for many years. James (1890) described attention as, "Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought" (p. 404). Over the past century, increasingly sophisticated models make it possible to quantify the attention more precisely.

Interactions with DVIs often involve search tasks, such as finding a location on a map, choosing between options on a menu, or selecting from a list of options. Search tasks require visual attention to find a specific target object in a scene. Such search tasks depend on both top-down, goal-driven and bottom-up, stimulus-driven processes. These processes have been considered in Posner's work (e.g., Posner, 1980) describing the orienting of attention, Treisman's feature integration theory (e.g., Treisman, 1988), and Wolfe's guided search models (e.g., Wolfe, 1994). Posner's work introduced "attentional shift", which directs visual attention to a point of interest to increase an efficiency of processing information. Treisman's feature integration theory showed how different sensory features could be coded and combined into a master module to perceive an object. Wolfe's guided search model introduced a concept of "activation" that guides attention. The concepts used in these theories provide a background for the general structure of the visual salience model applied in this dissertation.

2.3.1. Orienting of Attention

Orienting of attention can be defined as aligning of attention to a sensory input (Posner, 1980). The term "orienting" has been further divided into overt orienting, shifts of attention with changes in body position, and covert orienting, shifts of attention without changes in body position (Posner, 1980; Wright & Ward, 2008). In Posner's experiments, participants were presented with location cues including potential location of detection stimuli or neutral cues that do not provide any location information. Valid cues represent when the target appeared at a cued location, whereas invalid cues represent when the target appeared on the opposite side of a cued location. Efficiency can be calculated by comparing the mean response time of the neutral cue condition to the valid cue condition. A difference in mean response time between neutral cue and valid cue conditions is referred to as a "benefit", whereas a difference between neutral cue and invalid cue in mean response time is referred to as a "cost". "Benefit" can be explained by a covert shift to the cued location, because the attentional shift was initiated to the cued location before the target appears (Figure 4). For example, Posner and colleagues showed that the difference between the mean valid-cue trial and mean neutral-cue trial (i.e., "benefit") was 23 ms, and the difference between the mean invalid-cue trial and mean neutral-cue trial (i.e., "cost") was 40 ms (Posner, Nissen, & Ogden, 1978).

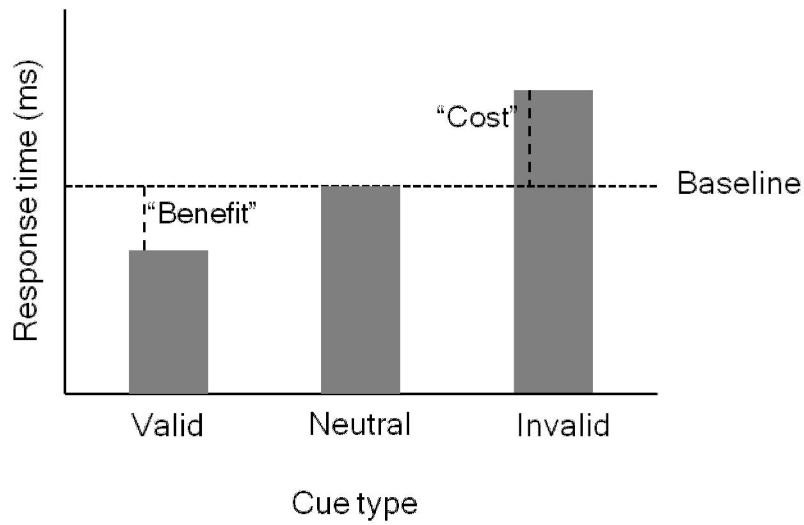


Figure 4. The costs and benefit of valid, neutral, and invalid cues.

One of the main contributions of Posner's work is that it demonstrated an accurate method to measure covert orienting with as much precision as overt shifts. Moreover, other parts of Posner's work focused on the relationship between overt and covert orienting of attention, and demonstrated that those two orienting systems are functionally related rather than completely independent.

2.3.2. Feature Integration Theories

Posner differentiated orienting and detecting, and focused on covert orienting, the stage before a stimulus enters short-term memory and conscious awareness. Treisman and Wolfe's work has extended the research focus to "detection". The main idea of Treisman's model (1988) is that "detection could be triggered by simple features" and "conscious awareness might depend on feature integration" (p. 201). Two stages are proposed in the model: preattentive and feature integration. At the preattentive stage, different sensory features (such as color, orientation, or size) are coded in specific modules and represented with "feature maps". These basic features are coded automatically and spatially in parallel (i.e., it does not need focused attention). The second stage involves focused attention where individual features are integrated into a “master map” (or “saliency map”) to perceive an object (or target).

Two types of visual search tasks typify this research: “feature search” where the target differs from the distractors by a single property such as color or shape, and “conjunction search” where the target and the distractors shares more than one single visual property (i.e., features must be conjoined to specify objects). The single feature search (e.g., pick out a red circle within a group of black circles) leads to faster reaction times and can be performed pre-attentively with parallel processing, whereas conjunction search (e.g., pick out a black horizontal line within a group of white horizontal lines and black and white vertical lines) requires conscious attention and results in slower reaction times due to a serial process of integrating information of multiple features.

Treisman (1985) conducted an experiment based on Posner’s work by using valid and invalid cue conditions to test the effects of “benefit” and “cost” on “feature search” and “conjunction search”. The result showed that the effect of spatial cues that indicate where the target will occur was trivial in “feature search” tasks, whereas there was a substantial benefit in “conjunction search” tasks, because the valid cue eliminated the serial checking phase. Treisman and Gelade (1980) also found that search time increased linearly with display size (i.e., the number of distractors) for conjunction search, whereas search times were hardly affected by the number of distractors in feature search.

An important contribution of the study is that it defined the two stages—feature and conjunction search—and their characteristics (parallel vs. serial). It showed that there is a limited set of features that can be processed in the first stage (e.g., the feature search).

2.3.3. Guided Search Model

Wolfe (1994) proposed guided search as a model of human search behavior based on Treisman's feature integration model and other previous work (e.g., Egeth, Jonides, & Wall, 1972; Hoffman, 1978). The main concept of this model is "… that attentional deployment of limited resources is guided by the output of the earlier parallel processes…." (Wolfe, 1994, p. 202). The model assumes limited capacity of visual system: the first stage discards partial input and the second stage selectively processes information by linking different features that can differentiate the target from distractors. The concept of feature maps is also incorporated into the guided search model. However, the guided search model elaborated each feature map into separate maps for each feature. The model implemented a concept of "activation" of location in each feature map, and the relationship between activation and attention can be described as "the greater the activation at a location, the more likely it is that attention will be directed to that location" (Wolfe, 1994, p. 205). The "activation map" (also called "activation field") is a promising concept to integrate two attentional control processes (i.e., bottom-up and top-down mechanisms), because it provides an intermediate level. Similarly, LaBerge and Brown (1989) also used the concept of "activity distribution" and developed a model in which attention can be directed to the location of one of the activity distributions in either a top-down or a bottom-up manner, and the intermediate-level activity distributions consist of goal-driven/stimulus-driven activity distributions (Figure 5). Top-down activation depends on goal-driven input (e.g., "attend to a red object"). Bottom-up activation is based on stimulus-driven input. For example, the saliency map models a process of the bottom-up activation to predict attended locations by computing the effects of color, orientation, and intensity. The saliency map will be described in more detail in the next section. The activation map that sums all activations from all feature maps is equivalent to the “master map” in the feature integration model. In a process of visual search, if the highest activation location does not include a target, attention moves to the next highest location until the target is found. By using the concept of activation map, the model can easily explain and predict preattentive parallel visual search and target salience (or similarity between targets and neighbors).

As an example of how the activity distribution model operates, Figure 5 has two circles on a visual scene (the bottom layer named “Visual Scene”). The “Stimulus-driven Field” represents an activation field for the visual scene. The peak height represents a level of activation. That is, higher peaks mean more activation, and the highest peak in the scene is likely to be attended at first. In this example, those two circles stand out from their surround (or background), but the activation levels for both objects are nearly the same (so the probability of being attended will be the same as well). The “Goal-driven Field” represents a top-down activation that can be promoted by goal-driven input such as “attend to a red object” or “target object will appear on right”. The “Intermediate-level” sums activations from the “Stimulus-driven Field” and the “Goal-driven Field”. After the combining process, the red circle on right becomes the highest peak on the scene, and is likely to be attended.

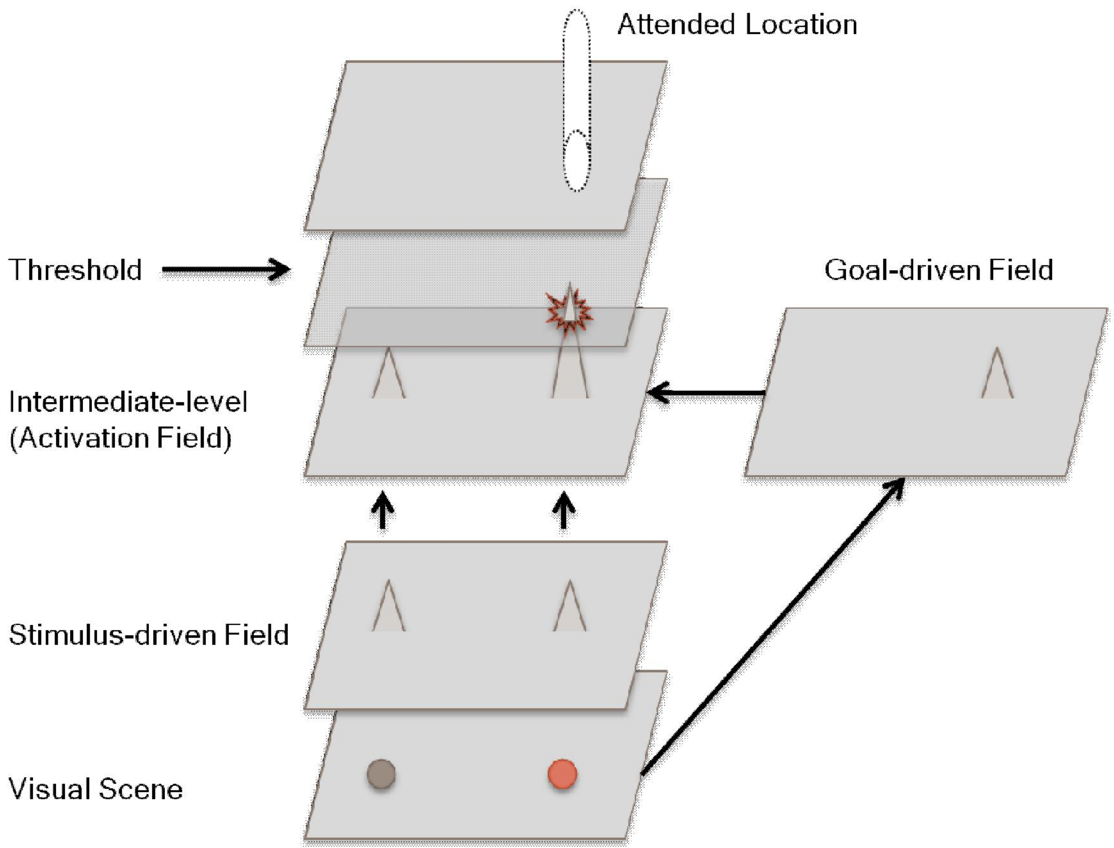


Figure 5. LaBergian Activity Distribution Model (adapted from Wright & Ward, 2008).

2.4. Modeling Visual Attention to Assess Distraction Potential of Vehicle Systems

This section describes a visual salience model and how it will be integrated into Distract-R, providing the system with a rigorous model of the factors that influence driver's visual attention. The model of visual salience is fine-tuned for describing visual attention to vehicle displays—a critical concern for the design of DVIs. Integrating the saliency model into Distract-R is a fundamental process to enhance the current version of Distract-R.

2.4.1. The Saliency Map

The saliency map is a promising approach to assess how features of a scene influence visual attention. The saliency map represents visual saliency of a scene by modeling the factors that attract attention (Koch & Ullman, 1985). One advantage of the saliency map is that it captures the influence of visual feature contrast, along with local absolute feature strength (Itti & Koch, 2001). Koch and Ullman (1985) originally introduced the concept of a saliency map to predict preattentive selection by encoding the saliency of the objects in the visual environment. Subsequently, Walther and Koch's model (2006) has been validated to predict sequential attention to objects in a complex scene by using three low-level visual features— color, intensity, and orientation—that may attract (and thus distract) drivers' visual attention.

The salience map builds on a series of computations that reflect the neurological processes associated with visual attention. The input image is sub-sampled and filtered by a Gaussian Pyramid that is widely used in image processing. Each pyramid level is decomposed into feature channels based on three low-level features (i.e., color, intensity, and orientation). Color is derived from red, green, blue, and yellow RGB values. Intensity is computed using an averaging function of the color features. Orientation (0^o, 45^o, 90^o, and 135^o) is computed by applying a Gabor filter (a linear filter for edge detection) to the images in the intensity pyramid. These features are summed by center-surround combinations and normalized. The resulting feature maps are combined into conspicuity maps—one for color, one for intensity, and one for orientation. Finally, the conspicuity maps are combined into a single saliency map that represents saliency as a scalar quantity at every location in the scene. In the saliency map, the location of the focus of attention is estimated using a Winner-Take-All (WTA) neural network that selects the most active location and suppresses the other locations.

The saliency map computes the influence of low-level visual features of a display that attract driver's attention, which can be the features that might draw driver's attention away from the area of the display that contains the information of interest to the driver. Predicting the salience of display features is central to estimating distraction potential, because it determines how likely an object is to attract or distract drivers' visual attention, and consequently, how many fixations might be required to find the desired information. If important information is highly salient relative to the background, it will be detected easily with a few fixations. However, misplaced salience—situations in which highly salient display features do not correspond to drivers' information needs—might lead to many fixations and long glances away from the road. The saliency map can identify instances of misplaced salience, which could help designers reduce the distraction potential of DVI displays. This study proposed an integration of the saliency map into Distract-R to enhance Distract-R's visual module. The saliency map has not been validated for predicting search time, and has a limited ability to predict variability. Therefore, these limitations need to be addressed before the integration.

2.4.2. Integrating the Saliency Map into Distract-R

This dissertation integrates Distract-R and the saliency map, and Figure 6 shows how the saliency map can support Distract-R and how those two models can be integrated. The shadowed box that includes "saliency map" and "activation map" on the left indicates how the modification will fit into the original structure of ACT-R.

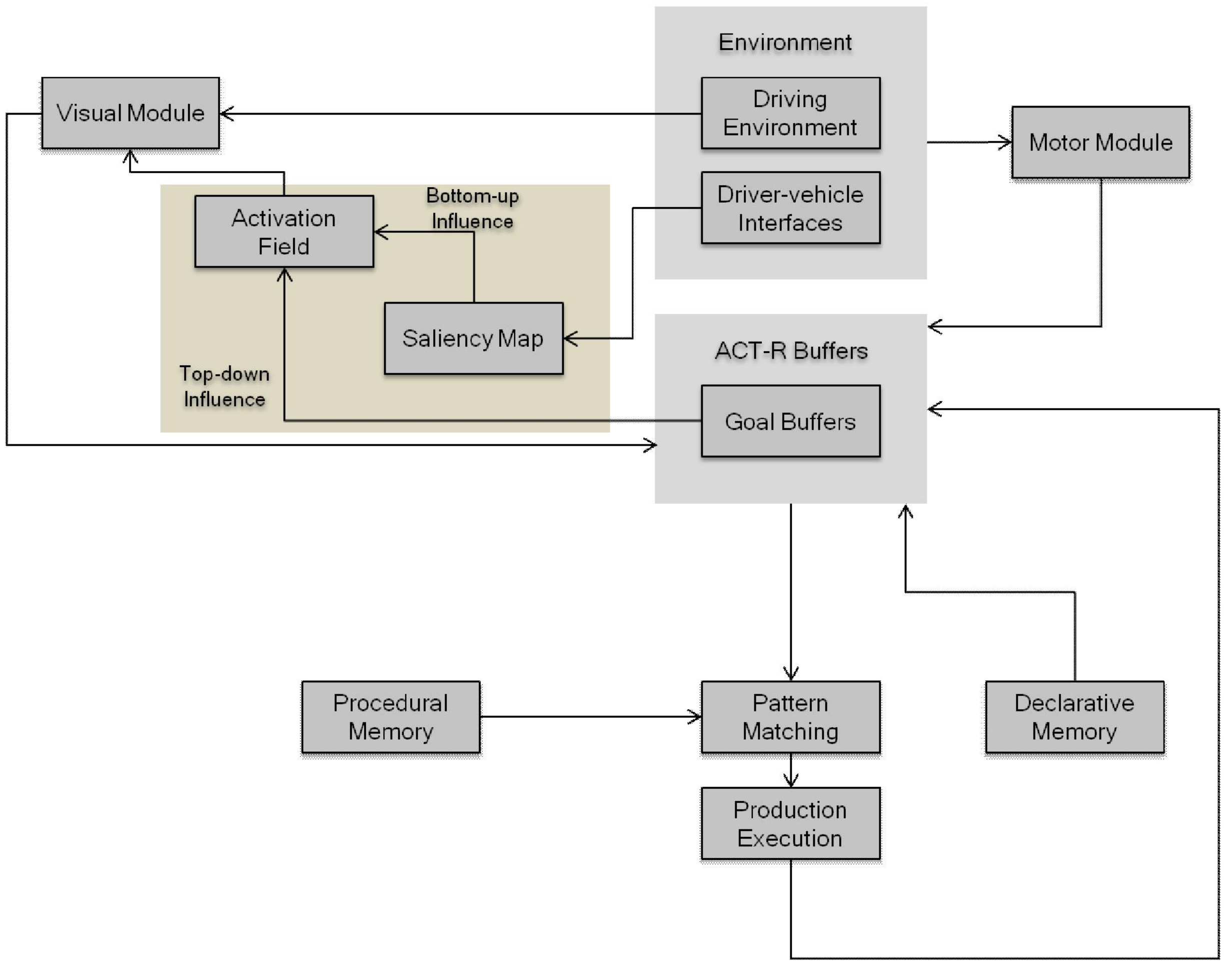


Figure 6. A structure of the integration between the saliency map and Distract-R models.

Figure 7 shows the integration of the saliency map and Distract-R, and how these models complement each other in guiding visual attention. Distract-R considers drivers' goal-driven, top-down processes and the saliency map considers drivers' stimulus-driven, bottom-up process.

The saliency map in the proposed model simulates the feature integration models (Treisman & Gelade, 1980; Wolfe, 1994) to account for an influence of visual features. Distract-R will be modified to model the influence of drivers' expectation on visual attention. The activation field that links the saliency map and Distract-R simulates this influence and specifies the most likely location of glances (Laberge & Brown, 1989). The role of the activation field is to link the top-down, symbolic influence on visual attention (e.g., expectation) from Distract-R to the bottom-up, spatial representation of the saliency map, and to enable Distract-R to simulate the path of visual attention.

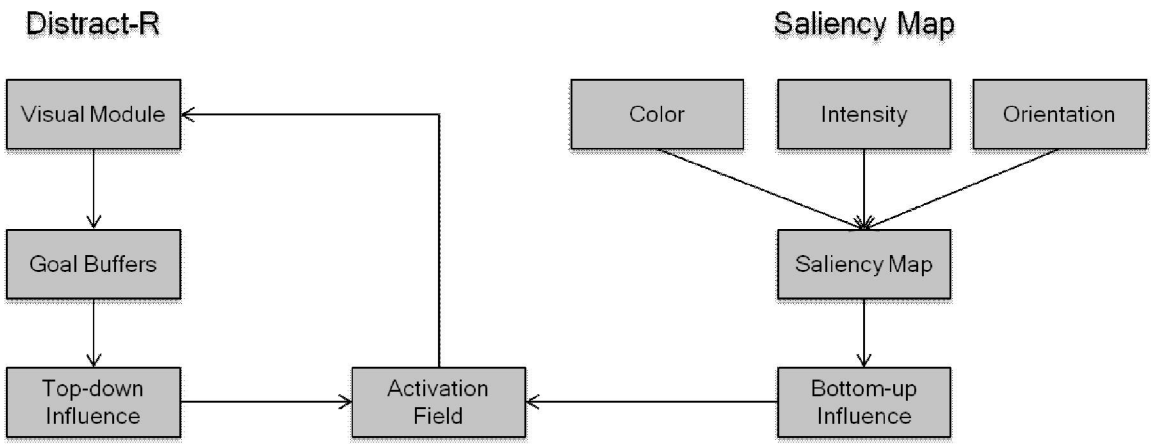


Figure 7. A structure of the integrated model that combines Distract-R and the saliency map.

2.5. Summary

As a first step to develop a computational model for evaluating the distraction potential of DVIs, previous models of driver behavior and performance were reviewed. Based on the review of the previous models, it is concluded that a model to evaluate driver distraction must include the ability to: (1) consider threaded cognition to predict switching between tasks, (2) provide measurements of driving performance, and (3) account for top-down and bottom-up influences that might affect drivers' visual attention. The Distract-R model can simulate and model driver's task switching with specified secondary tasks, and predict task and driving performance in specified task and driving situations. Distract-R can also account for top-down influences on drivers' visual attention. Therefore, Distract-R satisfies the first two requirements listed above and part of the third. The saliency map can account for the influence of bottom-up visual features—a feature missing from the Distract-R model and needed to satisfy the remainder of the third requirement. The proposed model will integrate Distract-R and the saliency map and will thus have a conceptual structure that is expected to satisfy all of the identified requirements.

This integration involves several steps. First, the saliency map is a deterministic model that does not account for variability, so a modification is required to simulate human performance variability. Second, the saliency map's ability to predict search time has to be tested and validated. Third, the model integration will require technical modifications of Distract-R and the saliency map to link them. Last, the integrated model must be validated to assess whether the integrated model can predict drivers' interaction with DVIs while driving. The following chapters describe the accomplishment of these requirements.

The integrated model will have several practical benefits, such as complementing existing design guidelines. For example, Federal Highway Administration guidelines (Campbell et al., 1998) suggest that highly saturated blue should be avoided, but this guideline only considers the symbol color rather than the contrast between symbols and background color. In contrast, the proposed model could identify the effect of using specific colors or other visual features based on given backgrounds. As another example, guidelines published by the University of Michigan Transportation Research Institute (Green, Levison, Paelke, & Serafin, 1995) suggested using "discriminable colors." However, what constitutes a "discriminable color" depends on its interaction with other colors and display features. It is difficult for guidelines to indicate how combinations of interface features will influence driver attention to display, but such effects can be captured with the salience map process in the proposed model.

The integrated model can also guide designers who might want to spatially group information of similar type in how to help drivers develop spatial expectations for specific target information. Given current DVIs' multiple sources of information and overall information capacity, spatial grouping/segregation will be an important design feature, and designers will need to consider the influence of top-down along with bottom-up attention mechanisms. The integrated model will be able to examine the effect of coordinating salient visual features and drivers' expectations. Moreover, the integrated model will indicate how potential misplaced salience might affect driving performance and how top-down processes might compensate for the effect of misplaced salience.

The subsequent chapters describe the process of model integration. Chapter 3 presents modification and validation of the saliency map for the integration. Chapter 4 describes technical and conceptual details of the model integration. These two chapters describe the successful integration of Distract-R and the saliency map.

3. Validation of the Saliency Map

3.1. Objectives

A main role of the saliency map in the proposed model is to estimate search time for predefined targets. The saliency map is one of only a few computational models well-suited for modeling visual search by accounting for low-level visual features. It represents visual saliency of a scene by modeling the factors that influence attention in a bottom-up fashion (Koch & Ullman, 1985). The saliency map has been validated for predicting gaze locations (e.g., Harel, Koch, & Perona, 2007; Peters, Iyer, Itti, & Koch, 2005); however, its ability to predict search time has been validated in only a few empirical studies (e.g., Itti & Koch, 2000). Moreover, the saliency map does not account for variability of human performance, which can be critical in identifying the likelihood of long glances. The objectives of this chapter are to discuss how a Monte-Carlo technique was implemented into the saliency map to account for glance duration variability and to discuss the validation of the saliency map's ability to predict search time before integrating it with Distract-R.

To evaluate distraction potential, predicting search time in various situations is essential. Visual features of DVIs can affect the search time and the length of time a driver's eyes are off the road. A relationship between search time and number of items (i.e., "set size") has been studied by numerous researchers (e.g., Wolfe, Cave, & Franzel, 1989). Highlighting information (or a target) using a different color for a target is well-known method to shorten search time (e.g., Fisher & Tan, 1989). Changing the set size and highlighting a target have important implications in designing DVIs. Excessive information can confuse drivers by obscuring important information, extending search times, and leading to long glances away from the road.

Highlighted information attracts drivers' attention and can decrease search time if the highlighted information coincides with the search target.

The basic experimental paradigm of targets and distractors is applied in this study by varying the set size and highlighting conditions (e.g., Treisman & Gelade, 1980). It is expected that search time will increase, as the set size increases (i.e., clutter will increase search time). Moreover, highlighting target information is expected to mitigate the effect of set size on search time. The rationale for choosing such well-known effects was to assess whether the modified saliency map could predict the effects observed in empirical data. The saliency map is expected to predict the effect of the set size and highlighting, and their interaction.

3.2. Saliency Map Modification

A previous study (Itti & Koch, 2000) compared model reaction time to human reaction time based on the relation between an average of 40 ms per model shift (e.g., shifts to the next most salient location) and an average of three shifts of attention per second for humans. The saliency map simulates the cognitive process of "integrate and fire neurons" and the parameters of the model "neurons" have been roughly adjusted to match real neurons. The previous research scaled the model's simulated time to real time and added 1.5 seconds to account for latency of human motor response. The previous research (Itti & Koch, 2000) found a poor correlation between human and model search times: the saliency map outperformed humans in search tasks. One explanation for the superior performance of the model was that top-down influences might play a significant role in the deployment of attention, in a way that the model fails to consider. The salience model used in this study also does not account for any top-down influences, because Distract-R will account top-down influences in the integrated model. However, minor modifications in parameter settings and a Monte-Carlo simulation are implemented to model variability of human response.

This study uses the MATLAB Saliency Toolbox (Walther & Koch, 2006), to generate the saliency map. Saliency maps were used in a Monte-Carlo simulation to predict search times for targets. The Monte-Carlo analysis used repeated random sampling to generate a distribution of glances for each of images. To replicate the perceptual noise that guides visual attention, random variance was introduced into the model.

The keystroke-level model (Card, Moran, & Newel, 1980) estimated specific time for low-level operations (e.g., button press from best typist: 0.08 seconds, button press from average skilled typist: 0.2 seconds, and button press from worst typist: 1.2 seconds). The saliency map was modified to generate data for a distribution of virtual participants based on keystroke-level data to model motor responses and to estimate search time. Implementing those distributions is an important difference from the approach used in the previous study (Itti & Koch, 2000).

Assumptions for the model include:

1. When the "focus of attention" region covers the target, it is considered detected.
2. When the model does not find the target within 30 seconds of scaled time, the simulation is terminated.
3. Inhibition of return (a function that restricts visual attention to areas except previously attended locations) suppresses the fixated locations until the simulation ends.

3.3. Method

3.3.1. Participants

Participants were required to be in good health, and were screened using a telephone script that included study inclusion/exclusion criteria. The study included 30 participants from four age groups: 8 between 18-24 years, 8 between 25-39 years, 7 between 40-54 years, and 7 between 55-75 years.

3.3.2. Apparatus

Data were collected using a PC-based driving simulator, the MiniSim (Figure 8), at the National Advanced Driving Simulator (NADS). The simulator consisted of three plasma widescreens (42” diagonal measurement, 130^{o} horizontal and 24^{o} vertical field of view at 48” viewing distance).

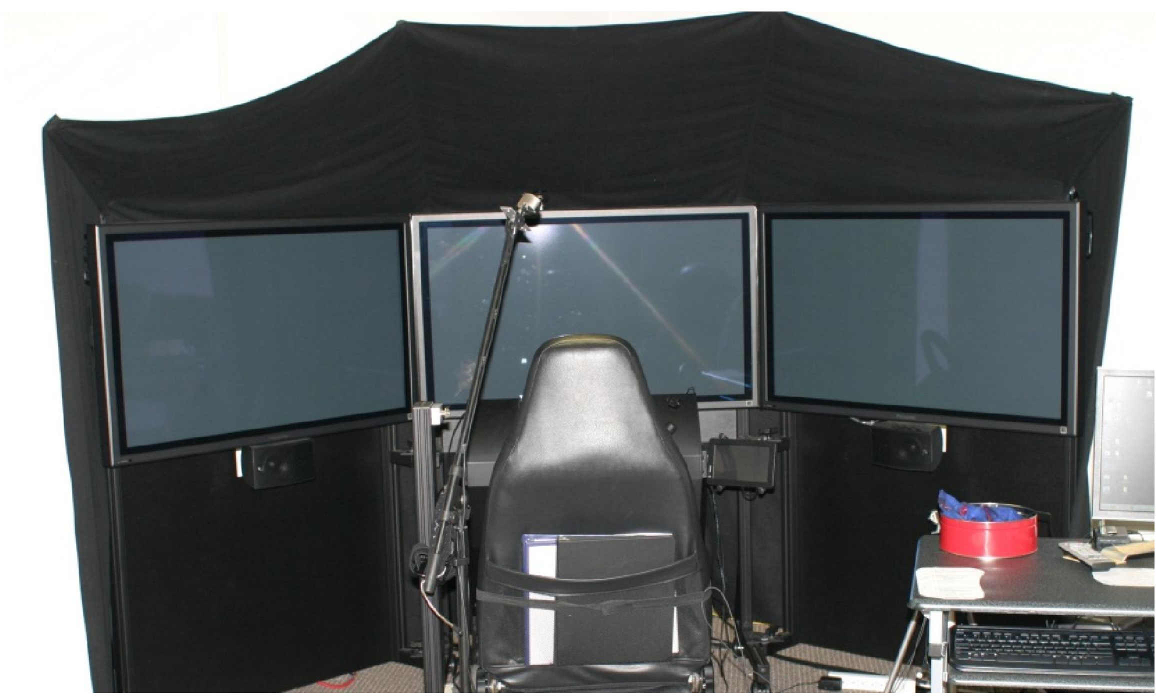


Figure 8. MiniSim simulator.

3.3.3. Design

Three icon set sizes (9, 18, and 27) were manipulated to test an effect of cluttered displays with icons on search time. Three highlighting conditions were also manipulated: (1) no highlighting, (2) correct highlighting, and (3) incorrect highlighting. No highlighting condition represents a display that has low target-distractor discriminability (e.g., all icons have the same color) and correct highlighting condition represents a display that has high discriminability (e.g., the target icon has a unique color). Incorrect highlighting represents a situation where a non-target icon is highlighted with a unique color. Therefore, a 3 (set size = 9, 18, and 27) x 3 (highlighting condition = no highlighting, correct highlighting, and incorrect highlighting) repeated measures analysis was employed. Numbers of trials were equal for correct highlighting (e.g., valid cue) and incorrect highlighting (e.g., invalid cue), and trials were presented in a random order. Highlighting validity was maintained as 50% to promote bottom-up influence from highlighting by inhibiting potential feature expectation from participants.

There were nine possible target locations (from cell 1 to cell 9 on the virtual grid), and all locations were presented to participants during each condition. A total of 81 (e.g., 3 x 3 x 9) screen images were available, and all images were presented to each participant on the center display of the MiniSim.

3.3.4. Materials

Icons. The map icons used in the experiment were taken from the Maki project (MapBox, 2013). Twenty-seven icons were selected from the icon pool. The icons were approximately the same size, and icon colors were the same (Figure 9). Each search task defined one target icon, and other icons were distractors.

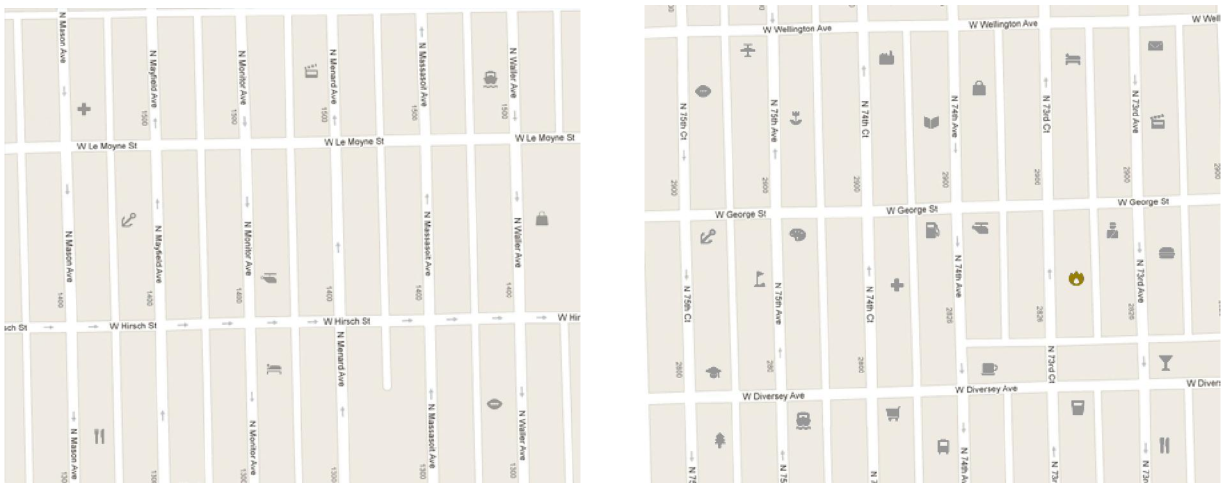


Figure 9. Examples of stimuli (Left: set size = 9, highlighting condition = no highlighting, Right: set size = 27, highlighting condition = correct highlighting)

Background Maps. Stimuli consisted of two layers: one layer contains a map as a background and the other layer contains icons. To select nine background maps, 21 maps (600 x 800 pixels) were initially captured from Google Maps, and previously embedded icons (e.g., bus-stop signs) on the map were removed using Adobe Illustrator, retaining only street names and road shapes. After this process, the feature congestion level for each map was used to control visual clutter for background maps (Rosenholtz, Li, & Nakano, 2007). Nine of 21 maps with feature congestion estimates close to the median feature congestion level were selected (the median congestion level for the final nine maps was 3.85 and standard deviation was 0.05). The remaining analyses assume this selection process controlled for the clutter of the background maps.

Color Manipulation. Google Maps uses several colors for its map icons. From these colors, three were chosen and each color's RGB values [red (151, 52, 52), yellow (151, 127, 0), and blue (48, 123, 191)] were extracted and used to color the targets and distractors.

Target Placement. Eighty-one full-color images of map displays were presented to participants. A target icon was placed on one of nine cells on a virtual 3 x 3 grid. The order of target placement was random, but the icons were evenly distributed so each cell had the same number of icons.

3.3.5. Procedure

Participants were instructed to search for target icons on a map display and press a button to indicate when they found the targets. Participants were seated in a driving simulator and a large map was displayed in the center screen, but they did not drive. In the experiment, two independent variables (set size and highlighting condition) were investigated to assess whether these factors affect participants' search performance, search time, and whether the saliency map can predict those effects.

Informed consent was obtained from each participant and then they completed a demographic survey. Participants were seated in the simulator and given instructions. The practice session consisted of five task instances and participants were allowed to repeat the practice session if they desired. A target icon was presented for one second in the upper left portion of the forward screen, and then a map display was presented on the center of the screen until the participant pressed the space bar to indicate detection. Following this, a 3 x 3 grid was presented and participants were asked to verbally report the target icon's location on the grid (Figure 10). The experimenter recorded the target location (i.e., cell number on the grid) after participants' verbal report.

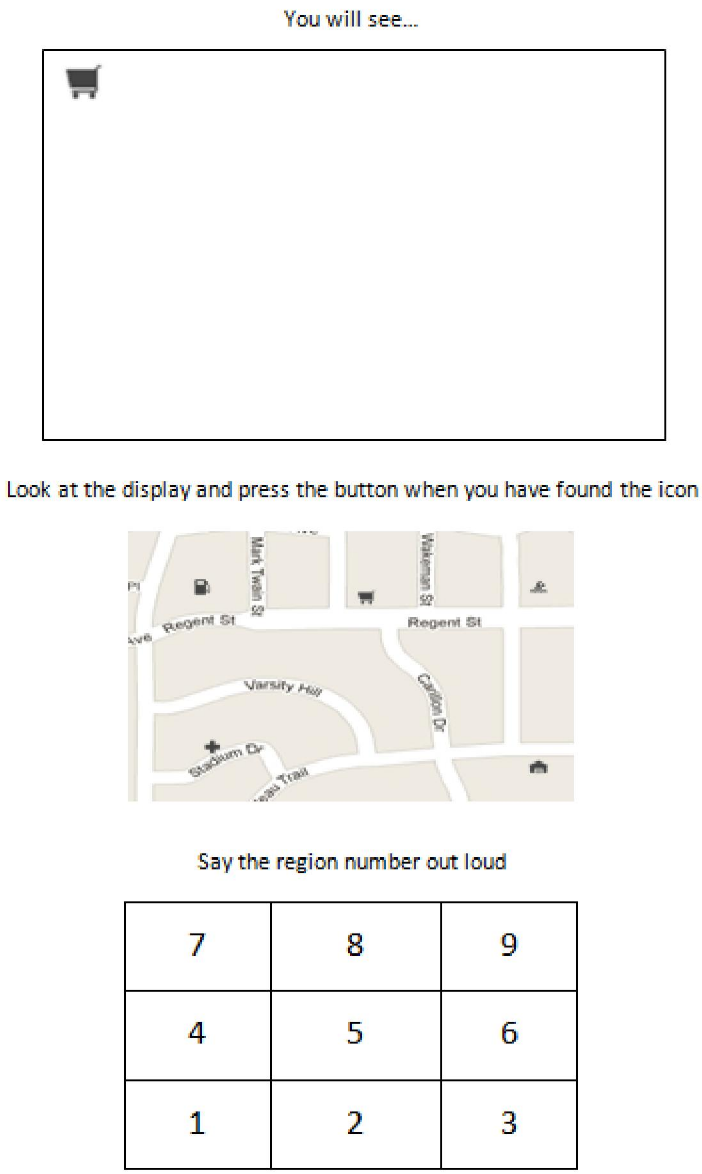


Figure 10. Screen images for the instruction.

3.3.6. Data Reduction

MATLAB (R2011b) and R 2.14.1 (R Development Core Team, 2011) were used for the data reduction and analysis. Consistent with previous research (Fleetwood & Byrne, 2006), incorrect response (e.g., when participants did not correctly identify the target icon) were replaced with the individual participant's overall mean. The rationale for this replacement is that the incorrect response might not reflect search time for the specified target. It was possible to find a correct target, but reported an incorrect target number. However, there was no way to identify the reason for the each incorrect response. Therefore, incorrect responses were replaced with the individual participant's overall mean.

3.4. Results

3.4.1. Response Time Data

This dissertation borrowed the terms, "cost" and "benefit", to explain effects of highlighting. Simply, a difference between the no highlighting condition and the correct highlighting condition (in response time) was represented by a benefit, and a difference between the no highlighting condition and the incorrect highlighting condition was represented by a cost in this dissertation. Figure 11 (left) shows mean reaction time. As expected, set size generally increased response time [F (2, 58) = 107.00, $p < .001$], and appropriate highlighting reduced response time [F (2, 58) = 5.43, $p < .01$]. Post-hoc analyses using paired t-test indicated that response time in the correct highlighting condition was faster than other two highlighting conditions (no highlighting and incorrect highlighting) [t (29) = -3.57, $p < .01$, and t (29) = -2.34, $p < .05$], but there was no significant difference between the no highlighting condition and the incorrect highlighting condition. A cost of incorrect highlighting (e.g., misplaced salience) was expected, but the difference between the no highlighting condition and the incorrect highlighting condition failed to reach statistical significance. In the experiment, only one highlighted item was presented both in the correct highlighting condition and in the incorrect highlighting condition, and this might reduce the magnitude of cost. It is expected that increasing the number of highlighted items on a screen (e.g., highlighting five icons including one target among 12 icons on a map) will increase the magnitude of cost. The interaction effect was significant [*F* (2, 58) = 7.09, $p < .01$], indicating that the effect of highlighting increased with increasing set size.

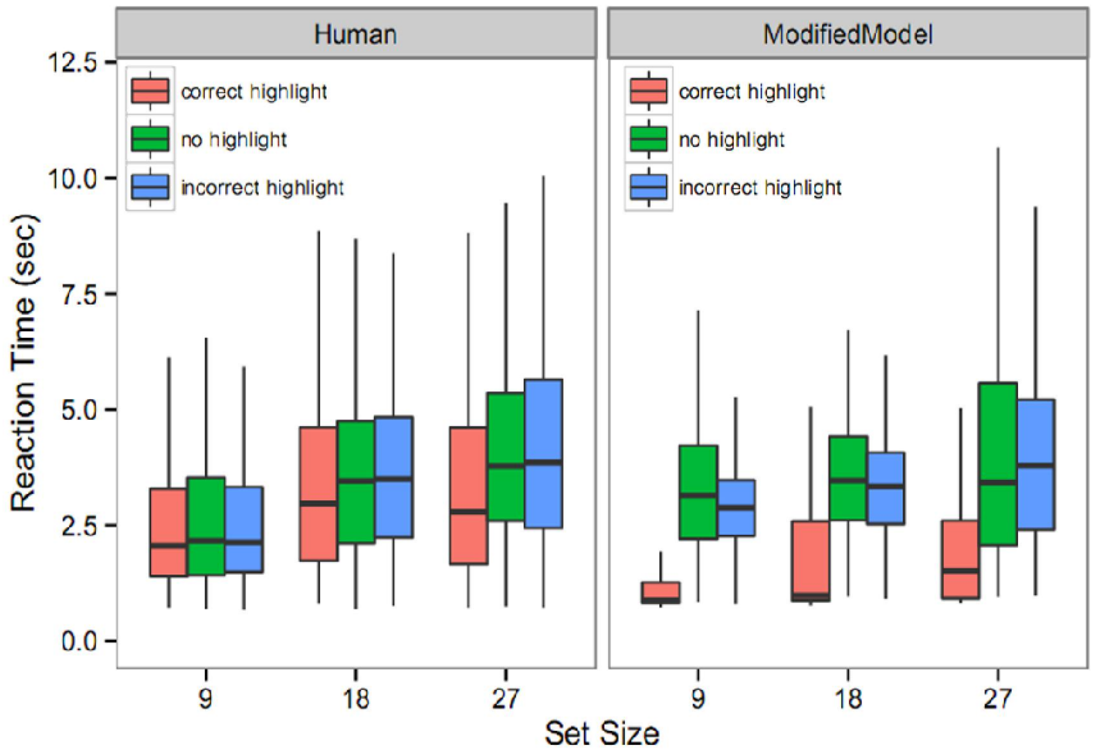


Figure 11. Mean reaction time of the empirical data by set size and highlighting conditions (left) and predicted mean reaction time from the modified saliency map (right).

Figure 11 (right) shows predicted mean reaction time by the modified saliency map. The main effect of set size was observed [*F* (2, 58) = 92.29, $p < .001$], indicating that set size increased reaction time. Appropriate highlighting reduced reaction time [*F* (2, 58) = 775.30, $p < .001$]. Post-hoc analyses using a paired t-test indicated that reaction time in the correct highlighting condition was faster than other two conditions (no highlighting and incorrect highlighting) [*t* (29) = -28.76, $p < .001$, and *t* (29) = -28.82, $p < .001$], and similar to human subject data, there was no significant difference between the no highlighting condition and the incorrect highlighting condition. The interaction effect was significant [*F* (4, 116) = 9.18, $p < .001$].

3.4.2. Response Time Comparison

Previous research (Salvucci, 2009) calculated root-mean-squared error (RMSE) to measure a difference between model predictions and actual observations. The same technique was applied here. An important consideration in comparing the model performance to the empirical data concerns the predictability of the empirical data. If there is little regularity in the data, then the model will inevitably perform poorly. To evaluate the predictability of search times, between-subject variability was calculated. The between-subject variability provides an upper-theoretical boundary. The upper-theoretic boundary is an ability of one set of participant's data to predict other set of participant's data.

The between-subject variability was calculated using split sample validation. Split sample validation consists of two samples from the available data: one sample for calibrating the model, and another one for testing predictability of the model (Pedhazur, 1982). Previous research (3M Commercial Graphics Division, 2010) compared a saliency-based model and human observers using a split-data design to measure the upper-theoretical performance boundary.

To find the upper-theoretical boundary, the first half of participants predicted the second half of participants, and RMSE between those two groups was 3.57. The RMSE between the data from the experiment and the model was 3.20, which was close to between-participants error.

The original (Itti & Koch, 2000) and the modified models were also compared to the participants' data using the Kolmogorov-Smirnov test. The test rejected null hypotheses for both models ($p < .001$). However, the modified model had a smaller distance (e.g., maximum distance between two cumulative probability curves) from the empirical data (D = 0.15 for the empirical data and the modified model, and D = 0.34 for the empirical data and the original model) (Figure 12).

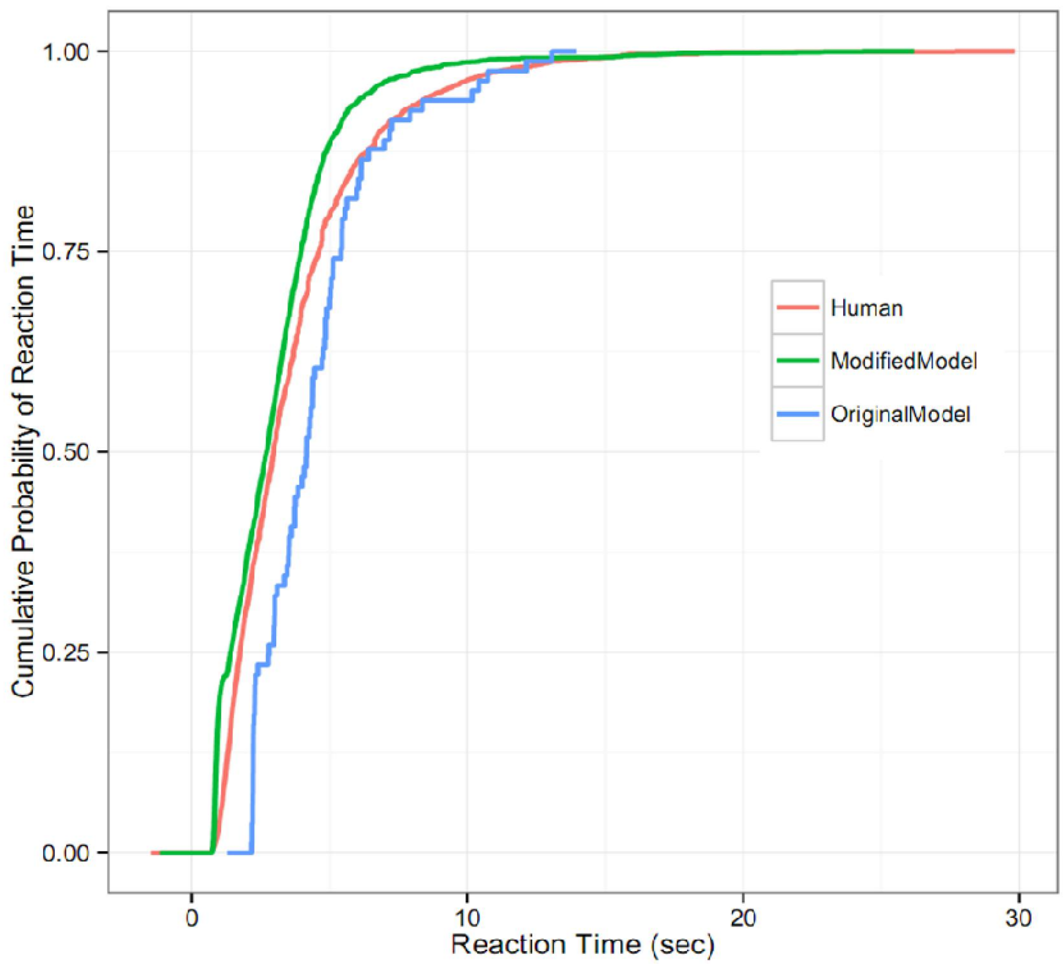


Figure 12. Estimated cumulative distribution functions for the original and modified saliency map models and for the human subjects' data.

3.5. Conclusion

An important contributor to driver distraction associated with complex in-vehicle displays is the search time required to find objects of interest in a cluttered display. Selecting interface elements that minimize this search time is a critical part of designing DVIs. The results showed that the saliency map could be applied to predict search time to a level comparable to between-subject variability—the theoretical limit of prediction. Specifically, the model describes how design features influence search time in terms of the bottom-up visual process. The model predicted effects of set size on search time, which is one component of visual clutter.

The model captured the effect of set size as the empirical data did, but was more sensitive to the effect of color (as established by highlighting condition) than the empirical data. The features tested by the model could substantially affect driver performance, but they are not well addressed by current guidelines. For example, modern vehicles can communicate with roadside infrastructures and other vehicles using the wireless communication system, and the system interface often displays a large amount of information from multiple sources. Highlighting can prioritize information based on a situation, and can be an efficient way to deal with the potential complex displays that might lead to long glances away from the road.

This experiment focused on the bottom-up influence of attention to the interface. The following chapter addresses top-down features (e.g., driver's expectation) and their interaction with bottom-up features. This experiment also focused only on the search task and excluded driving performance. The following chapter will model and collect data from a situation in which people search a screen while driving, making it possible to compare both predicted driving performance and predicted search time with experimental data.

This study extended the saliency map's application in driving research, and provided a foundation to integrate the saliency map into a cognitive architecture. The modified model can help designers who require a systematic approach to assessing and reducing distraction by providing estimated search time for different visual features of a DVI.

4. Integration of the Saliency Map with Distract-R

4.1. Objectives

The previous chapter demonstrated the ability of the saliency map to account for effects of bottom-up visual features. Based on the result, it is expected that integrating the saliency map into Distract-R can complement Distract-R's ability to account for visual features on DVIs. This chapter discusses the integration of the saliency map and Distract-R by modifying each model, and describes conceptual and technical details of the integration.

4.2. Modifications of the Saliency Map

Saliency Toolbox (Walther & Koch, 2006) has been modified to interact with Distract-R. The original model only requires specifying an input image to predict successive fixation locations. Additional information (e.g., target location, expected location for the target) was required to interact with Distract-R, and the saliency map was modified to receive that information from Distract-R.

4.3. Modifications of Distract-R

Distract-R has been modified to define additional parameters, such as target location, and to pass them to the saliency map. The estimated search time calculated using the saliency map is imported into Distract-R and used to model drivers' secondary task performance more precisely. The previous version of Distract-R implemented default parameter values (e.g., 135 ms for visual encoding and 20 ms for eye movements) for search tasks. However, in this version, the modifications enabled Distract-R to predict search time based on low-level visual features of DVIs.

Distract-R was also modified to include an additional parameter that accounts for "maximum glance duration". If the total search time is longer than the "maximum glance duration", Distract-R breaks it up using this parameter. In the modified model, the intention is to set this parameter based on empirical data, and the maximum glance duration is calculated from the empirical data to estimate a value for breaking up long single glances. The original Distract-R has an ability to determine timing to begin the secondary task based on the vehicle's stability. However, once the model driver is engaged in the secondary task, the driver continues the secondary task regardless of the vehicle's stability until completion of the secondary task. This situation might happen in real driving. However, drivers usually pause the secondary task, switch back to a driving task, and resume the secondary task to maintain vehicles stability or to monitor roadways. Drivers might use tactile cues from a steering wheel or a seat to detect vehicle's drift away from center of the road. Drivers might have time threshold in mind for the secondary task, so strategically switch back to the driving when they feel that they engaged in the secondary task for too long (e.g., over the threshold). This study assumes that drivers allocate maximum time for engaging to the secondary task and switch back to the driving task when task time over the allocated time.

Distract-R was originally developed as a rapid prototyping tool rather than a tool to evaluate predesigned DVIs. Prototyping components in Distract-R are sufficient to create very simple interfaces (Figure 13), but might not be sufficient to create complex interfaces that are employed by current vehicle systems. Therefore, a way to import screen images was implemented to enable potential users to import a predesigned screen image for evaluating distraction potential. Figure 14 shows a Distract-R screen that imported an image of the MyFord system, and added "hotspots" (the yellow boxes on the image) that store and pass coordinates of the items to Distract-R.

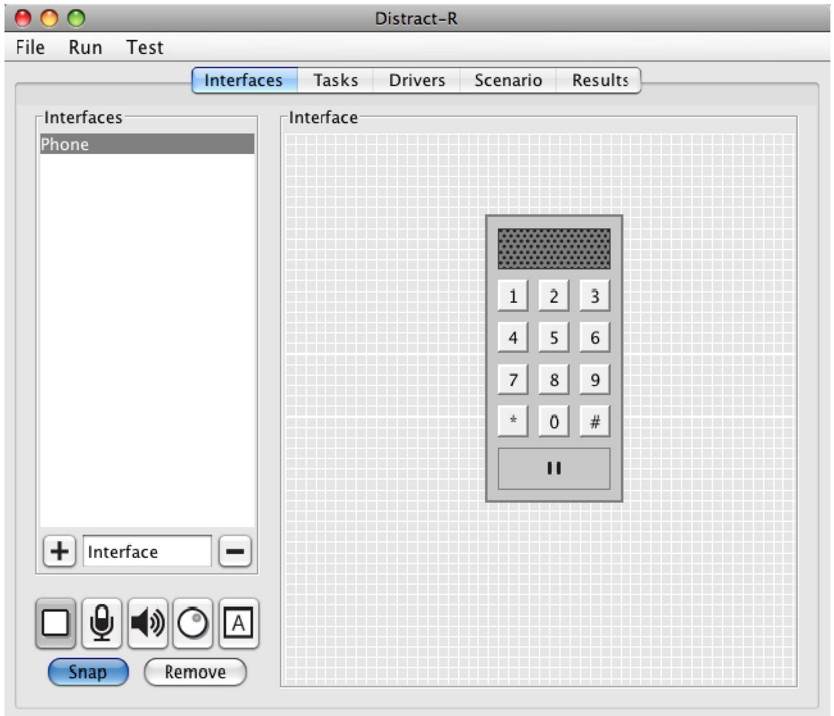


Figure 13. An example of interface prototyped from Distract-R.

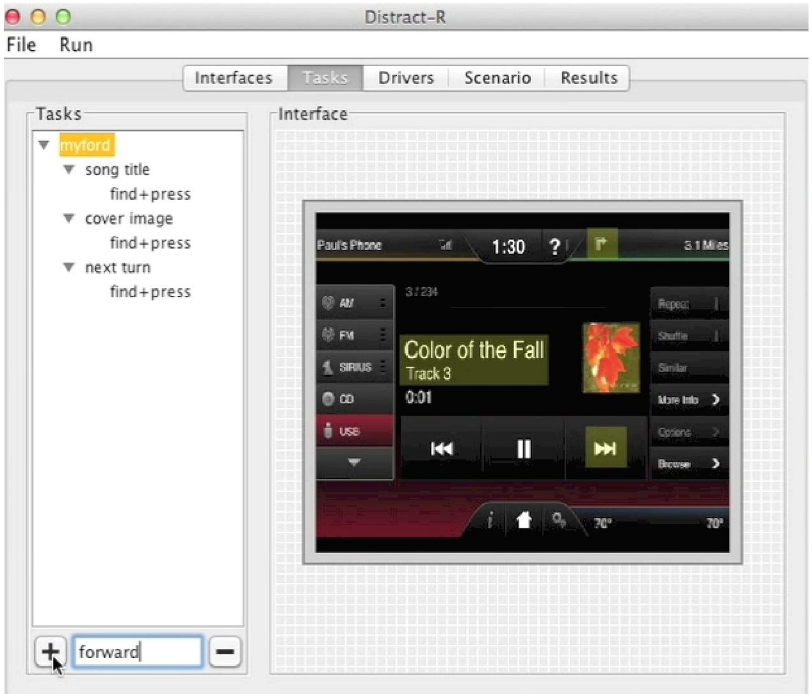


Figure 14. An example of imported screens in the modified Distract-R.

4.4. Integrating Two Models

The components of the models are coded in different programming languages—Distract-R in Java and the saliency map in MATLAB—, therefore a Java application programming interface was used for the technical integration of the two models. An additional Java class that calls the saliency map and imports the estimated search time was implemented in Distract-R.

4.4.1. Activation Field

To create the activation field that links Distract-R and the saliency map, four main modifications were required.

1. Distract-R was modified to define a location of the spatial expectation based on the task specification to build an expectation map.

2. The expectation map was generated based on the expected location by increasing activation values for the corresponding location.
3. The saliency map was modified to import the target information from Distract-R and to estimate search time for the defined target.
4. The activation field that combines the saliency map and the expectation map was created.

Figure 15 shows a process of generating the activation field with example tasks (e.g., finding hotels and finding restaurants from an interface that spatially segregated information by its source). In the example, the effects of bottom-up features were the same for those two tasks, but spatial expectation produced different results in the final activation filed that guide visual attention (the highest peak will be the location to be attended). To balance between top-down and bottom-up influence, the activation filed was defined with a weight as below.

$$Activation\ field = weight * saliency\ map + (1 - weight) * expectation\ map$$

When the weight equals zero, it represents when the model does not account bottom-up influence or top-down influence is dominant, and when the weight equals one, the model only considers the bottom-up influence excluding top-down influence.

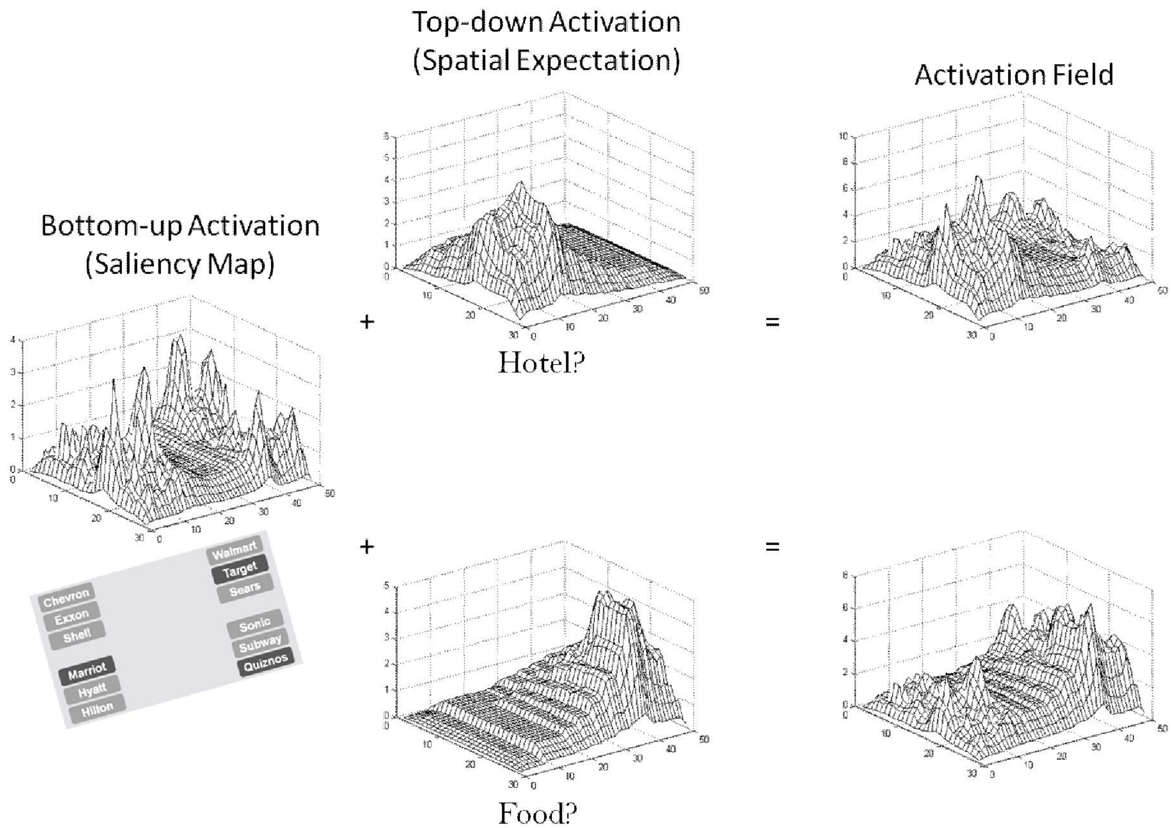


Figure 15. A process of generating the activation field.

4.4.2. Search Module

To simulate drivers' interaction with DVIs, the saliency map was modified to model search tasks by implementing a "search module" (Figure 16). The search module stores information of the predefined target (such as location of the target), compares an attended item (or location) and the target, and decides whether to continue the search task (i.e., when the attended item is not the target) or to terminate (i.e., when the attended item is the target).

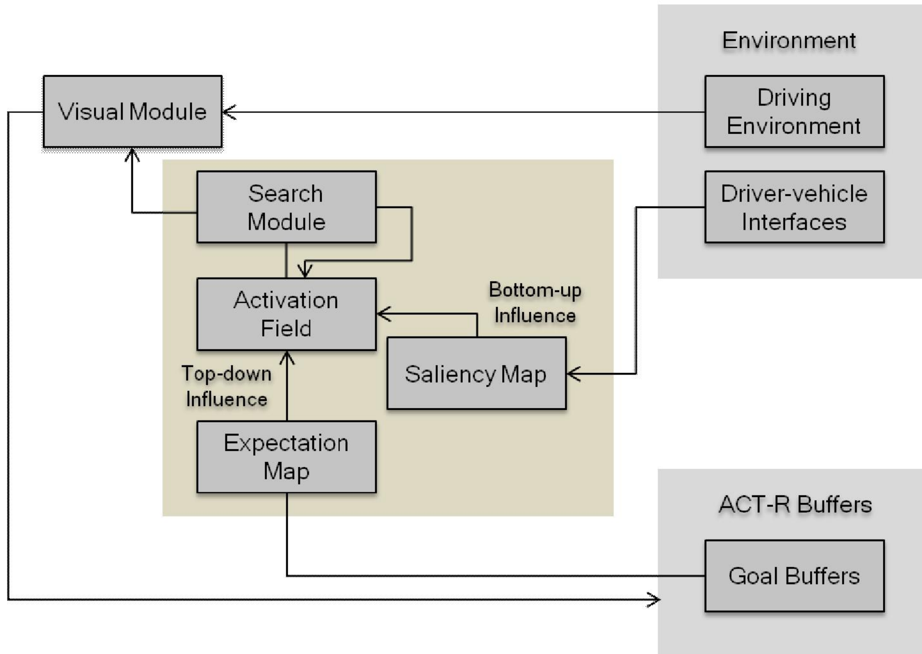


Figure 16. Search module and the modified model.

4.5. Summary

This chapter describes details of the model integration. Through the previous chapters, Distract-R and the saliency map's limitations were defined. The integration of the two models supplemented Distract-R's inability to account for bottom-up attention and implementation of the activation field and expectation map built a basic structure for the tool to evaluate distraction potential. The next two chapters will discuss the test of the integrated model's ability to predict drivers' performance, and the result of this test will be used to tune parameters and modify the integrated model.

5. Investigation of Driver Distraction

5.1. Objectives

The main objectives of this chapter are to (1) test the effects of design features that might influence drivers' performance through top-down and bottom-up attentional processes, and (2) determine a potential factor that might influence glance duration to the secondary task. The rationale for investigating these two aspects is based on an assumption that both design features and other factors related to driving (such as vehicle stability) can affect driving and secondary task performances. The results of this chapter will be used to develop the proposed model.

5.1.1. Design Features

Spatial grouping and highlighting of displays were selected as the factors to test the influence of design features. Spatial grouping refers a way to separate and consistently locate related information in a distinct location on the screen. Current DVIs have a potential to include multiple information sources, and categorizing this information on the interface might reduce search time and decrease distraction potential. Spatially grouped information can guide drivers' top-down attention (e.g., location expectation) and decrease search time. Figure 17 shows an example of a spatially grouped layout from the home screen of the MyFord system, and Figure 18 shows a counter example that did not use spatial grouping. There are four types of information sources (Phone, Navigation, Entertainment, and Climate) in Figure 17, and each information source is located at each corner (e.g., navigation information is always presented on the right-top corner). Therefore, drivers might easily find a target information if they remember the quadrant of the information source.

Figure 17. MyFord home screen, with consistently mapped information.

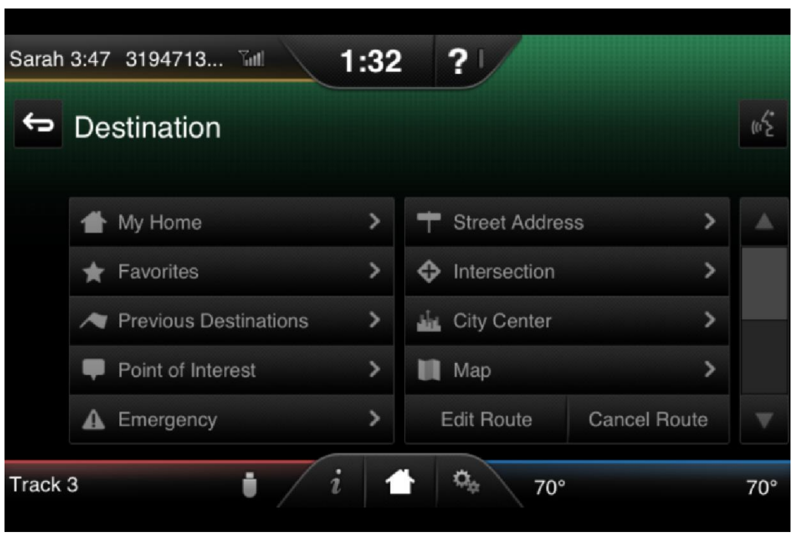


Figure 18. MyFord destination page, with inconsistently mapped information.

Interestingly, spatial grouping and semantic coherence are often confounded in guidelines and studies (Niemelä & Saariluoma, 2003). In this study, it is considered that spatial grouping tied with semantic coherence as an important factor. Specifically, this study focuses on lists of gas stations, hotels, food chains, and shopping centers used as targets and distractors. Figure 19 shows four combinations of spatial grouping and semantic coherence to present four types of information. The upper-left image in Figure 19 represents a spatially ungrouped and semantically incoherent layout, whereas the bottom-right image represents a spatially grouped and semantically coherent layout. The experiment in this chapter compared only these two cases, ‘spatially grouped and semantic coherent’ and ‘spatially ungrouped and semantic incoherent’ layouts. The reason for excluding ‘spatially grouped and semantic incoherent’ layouts was that this case is not likely to be designed in real applications. The ‘spatially ungrouped and semantically coherent’ layout was also excluded because this layout is spatially grouped in a rough way, and the effect of spatial distance from each semantic group is not the focus of this study. Therefore, ‘spatially grouped and semantic coherent’ and ‘spatially ungrouped and semantic incoherent’ conditions were only considered in the present study, and simply renamed as ‘spatially grouped’ and ‘spatially ungrouped’.

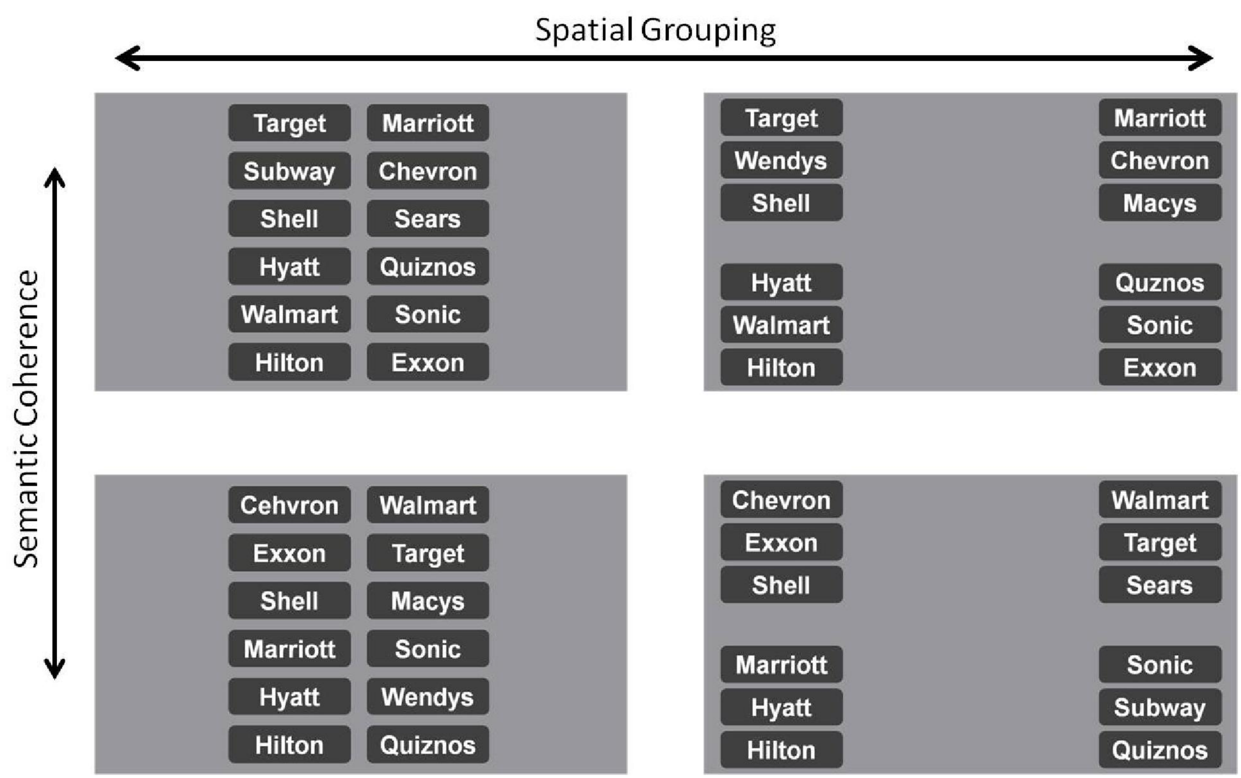


Figure 19. Examples of spatial grouping and semantic coherence.

Another factor that can affect visual search for in-vehicle displays is visual highlighting of items. Highlighting refers to changing the contrast and color to draw attention to locations on the screen. Highlighting can be used to give priority to information in situations where uncertainty makes filtering likely to eliminate important information. Because future vehicles are likely to have multiple sources of information that are not easily filtered, highlighting important information (based on the situation) can be a good application to test with the model. Figure 20 shows an example from a Macintosh operating system of highlighting needed information by manipulating contrast and color. A similar technique (e.g., highlighting specific information by changing contrast and color) was used in this study to assess saliency map's ability to account for bottom-up influences.

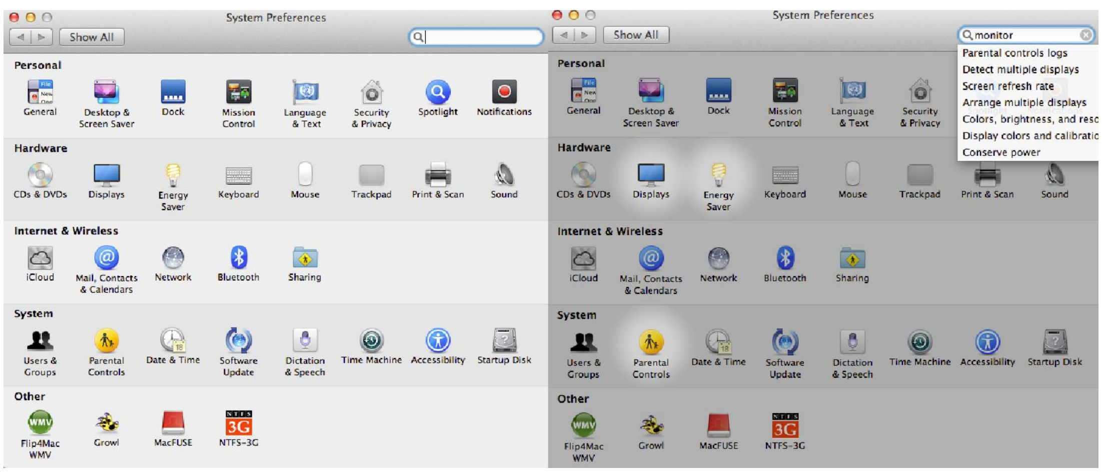


Figure 20. Highlighting function in MacOS.

A study was designed to test the two design features of information layout and highlighting. The layout factor included two levels: (1) a spatially grouped layout, and (2) a spatially ungrouped layout. It was expected that the spatially grouped layout promotes spatial expectation (which would increase activation level in the integrated model, and this would lead to faster search times), whereas the spatially ungrouped layout does not promote spatial expectation or a learning effect. The highlighting factor has three levels: (1) no highlight, (2) correct highlight (three out of 12 items are highlighted, and the target is one of the three highlighted), and (3) incorrect highlight (three out of 12 items are highlighted, and the target is not one of the three highlighted). If color highlights relevant information, this will lead to a high activation level for the highlighted item in the integrated model and will decrease search time for the target.

Spatial grouping and highlighting represent two important design features that could substantially affect driver performance, but are not well addressed by current guidelines. The experiment tested top-down influence (e.g., location expectation) on visual search task and driving performance. The outcome of the experiment was used to investigate the effect of design features on drivers' performance.

5.1.2. Predicting Long-Duration Glances

Drivers are adaptive, so they can actively manage their glance duration on the secondary task based on driving situations. Senders and his colleagues (1967) introduced the "uncertainty model" to help explain this adaptive process. In the model, the "uncertainty" refers to one's own vehicle position and the road environment (e.g., other vehicle's presence and potential obstacles). Figure 21 shows a model concept. For example, the amount of information about the road (e.g., other vehicles presence) reaches the maximum when the driver gazes to the roadway (e.g., the "vision" period in the figure) for enough time, and gradually decreases while glancing away from the roadway (e.g., the "occlusion" period in the figure). The uncertainty is equal to the amount of information that has been decreased (or lost) in this example, so when the uncertainty level reaches the maximum level (e.g., uncertainty threshold) that the driver is willing to accept, the driver requires visual sampling for the roadway again to acquire road information and decrease the uncertainty level. This visual sampling process and uncertainty threshold are closely related to factors such as vehicle dynamics, vehicle stability, and speed, along with the driving environment (e.g., road width, traffic, or headway). Senders and his colleagues described “if the vehicle is very stable, it does not need to be attended to as often as if it were very unstable; and if the uncertainty of steering is small, the driver does not have to look as often as he would if it were large” (Senders et al., 1967, p. 16).

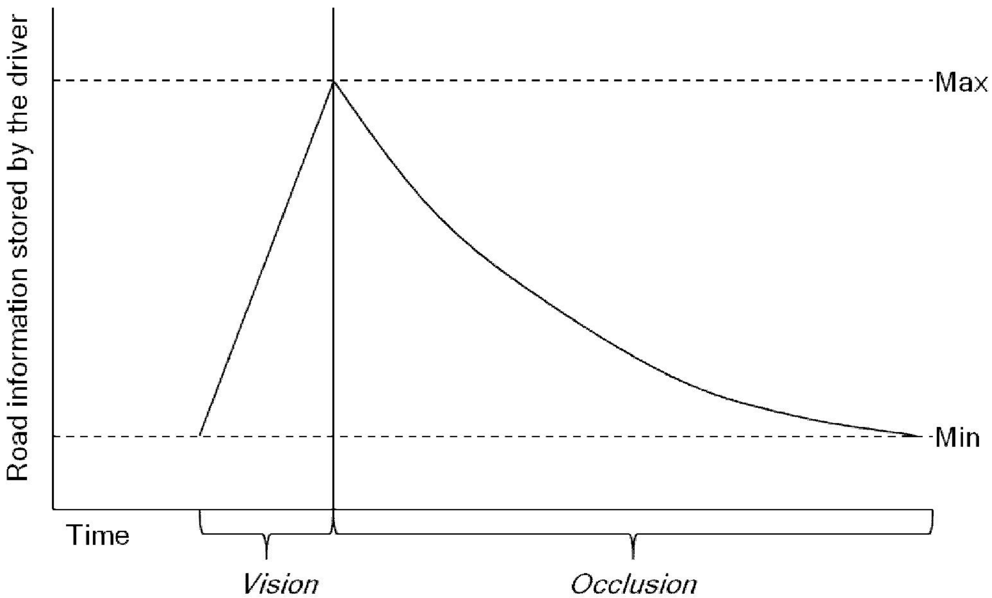


Figure 21. Senders' model (adapted by Senders, Kristofferson, Levison, Dietrich, & Ward, 1967).

Godthelp, Milgram, and Blaauw (1984) found a close correspondence between time-to-line-crossing (sometimes referred to as time-to-lane-crossing) and drivers’ self-chosen occlusion times. In the self-chosen occlusion (or self-paced occlusion) study, drivers decide duration of occlusion. This self-paced visual sampling is similar to multitasking situation in real driving. Chen and Milgram (2013) used the self-paced visual occlusion and revealed that the self-paced occlusion time which is equivalent to the glance duration on the secondary task in this study, could be predicted by lane deviation at the end of occlusion intervals. Drivers might rely on the vehicle's position at the last visual sampling to determine glance duration off the roadway. For example, if the vehicle's position at the last visual sampling was near the center, the driver might allow more time on the secondary task, whereas if the vehicle was near the lane boundary, the driver might allow less time on the secondary task or might not initiate the secondary task.

This study applied Chen and Milgram (2013)'s finding for driver distraction study with a more naturalistic setup. If lane deviation is correlated to long-duration glances (which are equivalent to self-chosen occlusion time) in the experiment, lane deviation at the last visual sampling can be used to predict glance duration and to enhance the integrated model's prediction given that the integrated model does not have an ability to predict single glance duration.

Both design features and vehicle's lateral stability will be investigated to test their effects on (1) driving performance, (2) total glance time to complete the secondary task, and (3) maximum glance duration. The results will provide a foundation for understanding driver distraction, and for building a computational model that can simulate distracted drivers.

5.1.3. Expected Effects of Design Features

Having one factor that promotes top-down influence and the other factor that promotes bottom-up influence will produce an interaction between two factors. The effect of highlighting was observed from the previous experiment. However, it is unclear whether the effect of highlighting will exist when top-down influence is engaged. In the experiment, the number of information categories was in a range of working-memory capacity, and there were multiple highlighted items for each stimulus. Therefore, it is expected that the influence of bottom-up process (e.g., highlighting) will be only significant when the top-down influence (e.g., layout) is not engaged, and when the top-down influence is dominant (e.g., in a grouped layout), the bottom-up influence will be limited.

5.2. Method

5.2.1. Participants

Participants were required to be in good health, and were screened using a telephone script that included study inclusion/exclusion criteria. The study included 12 participants (7 male and 5 female) ranging in age from 21 to 66 ($M = 41.75$) years old.

5.2.2. Apparatus

Experimental stimuli were presented on a touch screen (800 x 450 pixels). Participants were seated in a MiniSim driving simulator, and the touch screen was placed on a center console box (Figure 22). A faceLAB 5 eye tracking system was used to sample and record eye position at 60 Hz.

Figure 22. Driving environment for the experiment.

5.2.3. Driving Tasks

Participants were instructed to drive at 60 mph in the center lane, and to follow the lead vehicle. Senders and his colleagues (Senders et al., 1967) revealed that occlusion time (i.e., periods of not looking at the road) declines as the maximum velocity increases by controlling driver vision with an occlusion helmet. This study fixed travelling speed for controlling potential "speed compensation" by participants, because the focus of this study is to find effects of design features on glance duration for the search tasks.

When the drive started, they were required to accelerate to 60 mph, and from this point the cruise control was automatically engaged (e.g., participants did not have to control vehicle's speed or distance to the lead vehicle unless the lead vehicle braked). The driving task used a cruise-control system to limit participants' ability to compensate for their distraction (for example, drivers may decrease travelling speed and increase a distance from a lead vehicle when they are distracted by a secondary task). Participants were instructed to tap the brake pedal as quickly as possible when they saw the brake lights from the lead vehicle. After releasing the brake pedal, the cruise control was automatically resumed.

5.2.4. Search Tasks

Participants were instructed to interact with the touch-screen system while driving. Information about two types of layouts and information categories were given beforehand. For each secondary-task trial, participants heard a verbal instruction that indicated which button to press, and were instructed to find and press the correct button as quickly as possible while driving safely.

5.2.5. *Stimuli*

A total of 36 screen images (800 x 450 pixels) were presented on the touch screen. Stimuli were designed according to the independent variables: spatial grouping and highlighting. Figure 23, Figure 24, Figure 25, and Figure 26 show example stimuli.

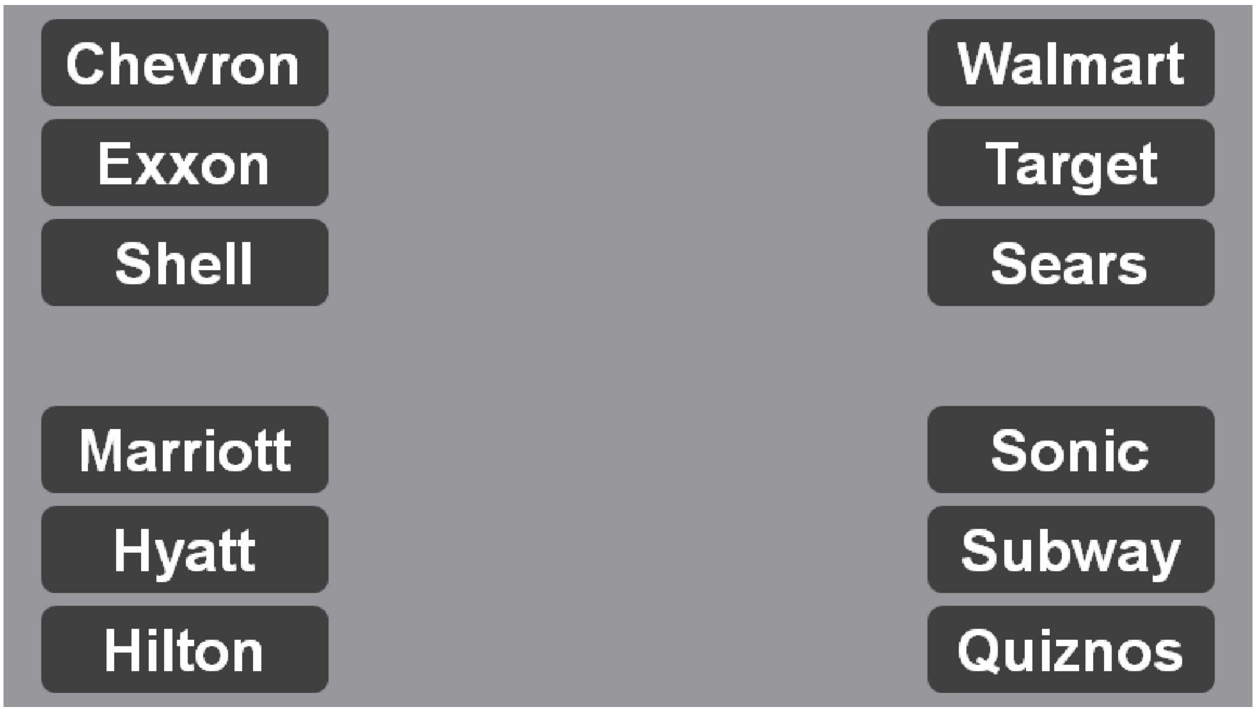


Figure 23. Spatially grouped and not highlighted.

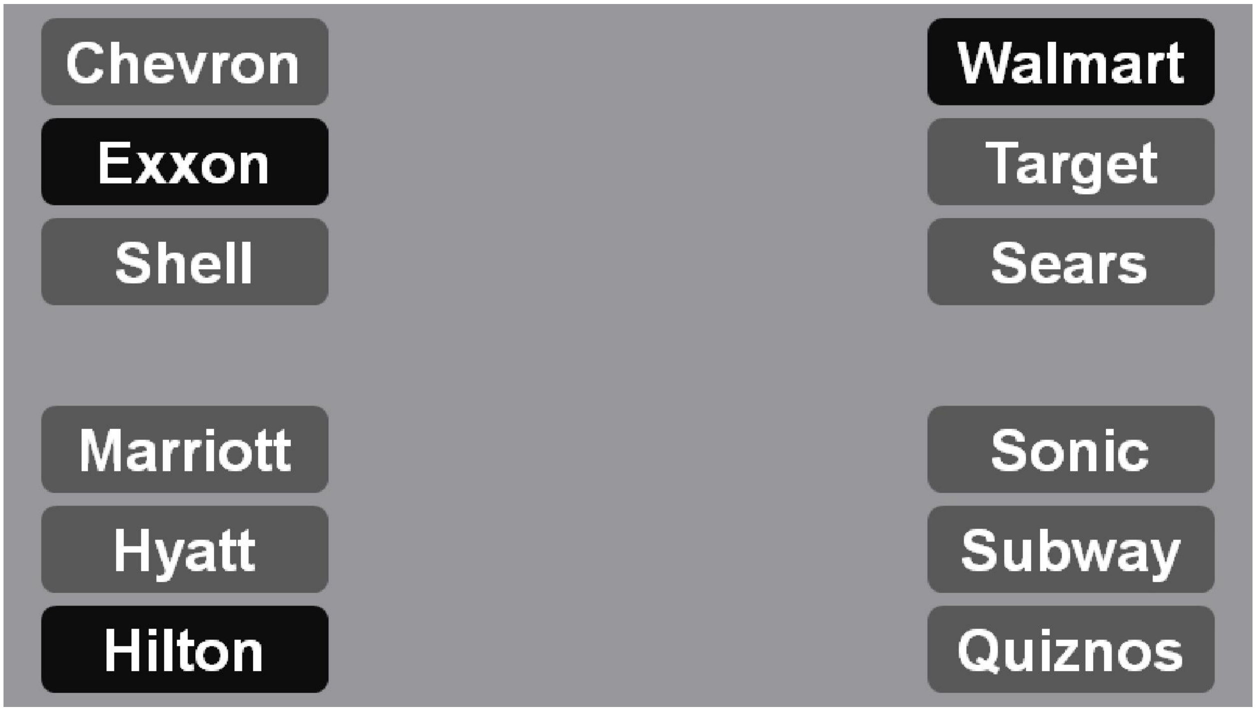


Figure 24. Spatially grouped and with correct and incorrect highlighting.

Figure 25. Spatially ungrouped and not highlighted.

Figure 26. Spatially ungrouped and with correct and incorrect highlighting.

5.2.6. Procedure

Four locations on the interface were used to present information, and four groups of information (e.g., gas stations, hotels, food chains, and shopping centers) were presented. Each participant experienced six conditions (two layouts x three highlight). Each condition consisted of six trials varying targets. Each participant drove for 15 minutes while performing the secondary task, which was triggered by an audio instruction (such as “Walmart”). If drivers did not respond in five seconds, the trial was automatically terminated and a new trial was started. Participants were instructed to interact with the touch-screen system while driving. Information about two types of layouts and information categories were given beforehand. For each secondary-task trial, participants heard a verbal instruction that indicated which button to press, and were instructed to find and press the correct button as quickly as possible while driving safely.

During the experiment, participants responded to two types of braking events, those that co-occurred with the search tasks and those that did not. After driving, each participant completed a post-experiment survey including nine questions that probed their strategy regarding how they performed the search task.

5.2.7. Design

A 2 (spatially grouped layout vs. ungrouped layout) x 3 (no highlight, correct highlight, and incorrect highlight) repeated measures design was implemented. Driving performance (e.g., lateral deviation and brake response time) and eye movement behavior (e.g., maximum glance duration and total glance duration) were measured as dependent variables.

5.2.8. Data Reduction

Alliance of Automobile Manufactures (Driver Focus-Telematics Working Group of Alliance of Automobile Manufactures, 2002) proposed distraction evaluations by evaluating glances to task-related areas such as displays, whereas NHTSA's criteria (National Highway Traffic Safety Administration, 2012) are based on eye glances away from the roadway. For this experiment, the world model (that measured and transferred size and relative distance of the objects in real world to the eye tracking system so the system can identify where an observer gazes) for the eye tracking system does not include any meaningful objects such as a roadway hazard or an oncoming vehicle from the side (it only included a center channel, an instrument panel, and a touch screen). Moreover, this study focused on total glance duration for the search tasks. Therefore, this study followed Alliance's criteria. The rationale for this classification is to calculate more accurate glance duration by enhancing the accuracy of classification for fixated objects.

One way to calculate glance duration is to sum total glance time to the touch screen. The faceLAB system provides a variable that represents fixated objects based on the gaze coordinate. The original model has nine possible objects (e.g., center channel, touch screen instrument panel, etc), and all objects are contiguous. This model requires a high accuracy of gaze tracking, because when glances fall on an adjacent area between two different objects, the glances have the possibility of being classified as either object due to tracking errors. If a glance to the touch screen is classified as a glance to another item, this can dramatically influence the calculation of single glance duration. For example, if a participant actually glances to the touch screen for three seconds in a single glance, but some parts of the glance are classified as other objects adjacent to the touch screen, the actual single glance duration might be split into multiple glances. Especially for distraction studies focused on long single glances, this issue should be carefully considered in interpreting the data. For these reasons, all participants' glance data were manually reviewed and re-classified.

5.3. Results

5.3.1. Driving and Eye Movement Patterns

Prior to the main analysis, this section introduces examples of driving and eye tracking data to help understanding the subsequent analyses. Figure 27 and Figure 28 show examples of the vehicle's lateral position and one participant's eye glances when completing two secondary task trials. As described earlier, for each secondary task trial, participants heard a verbal instruction that indicated which button to press. The "Instruction" period refers to the period from the beginning of the verbal instruction to when participants tapped the screen to indicate they were ready. The "Secondary task" period refers from the beginning of the secondary task to the end of the secondary task. In general, the secondary task periods accompanied eye glances to the touch screen, and sometimes the secondary task required more than two glances to the touch screen.

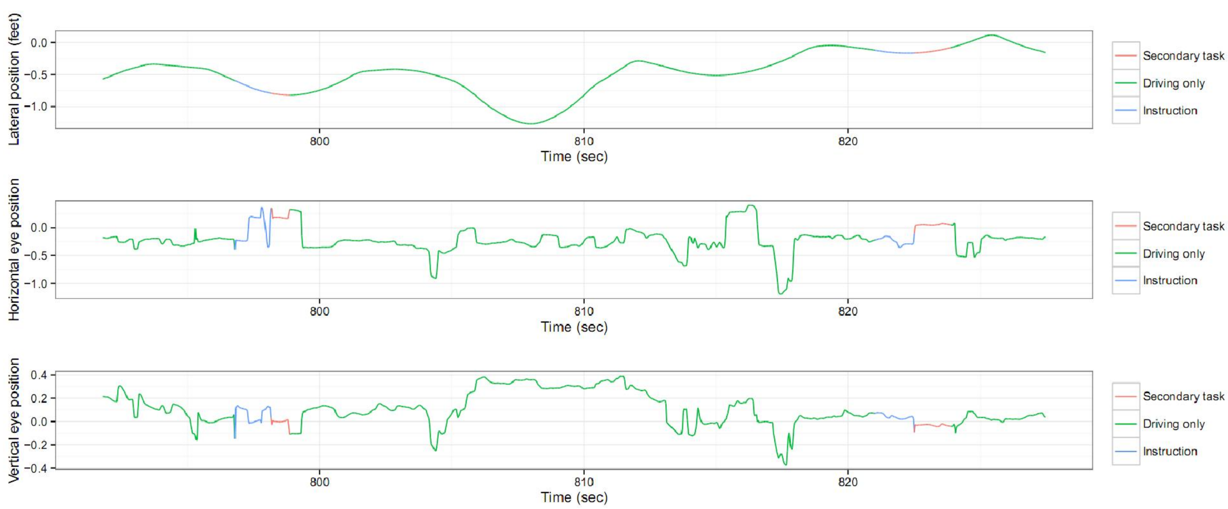


Figure 27. Examples of a driver's eye movements and lane keeping performance across time (participant ID = 2, trial ID = 10 and 11).

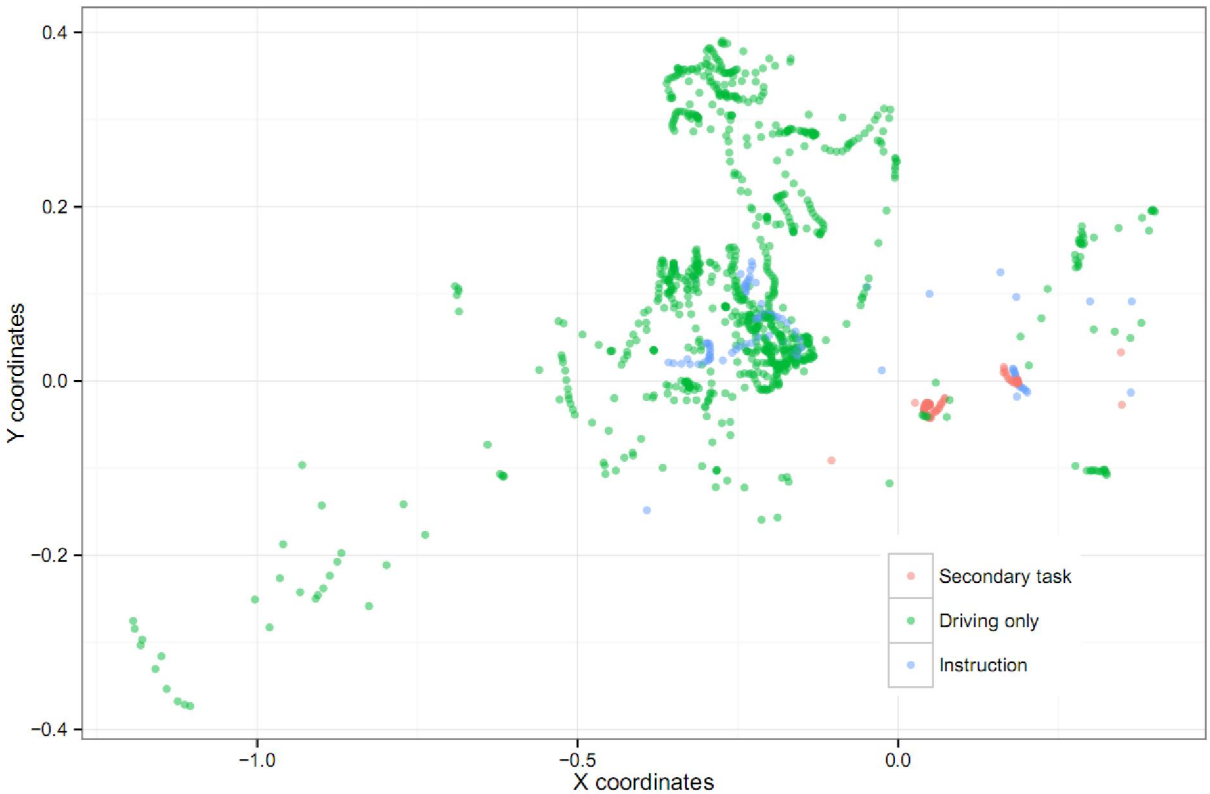


Figure 28. An example of fixation locations corresponding to Figure 28 (participant ID = 2, trial ID = 10 and 11).

Figure 28 shows all fixation locations to complete two secondary task trials. A majority of the fixations from the driving period fell on the area matched to the roadway, and a majority of the fixations from the secondary period fell on the area matched on the touch screen (e.g., the red points on the figure). It was technically impossible to differentiate focal vision and peripheral vision, but participants' gaze patterns suggest that the search task required glances to the touch screen.

5.3.2. Effects of the Secondary Task on Drivers' Response Time for Random Brake Events

It is expected that engaging in the secondary task imposes visual distraction and increases drivers' reaction time to the random braking of the lead vehicle. The mean response time during the driving only period was 1.18 seconds (SD = .14), whereas the mean response time of the driving with the secondary task period was 1.46 seconds (SD = .45). The distribution of response time for the driving with the secondary task showed greater number of long response time (Figure 29). Especially, response time during the driving only period did not exceed two seconds, whereas response time during the driving with the secondary task showed a proportion over two seconds.

There was a significant effect of the secondary task [t (286) = 8.82, p < .001], indicating that engaging in the secondary task increased response time to the random braking events. As the previous section indicated, the search task required visual attention and eye fixations, and this may cause slower reaction time to the braking lead vehicle.

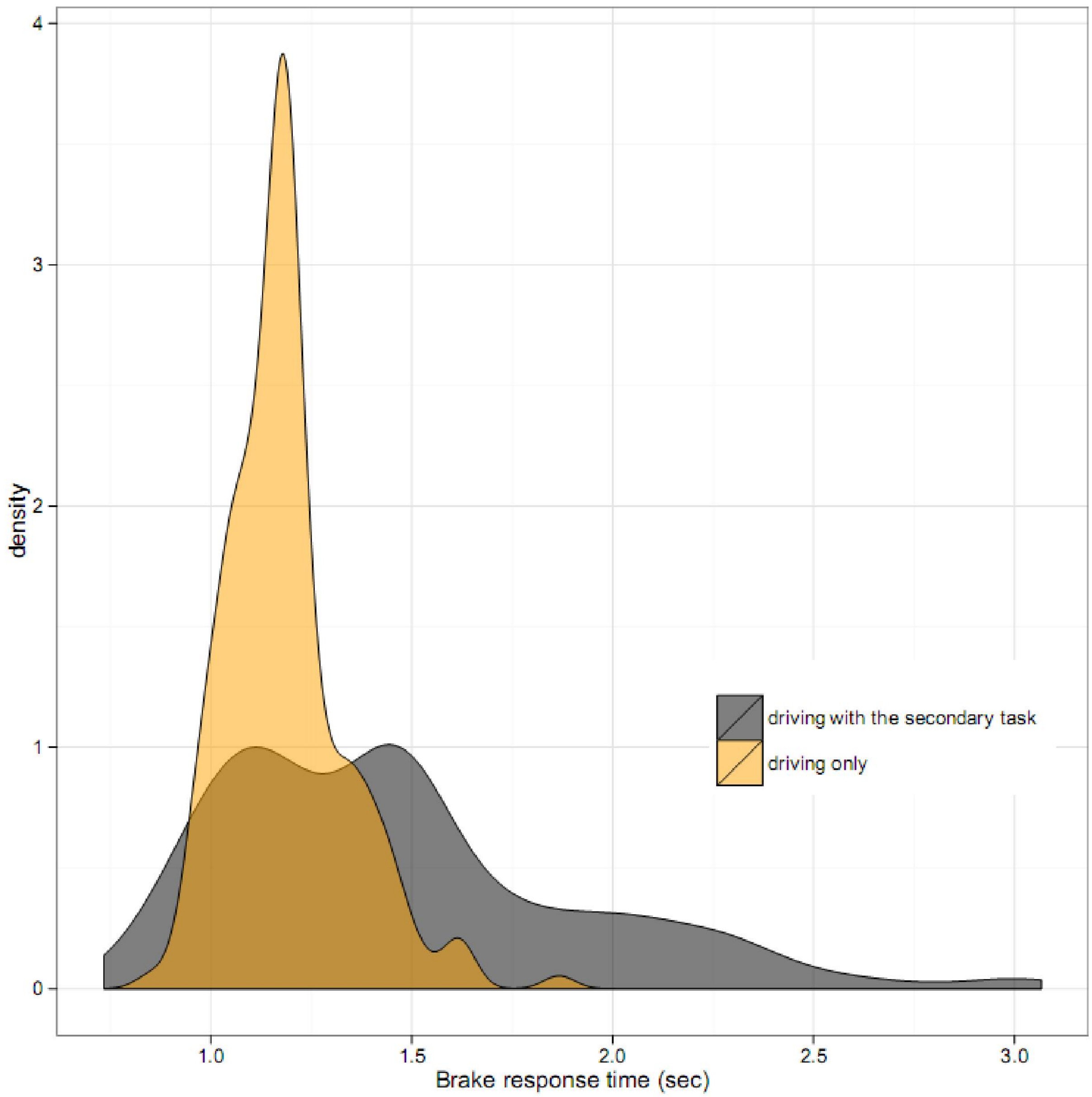


Figure 29. Distributions of brake response time (sec) from "driving only" phase and "driving with the secondary task" phase.

5.3.3. Effects of the Secondary Task on Driving Performance

Two dependent variables were measured for assessing driving performance: (1) lateral deviation—standard deviation of lateral position (feet), and (2) lateral velocity—standard deviation of lateral velocity (feet/second). Each trial consisted of two phases: (1) "driving only" phase and (2) "driving with the secondary task" phases. Figure 30 compares the lateral deviations for two phases and Figure 31 compares the lateral velocities. There was a significant effect of engaging in the secondary task on the standard deviation of lateral position [t (11) = -5.57, p < .001], indicating that engaging in the secondary task increased lateral deviation. There was also a significant effect of engaging in the secondary task on the standard deviation of lateral velocity [t (11) = -3.71, p < .001], indicating that engaging in the secondary task increased standard deviation of lateral velocity. Godthelp and his colleagues (1984) found that open loop control may occur during occlusion period (which is equivalent to the secondary task period in this experiment), and the results from this experiment also support the open and closed loop strategies.

However, there was no significant effect from different levels of highlighting and layout on the driving performance. Engaging in the secondary task decreased driving performance, but the different design features for the secondary task did not lead to any significant differences in driving performance.

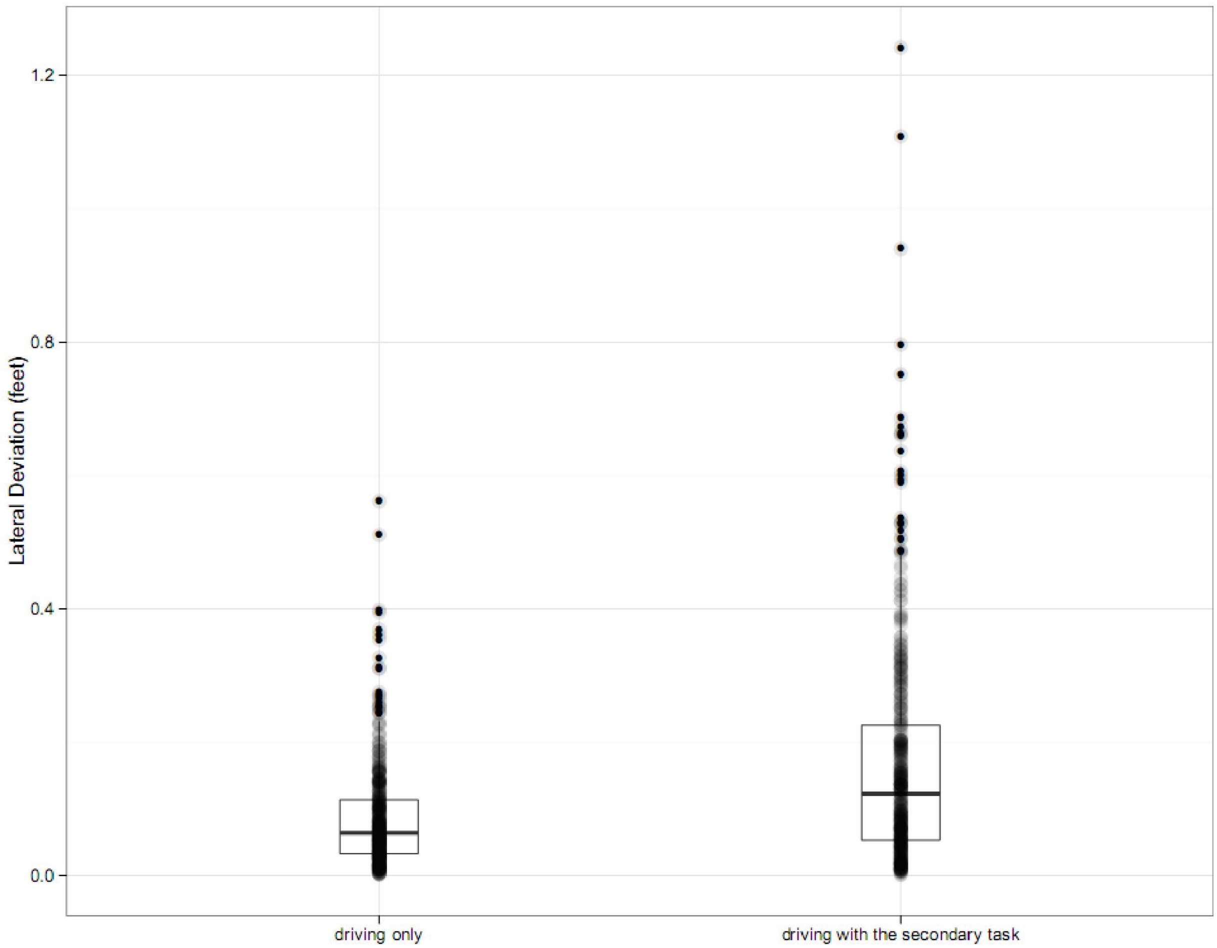


Figure 30. Lateral deviation of “driving only” phase and “driving with the secondary task” phase.

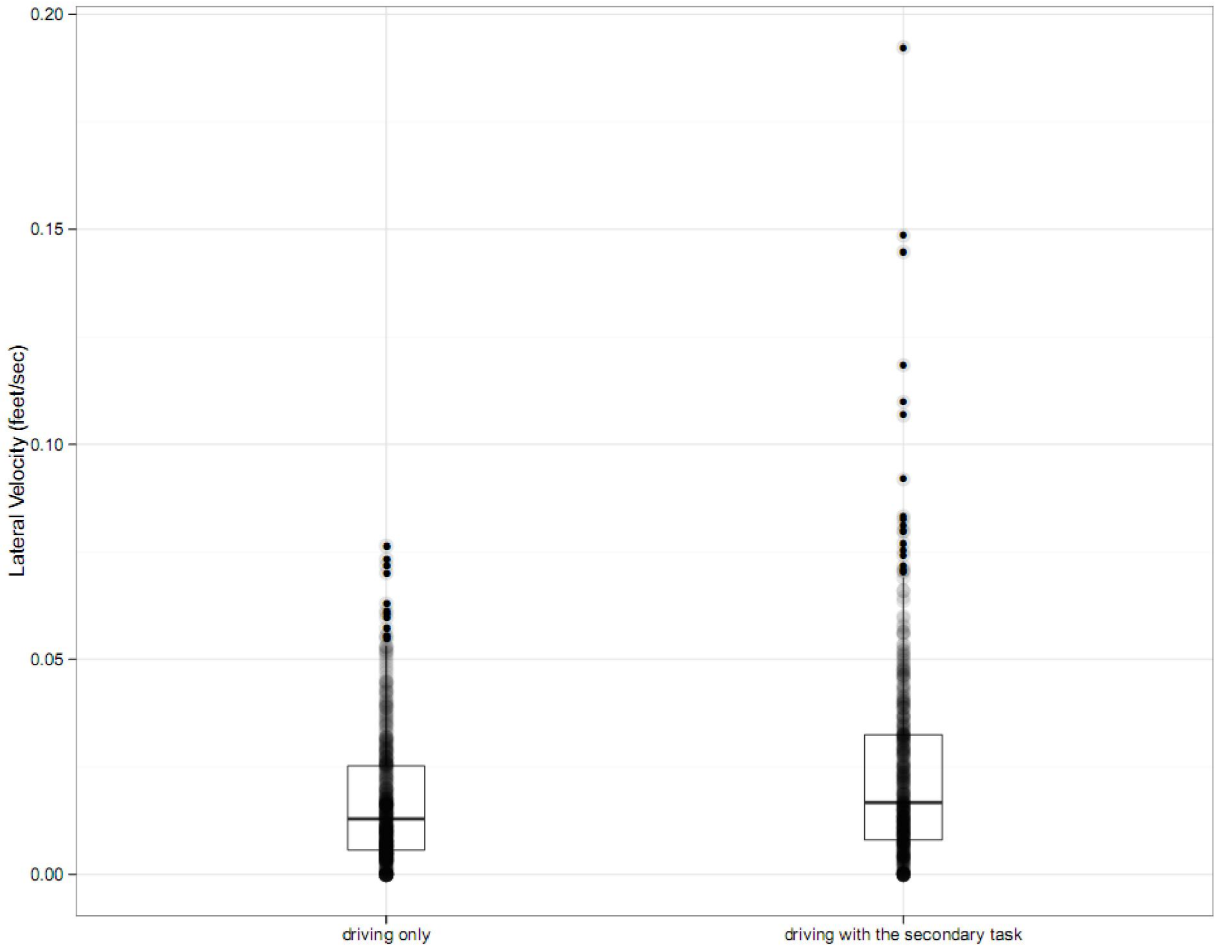


Figure 31. Lateral velocity deviation of “driving only” phase and “driving with the secondary task” phase.

5.3.4. Analysis of Eye Tracking Data

For analyzing eye-tracking data, two measurements are defined (Figure 32). Total glance duration (to complete the secondary task) is a sum of all the glance durations to the touch screen for each secondary-task trial. This measurement reflects the total glance time required to complete each secondary task. Maximum glance duration is the longest glance within each trial while performing the secondary task. A long single glance away from the roadway can be a particularly dangerous, because it increases uncertainty and the likelihood that drivers miss critical information or event on the roadway.

Figure 32 shows one participant's horizontal fixation coordinates during a task trial. Both horizontal and vertical fixation coordinates were used to classify glance locations, but in this example, only horizontal coordinates are presented for simplicity. In this example, glances over point 0.0 (usually between 0.0 and 0.2) are classified as glances to the touch screen. The figure on the top shows task completion time, and the shaded region indicates the beginning to the end of the secondary task. To complete the secondary task, this participant required three glances to the touch screen, and the figure in the middle represents the three glances (the shaded regions) and the switching between the primary task and the secondary task. The figure on the bottom shows the longest glance (the shaded region) of the three glances.

The subsequent section reports the percentage of glances away from the road, which is calculated by,

$$Percentage\ of\ glances\ away\ from\ the\ roadway = \frac{Total\ glance\ time}{Task\ completion\ time} \times 100$$

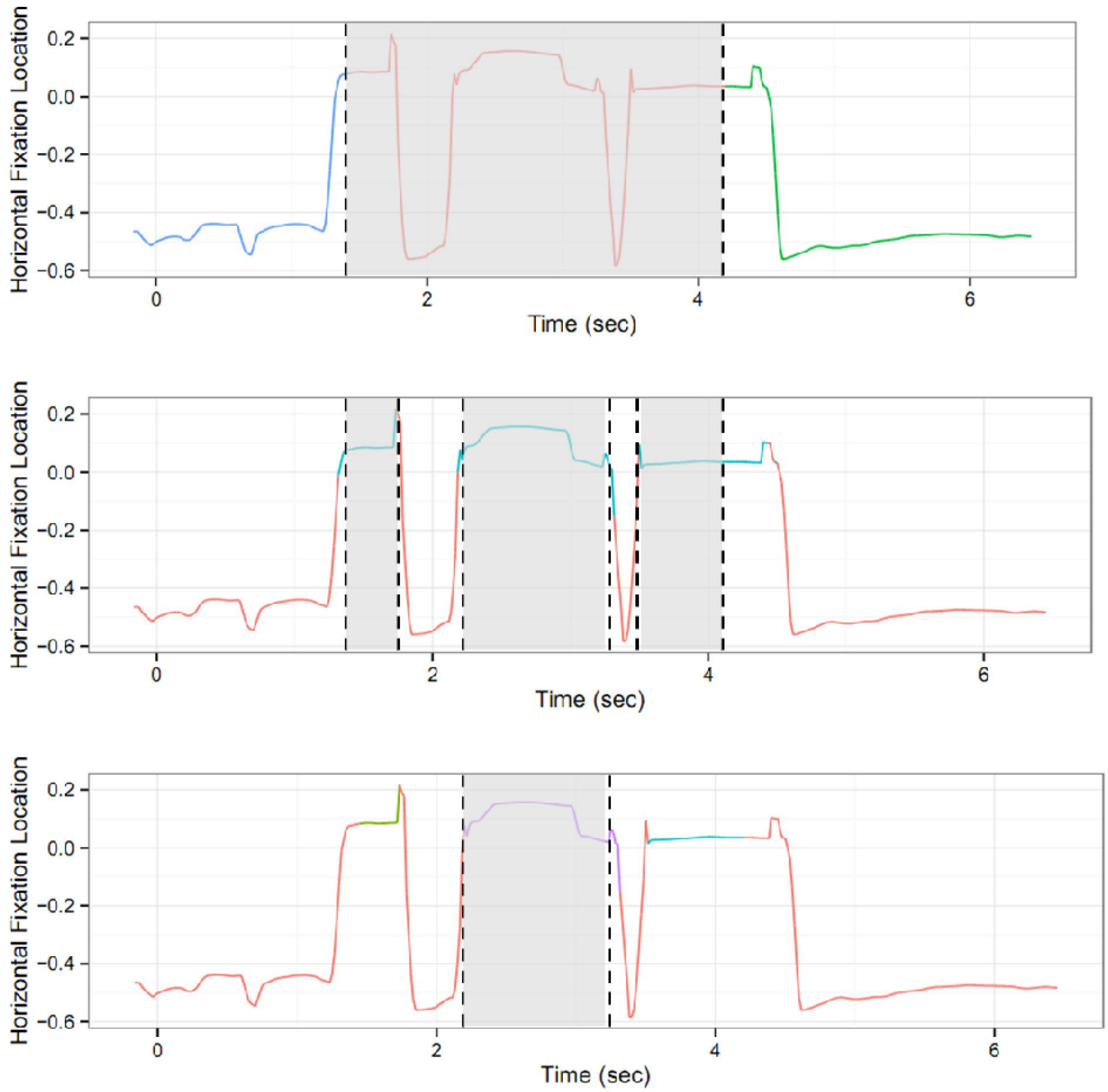


Figure 32. A comparison among Task completion time (top), Total glance time (middle), and Maximum glance duration (bottom).

Table 1 shows a summary of the eye glance data. The mean for total glance time to the in-vehicle display for each task was 1.24 seconds (*SD* = .55), and the longest total glance time was 3.75 seconds. The mean for the maximum glance duration was 0.90 seconds (*SD* = .37), and the longest single glance was 2.32 seconds. The mean percentage of glances away from the roadway was 59% (*SD* = 19.33), maximum value was 100% (i.e., when the secondary task was completed by a single glance away from the roadway), and minimum value was 12.20%.

Table 1. Summary of eye tracking measurements.

	Mean	Standard Deviation	Maximum Value
Total glance time (sec)	1.24	0.55	3.75
Maximum glance time (sec)	0.90	0.37	2.32
Percentage of glances away from the roadway (%)	59	19.33	100 (Minimum = 12. 20)

Repeated measures ANOVA was used to test the effect of highlighting on the eye gaze variables for each layout condition. In the grouped layout condition, there was no significant effect of highlighting as expected. However, in the ungrouped layout condition, there was a significant effect of highlighting on maximum glance duration [F (2, 22) = 5.34, $p < .05$], and total glance time [F (2, 22) = 4.53, $p < .05$].

Post-hoc analysis, using a paired t-test, showed that the incorrect highlighting condition led to longer maximum glance duration than no highlighting condition [t (11) = 4.22, $p < .01$], and longer total glance time [t (11) = 4.39, $p < .01$].

Both a benefit from the correct highlighting and a cost from the incorrect highlighting were expected, but only the cost was observed and there was no difference between correct highlighting and the no highlighting condition (Figure 33 and Figure 34). In the previous experiment, the number of highlighted items was set to one, and this experiment highlighted three items for the correct highlighting condition and the incorrect highlighting conditions. The increased number of highlighted items may reduce the effect of correct highlighting and increase the effect of incorrect highlighting as expected earlier. Previous research (Fisher & Tan, 1989) also expected that when the number of highlighted options is greater than one, highlighting may be worse than no highlighting, as observed from this experiment.

Post-hoc analysis showed a difference between grouped and ungrouped layout for total glance time in the incorrect highlight condition [t (11) = -2.26, p < .05], indicating that drivers' expectation might be particularly influential when the bottom-up features were not reliable (i.e., incorrect).

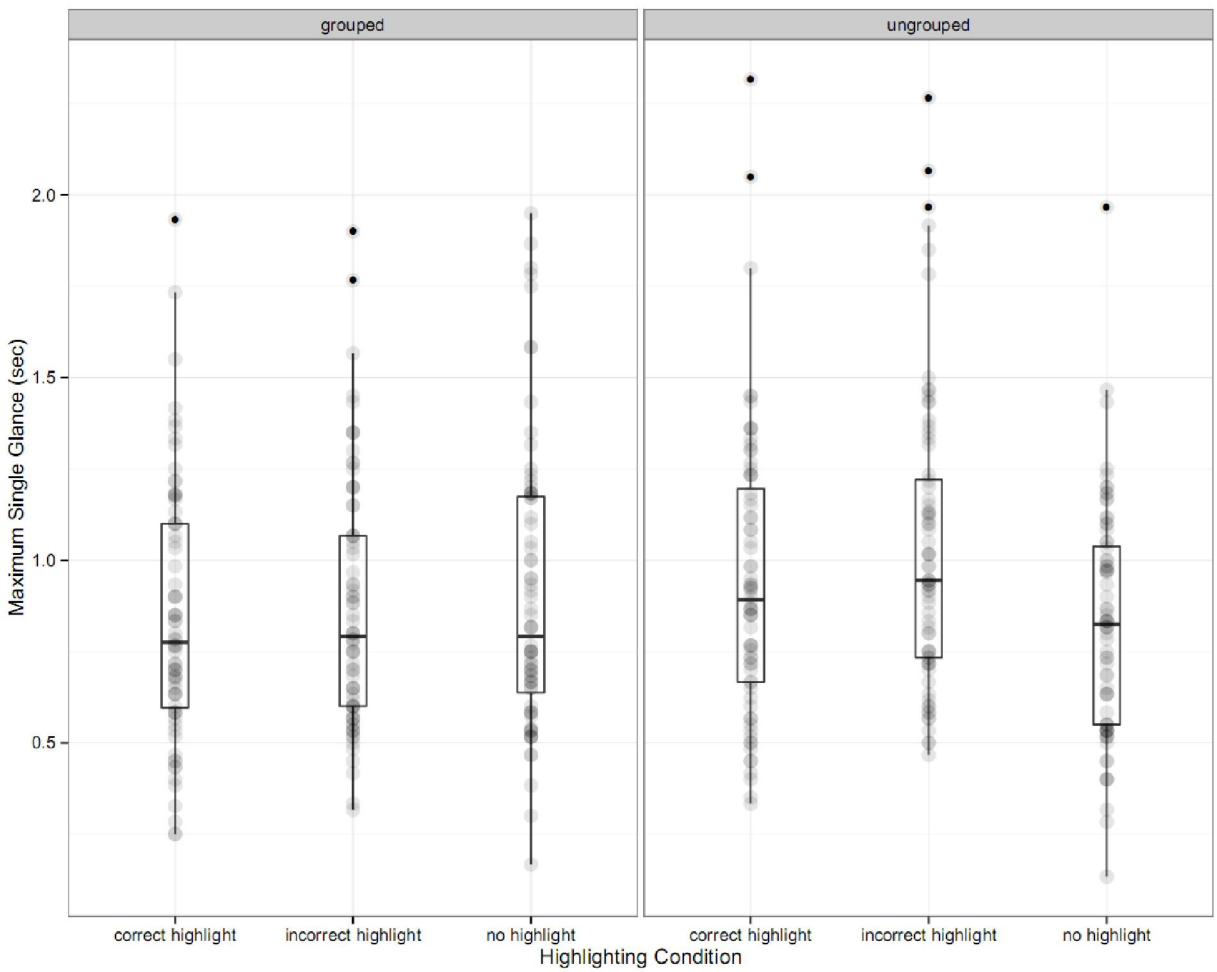


Figure 33. Comparison of maximum single glance duration across all conditions.

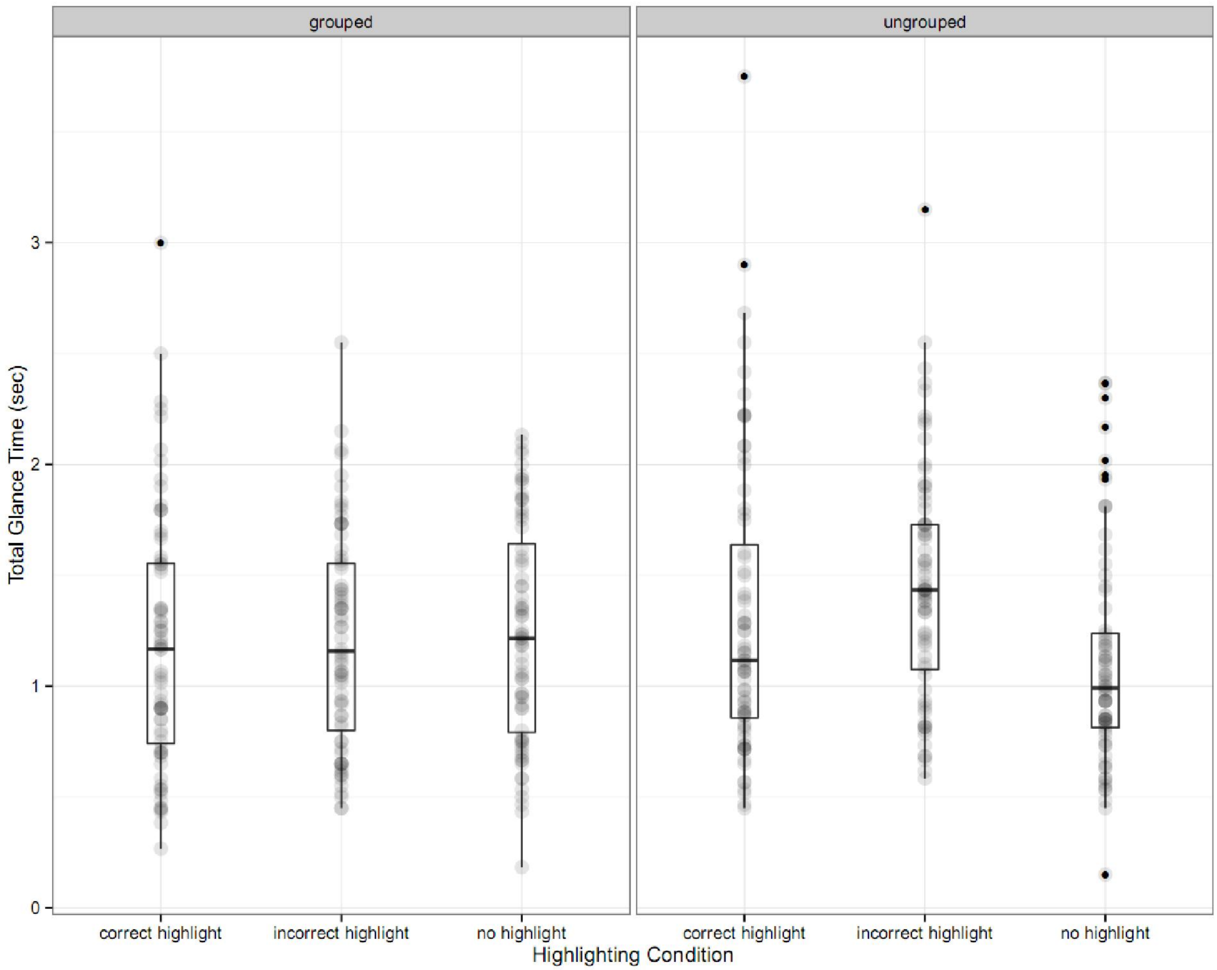


Figure 34. Comparison of total glance time across all conditions.

5.3.5. Analysis of Gaze Patterns

Additional analysis focuses on the relationship among the variables to investigate how drivers allocated their visual attention in a multitasking situation. The introduction of this chapter mentioned that design features or own vehicle stability might affect maximum glance duration. However, it is also reasonable to assume that some drivers tend to complete the secondary task with a few longer glances away from the road and minimize task switching, whereas other drivers tend to switch between the road and the secondary task more frequently with shorter glances to the secondary task. In the experiment, participants completed 47.69% of the secondary tasks with a single glance, and completed 52.31% of the secondary tasks with multiple glances. A correlation analysis indicated the maximum glance duration and percentage of glances away from the road (while performing the secondary task) were positively correlated [r (10) = .67, p < .05], and percentage of glances away from the road and the number of glances were negatively correlated [r (10) = -.81, p < .01]. This may reflect drivers who took shorter maximum glances and glances to the roadway more (i.e., low percentage of glances away from the road) to perform the secondary task, and this led more task switching (i.e., more fixations) (Figure 35).

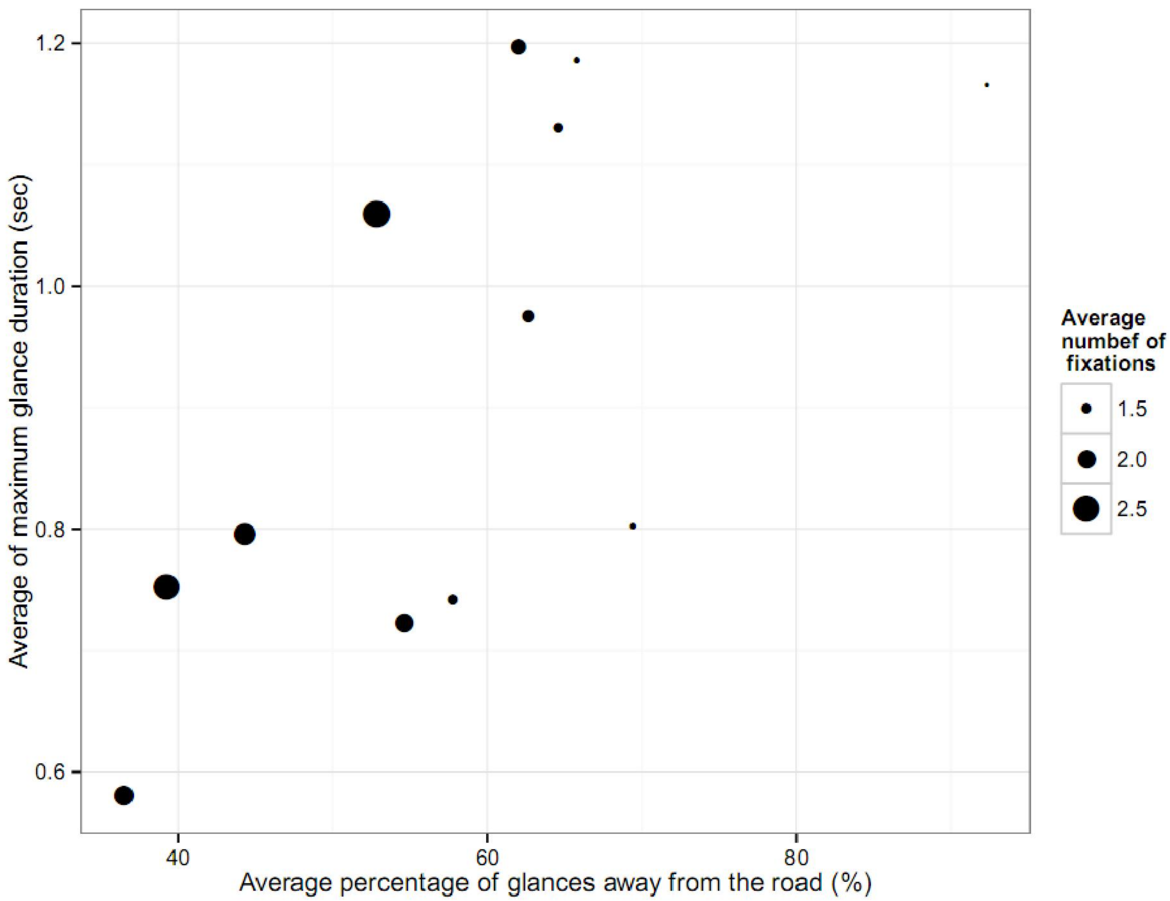


Figure 35. Individual averages of maximum glance duration and percentage of glances away from the road (note: each point represents one participant).

5.3.6. Lane Deviation and Long-Duration Glances

Previous research showed that self-paced occlusion time can be predicted by lane deviation (Chen & Milgram, 2013). Figure 36 shows a relationship between maximum glance duration and lane deviation at the start of the glance. In a self-paced occlusion experiment, drivers tried to maintain the occlusion period as long as possible (Chen & Milgram, 2013). However, in this experiment, drivers' maximum glance duration was determined by task completion time or drivers' strategies to handle the secondary task. For example, if a secondary task required one second-long glance to complete the task, a driver's maximum glance duration for the task might not exceed one second, even the driver was able to glance away from the road for two seconds. However, in the self-paced occlusion study, the driver might try to maintain the occlusion time until the threshold (e.g., two seconds in the example). Therefore, some of maximum glances in Figure 36 might not reflect the maximum capacity allowed by the vehicle's stability. For this reason, the 95th percentile of maximum glance duration was tested to assess the association between lane deviation preceding a long glance and the duration of the glance.

The results of the regression analysis indicated lane deviation at the last visual sample explained 34.4% of the variance [$R^2 = .34$, $F\ (1, 25) = 13.09$, $p < .001$)], and long-duration glances can be predicted by vehicle's lane deviation, similar to the relationship observed in the self-paced occlusion study.

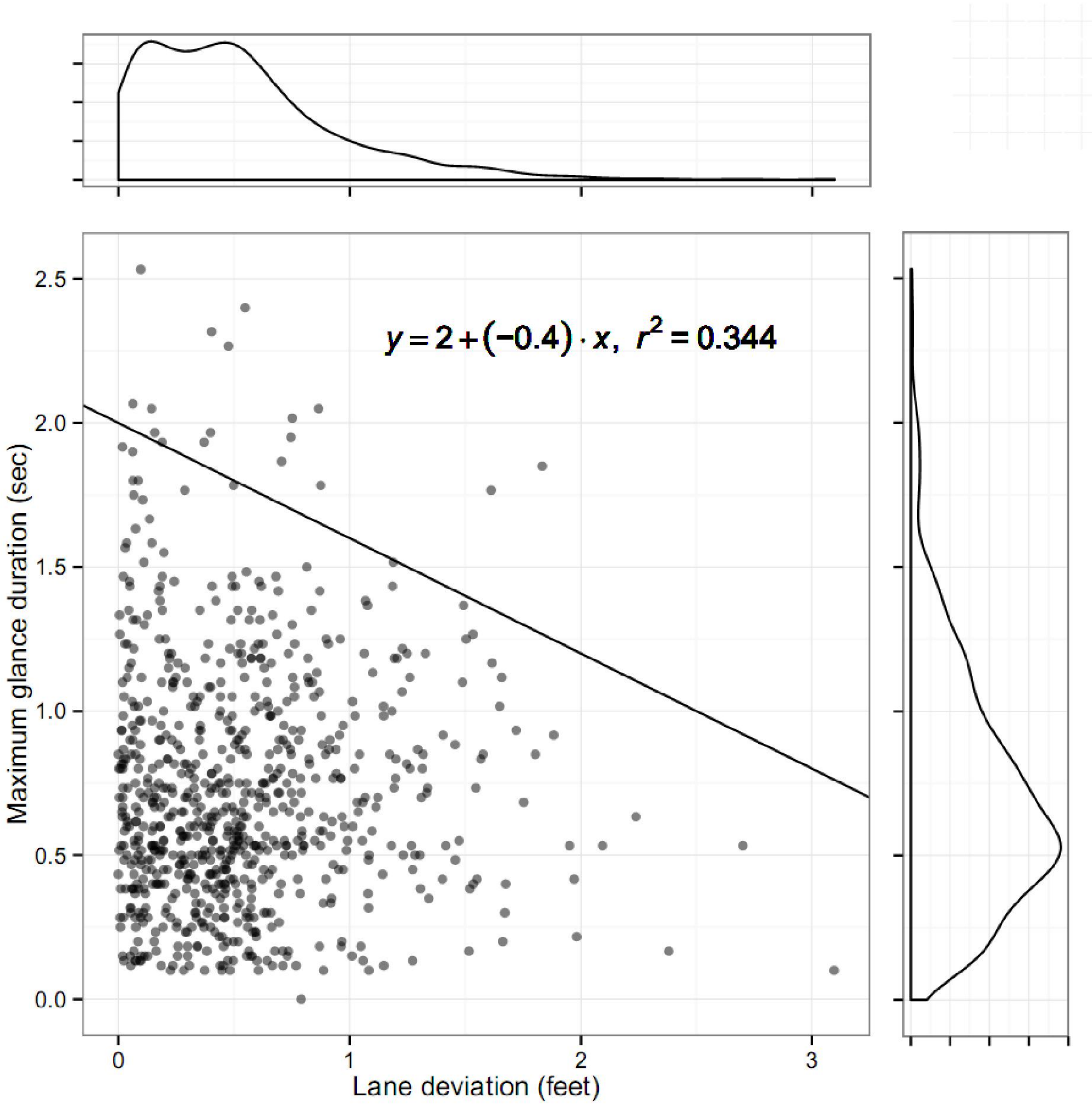


Figure 36. Scatter plot of maximum glance duration and lane deviation.

5.4. Conclusion

This chapter investigated driver distraction with empirical data to assess the interaction between top-down and bottom-up design influences and identify factors that lead to long glances away from the roadway. This study focused on two potential factors that might affect the off-road glances, design features of the secondary task interface and vehicle stability. For the design features, spatial grouping and highlighting were tested. Spatial grouping was expected to promote drivers' top-down attentional process such as expectation, and highlighting was expected to promote drivers' bottom-up process. These two factors can guide drivers' visual attention to important or critical information on the screen based on driving situations by minimizing search time. Therefore, testing these factors can provide in-depth understanding of driver distraction with various combinations of design features.

Senders and his colleagues (1967) explained that drivers have psychological threshold for determining travel velocity or off-road glance duration. More recent studies (e.g., Chen & Milgram, 2013) have shown that drivers' occlusion time was determined their vehicle's lateral deviation from the center of the road. Consistent with these results this chapter also revealed that long-duration glances could be predicted by lane deviation at the start of the glance—drivers tend not to look away from the road for a long period if the vehicle is not near the center of the lane at the start of the glance.

More generally, the results showed that there were a significant effect of engaging in the secondary task on driving performance and a significant interaction of design features on total glance time and maximum glance duration to complete the secondary task. The analyses revealed that drivers relied on both top-down and bottom-up attentional process while performing the secondary tasks. In the grouped layout where participants can easily memorize the location of items, there was no significant effect of highlighting target items. The effect of highlighting was only observed in the ungrouped layout condition where memorizing items' location was hard. In this condition, the cost from the incorrect highlighting conditions was observed, whereas the benefit from the correct highlighting conditions was not significant. The number of highlighted items can be a factor to affect the magnitude of cost and benefit of highlighting (e.g., Fisher & Tan, 1989), and three highlighted items on the correct and incorrect highlighting screens could increase the magnitude of cost in this experiment.

Engaging in the search task affected driving performance, but levels of the independent variables did not lead any significant differences. As mentioned in the introduction of this chapter, it is expected that spatial grouping and highlighting can decrease search time and help search performance. However, the search task (e.g., finding a target item among 12 items) was relatively easy in the experiment, so it might mask the effect of spatial grouping and highlighting. Moreover, this issue also reduced the mean of total glance time and the mean of the maximum glance duration. Especially, the mean of the maximum glance duration was 0.9 seconds and this might not reflect the maximum capacity of glancing away from the roadway, because a driver's maximum glance duration might not exceed task completion time.

Design features showed an interaction effect of top-down and bottom-up attentional processes on total glance time and maximum glance duration. The vehicle's stability predicted subsequent maximum glance duration significantly. Findings from this chapter can be used to tune and the integrated model, and evaluate the model's ability to predict driver distraction.

6. Validation of the Integrated Model

6.1. Objectives

The previous chapter analyzed driving performance and gaze patterns while drivers engaged in a secondary task. A major finding from the experiment was that drivers relied on both top-down and bottom-up attentional processes. The original Distract-R does not have an ability to account for bottom-up influence. Moreover, the original model does not have a function to simulate spatial expectation that was tested in the experiment. However, this dissertation augmented Distract-R by integrating the saliency map and the activation field that can combine both top-down and bottom-up influence, and built the integrated model.

The experiment also found that drivers used multiple glances to complete the secondary task while driving, especially when the secondary task required long completion time. To implement this feature into the integrated model, a regression model to predict long-duration glances by vehicle's lane deviation was developed.

Distract-R simulates multitasking situations and the model driver determines timing to initiate the secondary task based on the vehicle's stability. Senders (1967) introduced a concept of "psychological threshold" in his research. For example, drivers have "psychological limit", and when the vehicle's velocity is over the limit or when glances away from the road are particularly long, they feel uncomfortable or unsafe. The previous chapter also found that vehicle's stability can predict long-duration glances, and this supports drivers might have psychological threshold to determine duration of glances away from the road.

As mentioned above, Distract-R determines timing to engage in the secondary task, but it does not have an ability to switch back to the driving task. The stability in Distract-R is purely a function of the movement of the visual near and far points, and when the model is doing the secondary task, nothing about the driving task is considered. Therefore, in the previous model, the model driver completed the secondary task within in a single glance. It has not been a critical issue because the secondary tasks that were used previously (e.g., Salvucci, 2009), required short completion time or consisted of several simple actions, such as a task that required three key presses , "press 9, press 1, and press 1". The original Distract-R assumes the model driver already knew the location of all components on the interface. Therefore, pressing any button (or icon) on the interface usually takes little time to complete. This study extends the scope of the tasks considered by the model by implementing search tasks that require longer completion times. In this study, the model driver needs to "find" a required icon on the interface, and this extension can contribute to a more realistic driver model.

In the previous chapter, participants completed 47.69% of the secondary tasks with a single glance, and completed 52.31% of the secondary tasks with multiple glances. It is possible to assume that they intentionally break a long glance into multiple short glances, and they might have a time threshold that governs maximum glance duration to the secondary task.

This chapter aims to implement a function to switch back to the primary (e.g., driving) task into the integrated model, and this will enhance the model's ability to manage multitasking. The goal of this implementation is to build a model that can determine maximum glance duration based on vehicle stability from the most recent visual sample, and can actively split a long glance into short glances.

The integrated model has three potential enhancements by implementing: (1) the saliency map to account for effects of stimulus-driven (bottom-up) inputs, (2) spatial expectation to account for effects of goal-driven (top-down) inputs, and (3) task switching module. The model prediction was compared with the empirical data collected for the previous chapter and the original Distract-R.

6.2. Method

6.2.1. Driving Scenarios

Distract-R provides a construction-zone environment (Figure 37) that forces drivers onto a single lane behind a lead vehicle and the same driving environment (Figure 22) was implemented and provided to participants. (Lane width of the road is 3.65 meters and traffic cones are placed every 30 meters) for the previous experiment. The lead vehicle drove at a constant speed at 60 mph, and the model driver has desired traveling velocity as 60 mph.

Figure 37. Distract-R driving environment.

6.2.2. Model Enhancements

Table 2 shows the enhancements of the integrated model on the bottom row with comparison with the original Distract-R in the top row. As mentioned earlier, the original Distract-R does not have an ability to account for any bottom-up influence, and it does not simulate drivers' expectation. Along with these, a pattern to break a long glance into several short glances was also observed, and the integrated model was modified to generate realistic glance sequences to estimate the length and frequency of long-duration glances.

Table 2. Comparison between the original Distract-R and the integrated model.

	Ability to account for bottom-up features	**Ability to account for top-down features**	**Ability to determine timing to switch back to driving**
Distract-R	Does not have	Does not have	Does not have
Integrated Model	Implemented	Implemented	Implemented

It was revealed that drivers mainly relied on their spatial expectation when items are spatially categorized. It was expected that the enhancements listed above would improve the integrated model's prediction and could assess the effect of design features included in the experiment in a similar manner.

6.2.3. Model Training and Upper-theoretic Boundary

To test the integrated model's enhancement, two-fold validation was applied, where each fold was composed of a random sample stratified on the drivers. The rationale for choosing two-fold validation from other candidates were it provides both large test and training sets, and it enables to quantify the upper-theoretic boundary applied in the previous experiment by comparison between two data sets. The data set was stratified by randomly assigning each participant's data (including all trials) to either test or training set. Stratifying based on trials rather than drivers is also possible, but stratifying by drivers provides more generalizability of results. Therefore, half of the participants' data was randomly assigned to a set for training of the model, and the other half of the participants' data was used to test the model's predictions. The upper-theoretic boundary was calculated by predicting the test set's performance from the training set, and it provides an upper limit of the model's predictive capacity.

6.3. Results

This section compares the integrated model's predictions to the empirical data and to the original Distract-R predictions. The results consist of two parts: (1) testing the integrated model's predictions of the driving performance, and (2) testing the integrated model's predictions for the secondary task performance. The focus is more on the second part because most of the model modifications focused predicting secondary-task performance. The previous experiment showed that the design features affected to total glance time, and this is what the original Distract-R could not predict well. Therefore, it is expected that the integrated model that includes the saliency map could account the effect of the design features observed in the empirical data. It is also observed that drivers actively determine single glance duration to the secondary task based on vehicle's lateral deviation from the center. The integrated model was modified to break long-duration glances based on vehicle's lateral deviation.

6.3.1. Effects of the Secondary Task on Driving Performance

Lateral deviation (standard deviation of lateral positions) and lateral velocity (standard deviation of lateral velocity) were compared to test the integrated model's ability to simulate distracted drivers' driving performance. Lateral deviations for two phases (e.g., "driving only" and "driving with the secondary task") were compared in Figure 38, and standard deviation of lateral velocities for two phases were compared in Figure 39. There was a significant effect of engaging in the secondary task on lateral deviation [t (5) = -21.89, p < .001], indicating that engaging in the secondary task increased lateral deviation. There was a significant effect of engaging in the secondary task on lateral velocity [t (5) = -11.37, p < .001], indicating that engaging in the secondary task increased standard deviation of lateral velocity. This pattern matched that observed in the experiment.

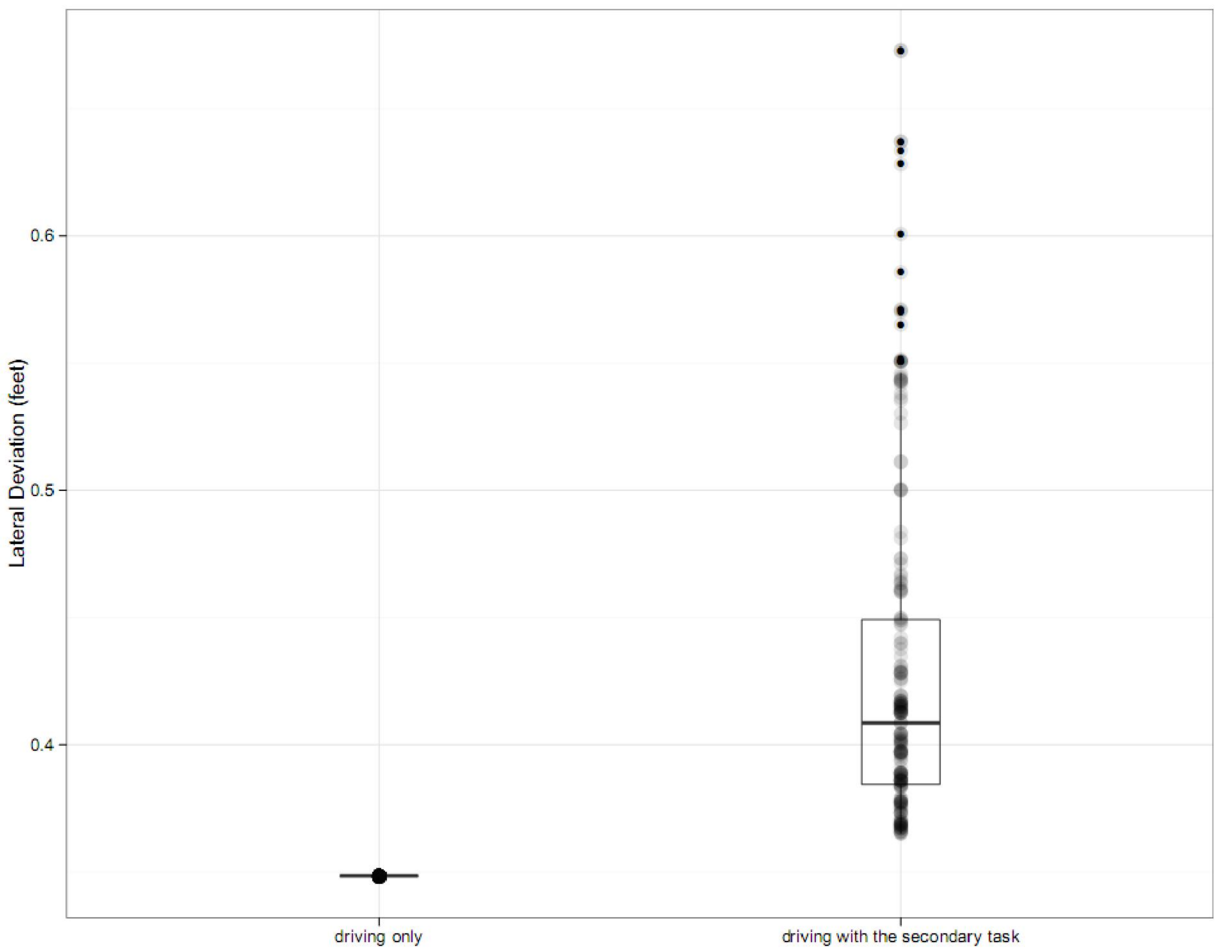


Figure 38. Predicted lateral deviation of "driving only" phase and "driving with the secondary task" phase.

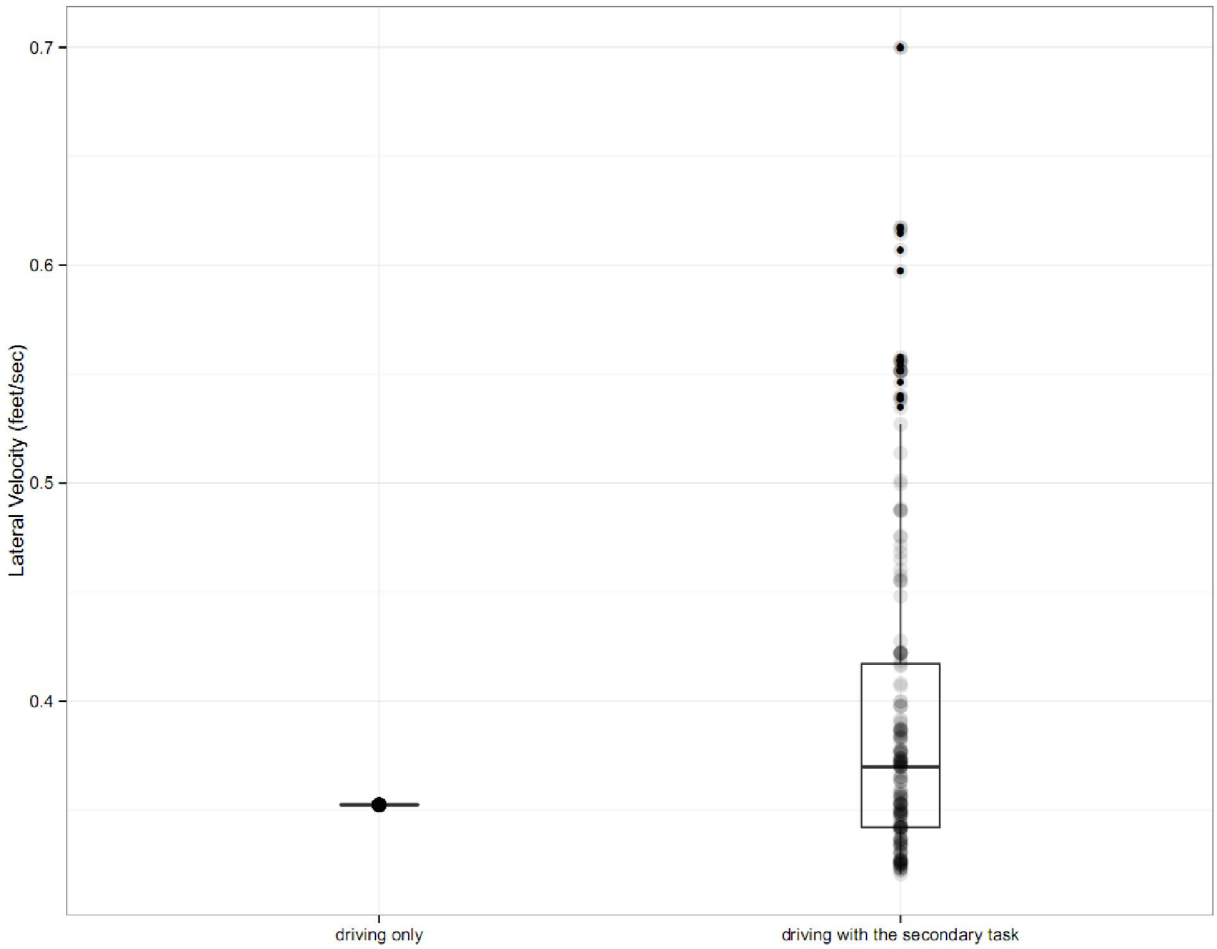


Figure 39. Predicted lateral velocity deviation of "driving only" phase and "driving with the secondary task" phase.

The model predicts the effect of secondary task on driving performance similar to the one observed in the simulator experiment. However, there was a difference between the model prediction and the data. In the experiment, there was a significant effect of the secondary task, but there were no significant differences across all levels of design features. However, the model showed significant main effects of highlighting [F (2, 10) = 6.80, $p < .05$], and layout [F (1, 5) = 89.35, $p < .001$]. Figure 40 compares human performances (upper plots) and model predictions (lower plots). Differences might represent a possible limitation of the driver model in the integrated model. Human drivers may use peripheral vision to sample the roadway or they might use tactile cues from the steering wheel to maintain vehicle's stability while interacting with the secondary task. However, the model driver does not have an ability to adjust steering wheel without vision, and it might cause the model to be more sensitive to the effect of the secondary task.

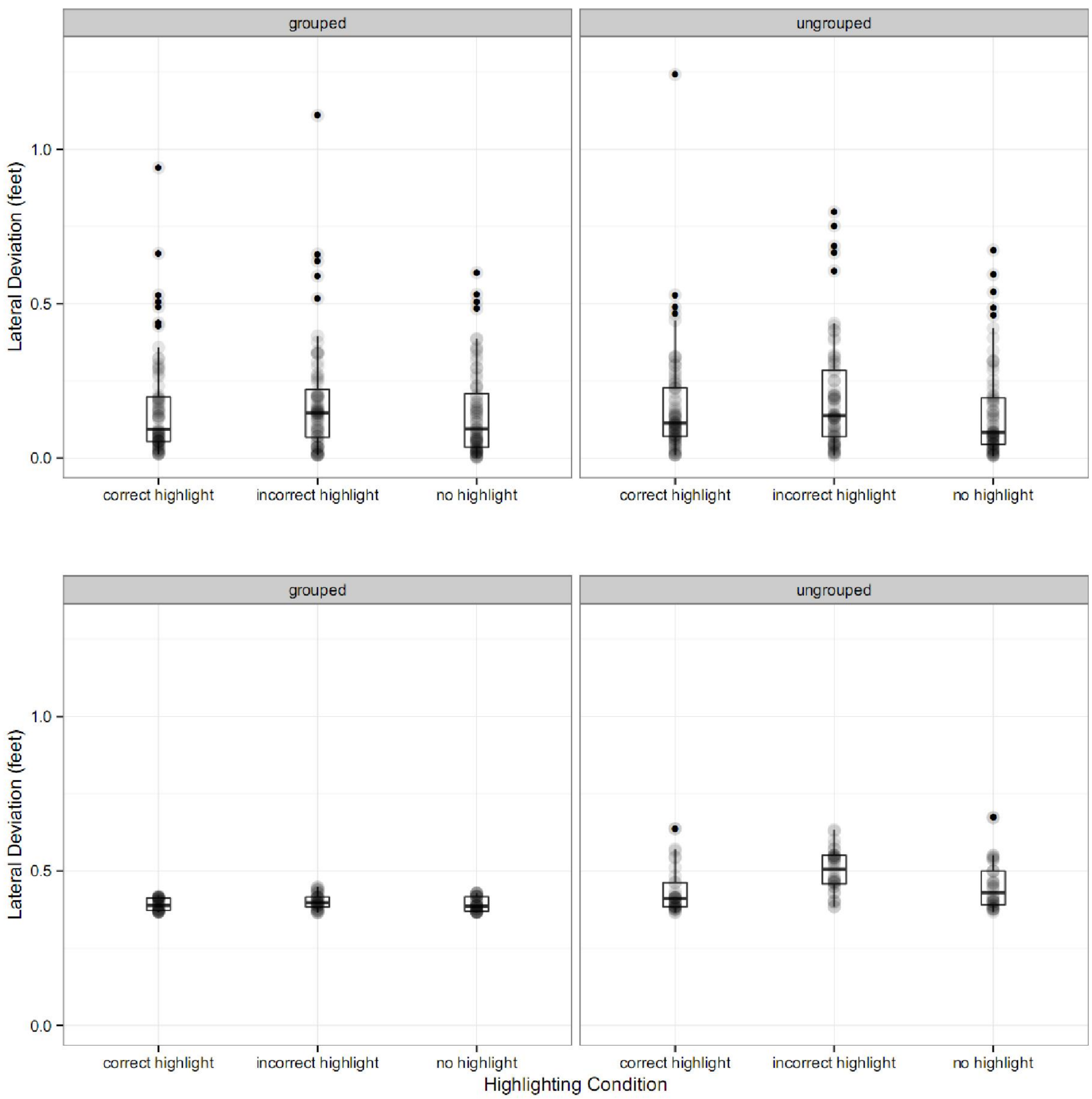


Figure 40. Comparison between human drivers' performance (top) and model prediction (bottom).

6.3.2. Comparison of Total Glance Time

A two-sample Kolmogorov-Smirnov test was compared the cumulative distribution of the empirical data and the model prediction. The D value (e.g., maximum distance in the cumulative distributions) between two empirical distributions (e.g., test set and training set) was 0.09 ($p = .37$) indicating that the two groups were sampled from populations with a similar distribution. After finding the upper-theoretical boundary, the empirical data and the original Distract-R's prediction were compared. D value between the empirical data and the original Distract-R's prediction was 0.23 ($p < .001$) indicating different distributions. However, comparison between the integrated model's prediction and the empirical data showed that the integrated model's prediction was closer to the empirical data and was not significantly different from the empirical data ($D = 0.11$, $p = .14$), indicating that the integration of the saliency map enhanced the predictions of total glance time (Figure 41).

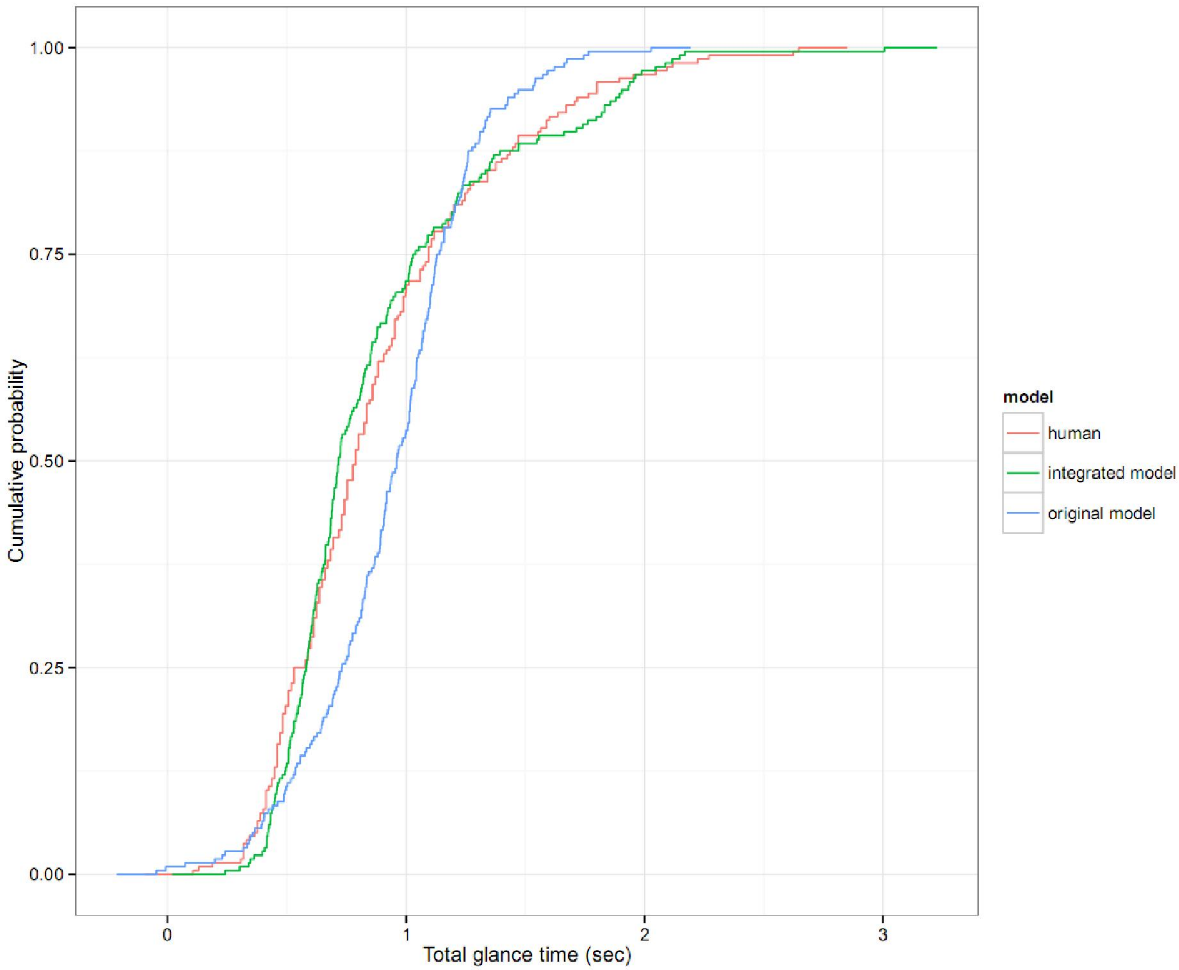


Figure 41: Comparison of total glance time between the empirical data and model predictions.

Previous research (Salvucci, 2009) calculated root-mean-squared error (RMSE) to measure a difference between model predictions and actual observations. The same technique was applied here. RMSE between two empirical data sets was 0.62. RMSE between the data from the experiment and the integrated model was 0.62, which is close to the between-subjects error.

The integrated model's ability to predict the effect of the design features was tested. The effect of highlighting was only significant when the items were randomly located on the screen (e.g., ungrouped layout), whereas the effect of highlighting was not significant when the items were spatially categorized (e.g., grouped layout).

The predictions of the original Distract-R were analyzed and the results showed no significant effects of highlighting across the two layouts (e.g., grouped layout and ungrouped layout), whereas the integrated model showed a similar result to that observed in the empirical data. There was a significant effect of highlighting for the ungrouped layout [F (2, 10) = 96.19, p < .001], but the effect of highlighting was not significant in the grouped layout (Figure 42 and Figure 43). One difference between the empirical data and the integrated model's prediction was that there was a benefit for the highlighting for ungrouped layout, which was not observed in the empirical data. Correct highlighting led to significantly shorter total glance time than no highlighting condition [t (36) = -3.95, p < .001], and no highlighting led shorter total glance time than incorrect highlighting [t (36) = -4.36, p < .01]. In the ungrouped layout, the integrated model only relied on the bottom-up influence, because the only available top-down influence was spatial expectation. However, in the experiment, the participants might use other top-down factors not included in the integrated model.

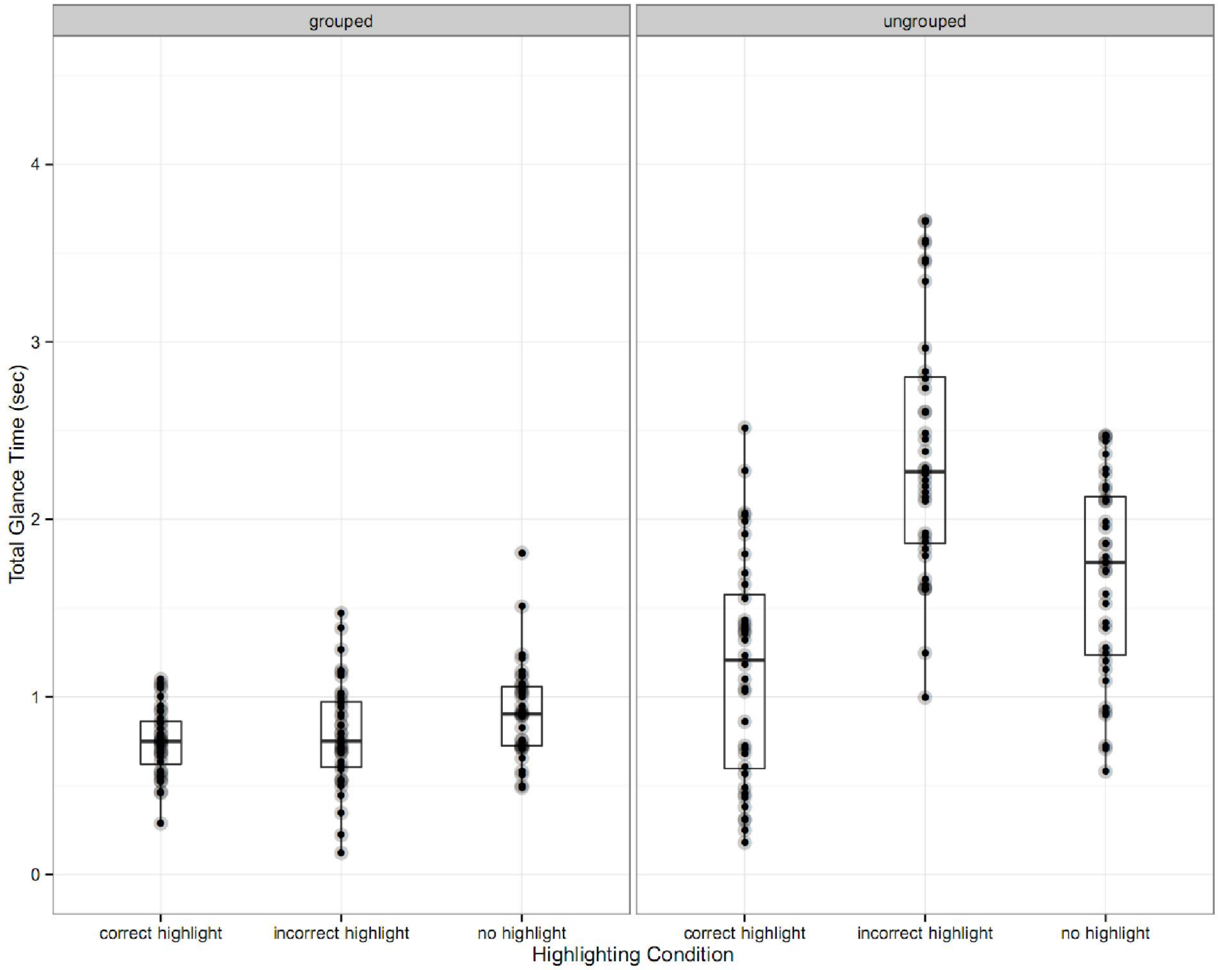


Figure 42. Comparison of predicted total glance time across all conditions by the integrated model.

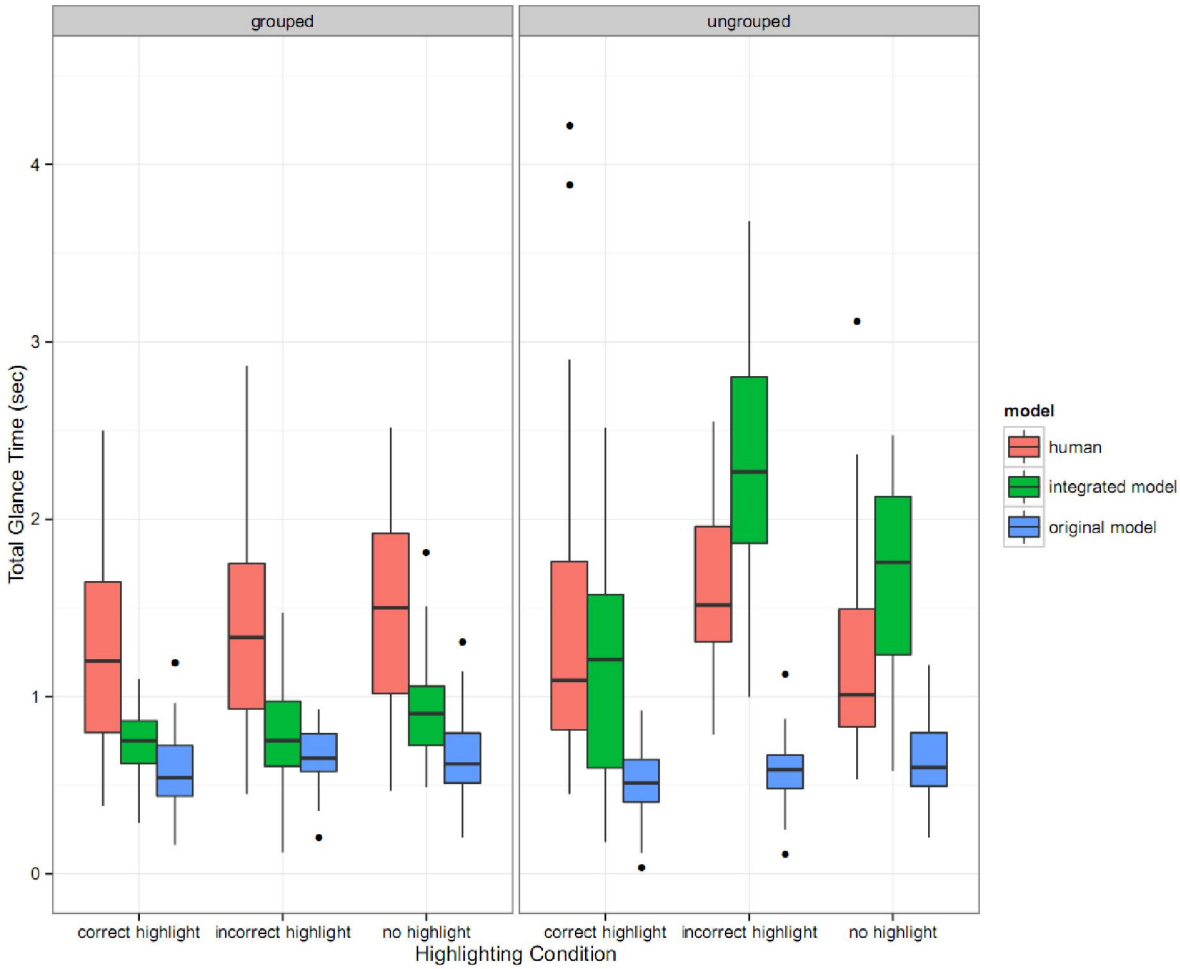


Figure 43. Comparison between model prediction and the empirical data.

6.3.3. Comparison of Maximum Glance Duration

Following the previous data analysis, the 95th percentile of maximum glance duration from the test data from the driving simulator was compared with the integrated model's prediction to test its ability to break long-duration glances based on vehicle's stability. Figure 44 shows long glances from the data on the top and long glances from the integrated model on the bottom. The results of the regression analysis indicated lane deviation at the last visual sample explained 46% of the variance [$R^2 = .46$, $F(1, 8) = 8.80$, $p < .05$] in the test set. Similarly, the results of the regression analysis indicated lane deviation at the last visual sample explained 44%

of the variance [$R^2 = .44$, $F\,(1, 37) = 31.01$, $p < .001$] in the integrated model's prediction (note: the original Distract-R does not have the function to determine maximum glance duration, so comparison with the original model was impossible).

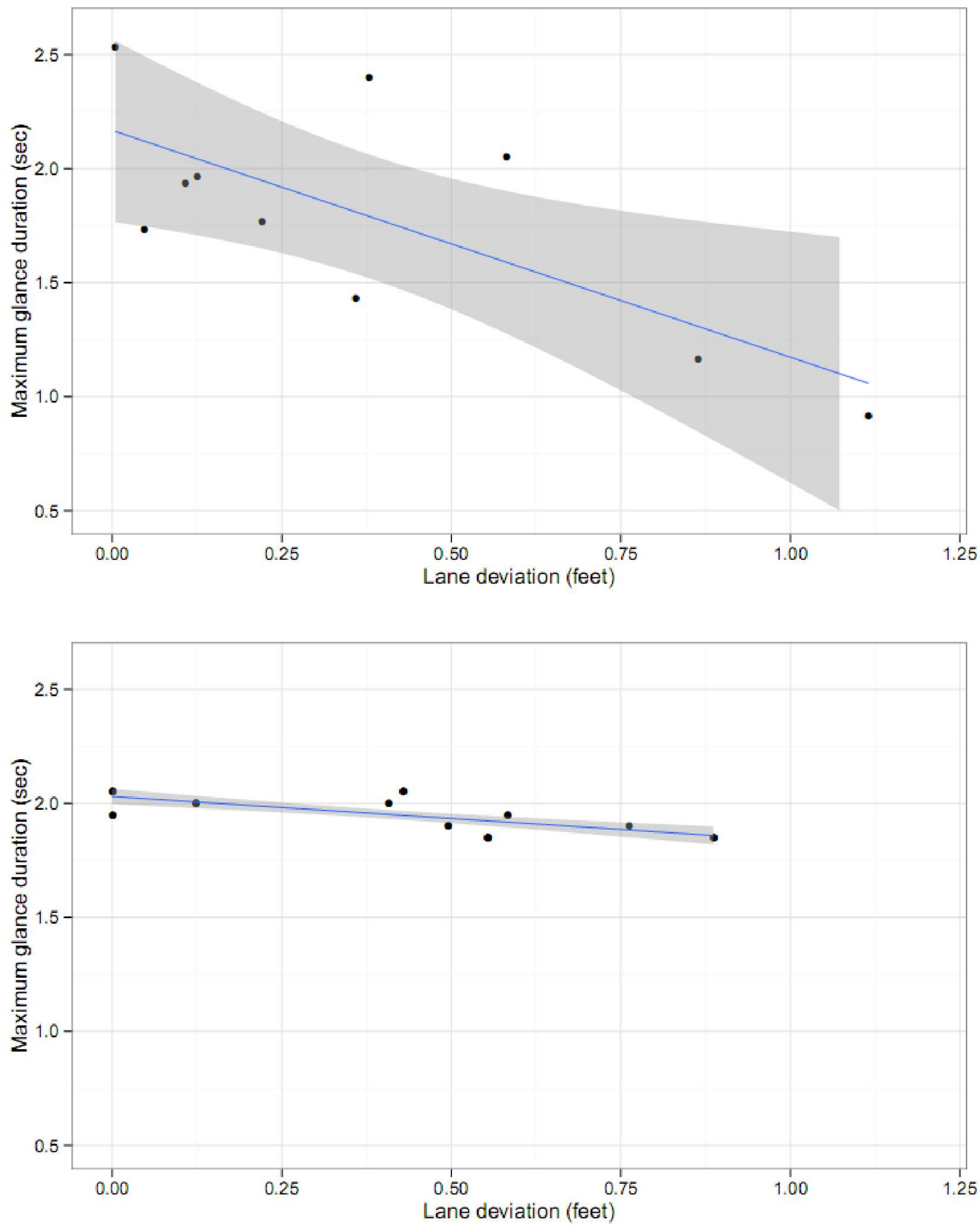


Figure 44. Scatter plot of maximum glance duration and lane deviation from the test data (top) and the integrated model (bottom).

Figure 45 compares the number of fixations from the test data, the original Distract-R, and the integrated model. As expected, the original Distract-R always completed the secondary task within one glance. However, the test data and the integrated model used multiple glances. The figure showed that human drivers more actively break long-duration glances to complete the secondary task.

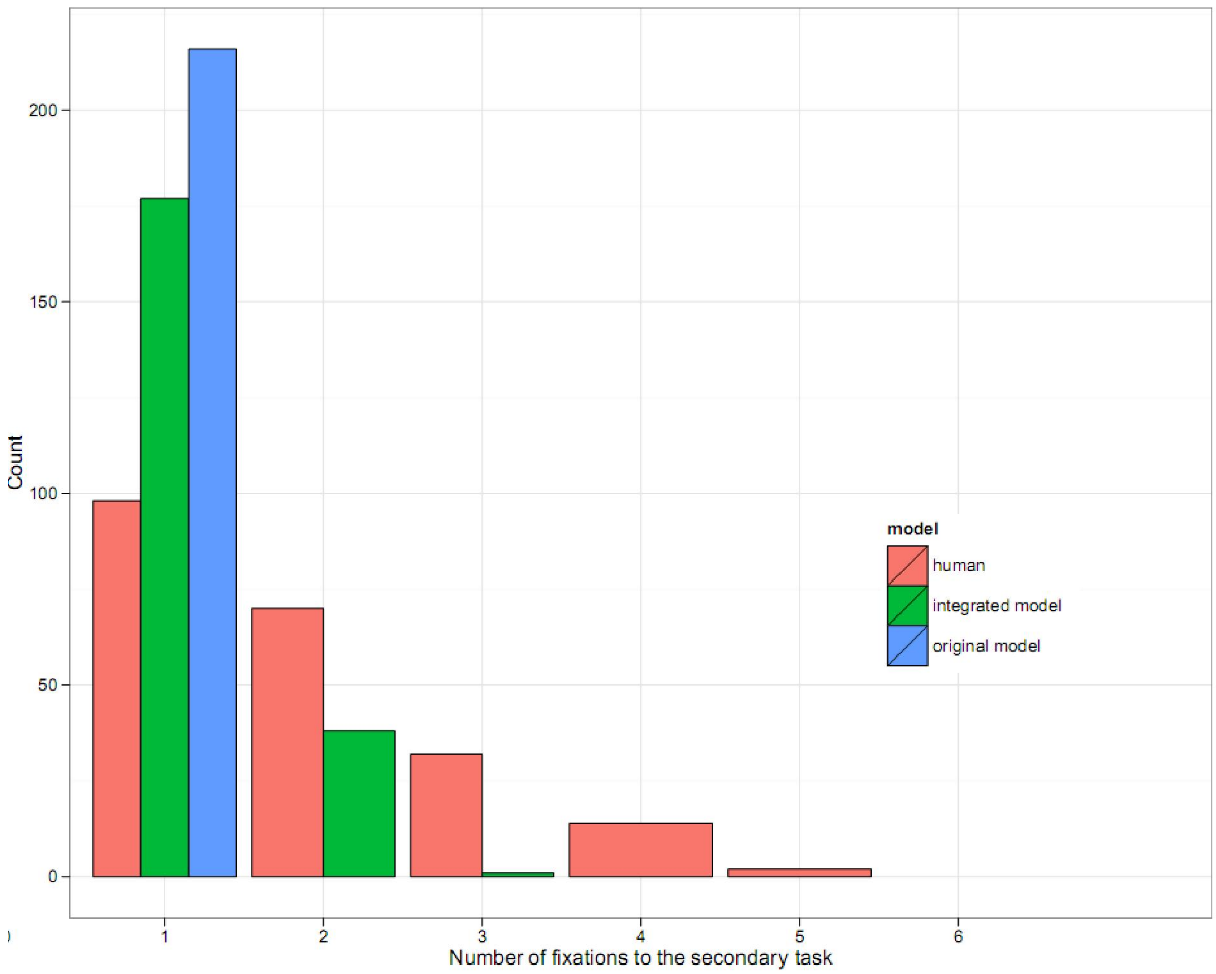


Figure 45. Comparison for number of fixations to the secondary task.

6.4. Conclusion

This chapter tested the integrated model by comparison with the empirical data and the predictions of the original Distract-R. The integrated model could predict the effect of highlighting in the ungrouped layout as observed from the experiment, and the detrimental effect of the secondary task on driving performance.

In the previous experiment, two design features (one promoting top-down process and the other one promoting bottom-up process) were tested. Drivers relied on both top-down and bottom-up attentional processes to complete the secondary task, and the original Distract-R does not have an ability to evaluate tasks that include a combination between the two processes.

In general, the integrated model outperformed the original Distract-R. Implementation of the saliency map accounted for the effect of visual features, and the activation field (with spatial expectation) accounted for both top-down and bottom-up attentional processes along with the interaction with driving tasks. The task switching function was a critical enhancement included in the integrated model, because it resembles drivers' gaze pattern while interacting with DVIs. The integrated model allowed more time to the secondary task when the vehicle is close the center of the lane and less time to the secondary task when the vehicle is far from the center. This pattern was observed in the experiment and other findings, such as self-paced occlusion study (e.g., Chen & Milgram, 2011). Although there is still a gap between human drivers and the driver model, these enhancements produced a more robust driver model and a more realistic evaluation tool for distraction potential.

7. General Discussion

This dissertation augmented Distract-R driver model to build a more robust model to assess distraction potential of DVI systems. Computational models can supplement the current guidelines for DVIs. Especially, a cognitive architecture that simulates human cognition and perception has the potential to simulate more complex and dynamic driving situations, such as multitasking while driving. The integrated model could provide a powerful way to describe the distribution of drivers' visual attention and predict potential distraction, because it accounts for symbolic representations of drivers' task goals and expectations along with a continuous field of activation associated with the salience of visual stimuli.

To build the integrated model, previous efforts to build a computational model to assess driver distraction were reviewed and requirements for the model were identified. Distract-R satisfies most of the requirements, but cannot address effects of low-level visual features, which can be critical DVI design features. The saliency map has an ability to address this limitation and to integrate bottom-up influences into Distract-R. However, the saliency map has not been validated for its ability to predict search time, and it has limited ability to account variability of human performance. To address this limitation, a Monte-Carlo approach was applied to the saliency map augmented with a stochastic process, and the augmented saliency map's predictive capability was compared with empirical data. The result showed that the augmented saliency map could be applied to predict search time to a level comparable to between-subject variability.

After validating the saliency map, model integration was conducted. The "activation field" was the conceptual basis for integrating the two models and technical model integration was conducted based the concept. In the integrated model, the saliency map interacts with visual stimuli in the world and estimates search time based on goals from Distract-R. Spatial expectation and symbolic representation was converted to a format of an activation field that could be combined with the saliency map. This combination of functions enables the integrated model to determine total glance duration to in-vehicle devices based on visual features, whereas the original model used a fixed value for all stimuli.,

The integrated model also included a function to break long-duration glances into several short glances to maintain vehicle's stability. The experiment was conducted to investigate driver distraction and to test the integrated model. The result showed that both top-down and bottom-up influences were engaged in drivers' search task, and the effect of highlighting was only significant when spatial expectation was not available to use. The integrated model's prediction and the empirical data were compared based on the experimental results. The results showed that the integrated model outperformed the original Distract-R, and the integrated model showed a similar pattern of total glance time across the experimental conditions. Moreover, the integrated model could break a long-duration glance into multiple glances to complete the secondary task as observed from the experiment.

The integrated model showed ability to account bottom-up influences and supported previous findings such as "feature search" (e.g., Treisman 1985) and the effect of display size (e.g., Treisman & Gelade, 1980). The integrated model also applied the concept of "guided search" (e.g., Wolfe, 1994) and "activation field" (e.g., LaBerge & Brown, 1989) to simulate the interaction between top-down and bottom-up influences. The integrated model might be helpful to understand the effect of design features on drivers' glances to the in-vehicle system and driving performances, and to enhance design processes by providing a tool to assess distraction potential.

The integrated model showed that it could predict the effect of design features. However, the results were obtained specific task settings, populations (especially for the Experiment 2), and simple stimuli. The subsequent section discusses limitations of this research.

7.1. Limitations

7.1.1. Limitations of Experiments

There was a significant effect of engaging the search task on driving performance, but levels of the independent variables failed to have any statistically significant effect on driving performance. In Experiment 2, the search task (e.g., finding a target item among 12 items) was relatively easy, so it might mask the effect of spatial grouping and highlighting. This dissertation measured total glance time and maximum glance duration. Around half (47.69%) of the search tasks were completed with a single glance, and these cases made it difficult to determine whether the maximum single glance was effected by drivers' capacity to control the vehicle without glancing to the roadway. The Experiment 1 revealed that magnitude of the effect of highlighting increases as the set size increases, and the future research may redesign experiment stimuli by increasing the number of items on the screen, or having combinations of sub-tasks that might require longer completion time.

The Experiment 2 fixed the vehicle velocity to control drivers' potential speed compensation. However, speed also affects glance duration as observed in the previous occlusion study. Moreover, the experiment did not account for any external factors that can affect driving performance and secondary-task performance such as amount of traffic, road width, or curvature of the road. The setting of the experiments increased controllability and simplified interpretation of the result, but might decrease generalizability of the results. Lateral deviation of the vehicle was the only factor that could affect glance duration, except design features of DVIs.

7.1.2. Limitations of the Model

The integrated model only implemented spatial expectation as a source of top-down influence. However, there was a possibility that participants relied on other top-down features for the search task. The experiment employed a within-subject design and participants might ignore highlighting during trials of correct highlighting after experiencing trials of incorrect highlighting. Intentional ignoring both correct highlighting and incorrect highlighting was also plausible. The integrated model does not have ability to account for these possible top-down influences, and this limitation might bring the difference with the empirical data. This dissertation treated highlighting as a source of bottom-up influence by using 50% of highlighting validity and relatively fewer numbers of trials to inhibit learning. However, previous exposure to other devices or other designs that applied highlighting might promote feature expectation from participants, and this possibility was not considered in the integrated model.

There was a clear difference between drivers' lane keeping behavior and the model driver's performance. For example, the model driver did not adjust vehicle's trajectory until the vehicle reached at predefined stability threshold. Therefore, the model vehicle's driving path was similar to a sine wave. However, human drivers frequently adjust steering angle and did not show a regular oscillation. The other limitation of the model is that it cannot address vehicle dynamics. For example, driving an awkward truck and driving an agile sedan might bring significant differences in driving performance and glance duration to the secondary task, but this issue was not considered in the model. This study compared data collected from a driving simulator and the integrated model's prediction by using Distract-R driver model. However, differences in vehicle dynamics between the driving simulator and the modeled vehicle were not directly compared and tuned. This limitation might affect predictions of both driving performance and secondary performance.

One enhancement of the integrated model was the ability to simulate search processes that are often required by interaction with current in-vehicle systems. The original Distract-R's model driver remembered all items' location on the screen, whereas the integrated model's model driver needed to find and search a required item on the screen. Therefore, the original Distract-R and the integrated model's model drivers represent two extremes in terms of driver's familiarity to the vehicle displays, so they could not simulate intermediate stages.

7.2. Future Research

Based on the discussion in the previous section, the present work can be extended further in several ways. As discussed above, a difficulty level of the search task in the experiments was low, so future research may increase the difficulty level by increasing the number of distractors, or may test different levels of difficulty to find critical number of items of DVIs. Current vehicle systems often involve navigation between pages and this requires more complex tasks or combinations of several actions. Modeling consecutive search tasks can be a good way to increase a level of difficulty for the secondary task while increasing reality of the task.

In the integrated model, spatial expectation was the only top-down influence. Future research should implement additional top-down influences, such as search strategies and familiarity with the system. To test familiarity of practice effects, the number of trials for each experimental condition would be increased to capture the effects. The number of highlighted items on a screen was not systematically manipulated as an independent variable for this dissertation. However, the results from the both experiments showed that the effect incorrect highlighting was significant especially when the number of highlighted items is greater than one (e.g., Experiment 2), whereas the effect of correct highlighting was significant when the number of highlighted item is one (e.g., Experiment 1). As current vehicle systems present information from multiple sources and include information not related to driving (e.g., advertisements), the number of highlighted items and ways to highlight items should be factors for further investigation.

The integrated model already built a platform to combine top-down influences with bottom-up influence, and ways to quantify the top-down factors would be challenges. The weights for top-down and bottom-up influences are implemented to compute the activation field, and this may facilitate the future work. As mentioned before, investigating potential top-down and bottom-up influences will be a useful way to extend the current research, and interaction among those influences will provide better understanding of the effects of design features on driver distraction. Moreover, cognitive architectures can easily inherit improvements from any model developed in the context of a cognitive architecture. For example, Trafton and his colleagues (2012) developed ACT-R/E (Embodied) based on ACT-R. ACT-R/E enhanced its ability to model spatial reasoning in a world, such as tracking people (or objects) as they move around, by implementing a theoretical framework of Specialized Egocentrically Coordinated Spaces. Referring applications in the modeling community should be a way to extend the present work.

This dissertation did not fully address variability between drivers and it limited generalizability of the results. Future research may extend a scope of modeling to factors related to variability between drivers such as age, gender, driving style (such as lane keeping style), or task switching style (such as several short glances to the secondary task vs. few long glances to the secondary task).

8. Conclusions

The goal of this dissertation is to develop a computational model to assess distraction potential of in-vehicle information systems by simulating both top-down and bottom-up attentional processes. The dissertation produced an integrated model that built upon Distract-R by adding the ability to (1) account for visual salience, (2) account for spatial expectation, and (3) determine glance duration to the secondary task. The model validation showed that the integrated model outperformed the original Distract-R model, and the integrated model could predict drivers' glance duration and glance pattern to a level comparable to the between-subject variability of drivers. The integrated model showed the interaction between design features that promoted both top-down and bottom-up influences. Especially, the bottom-up influence was limited when the top-down influence was dominant (e.g., in the grouped layout) in both the experiment and the model prediction. The integrated model also showed the ability to model drivers' task switching by implementing the function to determine maximum glance duration to the secondary task, and this might be helpful to understand the dynamics of drivers' visual attention.

8.1. Practical Contributions

The practical contributions of this dissertation include model modifications and integration of two models to predict distraction potential that a particular design might produces. Table 3 links each contribution to the associated chapter.

Table 3. Practical contributions of the dissertation.

Chapter	Contribution
Chapter 3	The augmented saliency map with the Monte-Carlo technique.
Chapter 4	Integration of Distract-R and the saliency map. Implementation of the activation field.
Chapter 6	A computational model that can account for both top-down and bottom-up attentional process with an ability to simulate driving performance. Implementation of the function to determine glance duration to the secondary task.

8.2. Theoretical Contributions

The theoretical contributions of this dissertation include an investigation of driver distraction and building a more robust computational model for assessing driver distraction. Table 4 links each contribution to the associated chapter.

Table 4. Theoretical contributions of the dissertation.

Chapter	Contributions
Chapter 3	Testing the effects of potential design features. Validation of the saliency map's ability to predict search time.
Chapter 5	Understanding of the impact of driver distraction. Understanding drivers' glance pattern while multitasking.
Chapter 6	A comparison between the empirical data and the model prediction. Validation of the integrated model that can account for both top-down and bottom-up attentional influences.

8.3. Summary

The integrated model with the saliency map provides an ability to predict variations in off-road glances based on saliency cues, and thus predict downstream effects on driver performance. For example, in contrast to the "saliency-blind" original system, the saliency map integration allows the model to predict how coloring and/or highlighting of a screen affect driver performance. Thus, the integrated model can equip designers with a systematic approach to assessing and reducing distraction by providing estimated search time of specific items varying visual features.

The inclusion of highlighting tested the saliency model's "bottom-up" influence on visual attention. The saliency map computes the influence of low-level visual features of a display that attract driver's attention, which can be the features that might draw driver's attention away from the area of display that contains the information of interest to the driver. Predicting the salience of display features can be central to estimating distraction potential, because it determines how likely an object is to attract drivers' visual attention. If important information is highly salient relative to the interface background, it will be detected easily with a few glances. However, misplaced salience—situations in which highly salient display features do not correspond to drivers' information needs— might lead to long glances away from the road. The saliency map can identify instances of misplaced salience, which could help designers reduce the distraction potential of connected vehicle displays. The inclusion of spatial grouping through the activation field demonstrated how the integrated model combines top-down with bottom-up influence. As drivers learn the location of items on the screen, the integrated model can dictate these spatial expectations to the activation field, again influencing visual scanning time and thus driver performance.

By implementing the saliency map and the activation field, the integrated model could test the interaction between highlighting and spatial grouping as observed from the experiment. This implementation enhanced the original model, and the enhancement was tested by comparison with the empirical data and the original model. Implementation of the function to determine maximum glance duration to the secondary task brought the ability to the model driver so the model could allow glances to the secondary task based on vehicle's stability.

The integrated model can be applied by system designers and in evaluating the distraction potential of candidate displays and tasks. Consider an example of an in-vehicle system with dynamic re-routing. When the system alerts the driver to a potential issue, it must now inform the driver about the reason for the alert, possible alternate routes, their associated travel times and costs, and so on. The visual display, as the most likely way to provide this information, becomes the critical bottleneck for behavior: a well-designed display can afford the information in a quick convenient way, whereas a poorly designed display could confuse the driver and result in potentially hazardous distraction. For example, designers may use salient colors for more important information such as warnings and emergent information. This could promote faster search and reaction times, because drivers could have expectations for specific types of information after learning this color scheme. However, if the visual salience of the display runs counter to these expectations, the interface will need more time to guide the driver's attention to other information. In other situations, designers may spatially segregate information by their sources or types, and drivers could have spatial expectations for specific target information. The integrated model combines salient visual features and drivers' expectations, and generates robust predictions of performance. The model should thus help to evaluate the display at this critical juncture, facilitating display design so that attention is drawn directly to the most informative parts of the display, minimizing the time drivers take their eyes off the road.

Appendix

Experiment 2 Instructions

Task Instructions

During this part of the study, you will be interacting with a touch-screen system while driving. On the touch screen, there will be 12 buttons at a time, with the following labels:

Gas: Chevron, Exxon, Shell

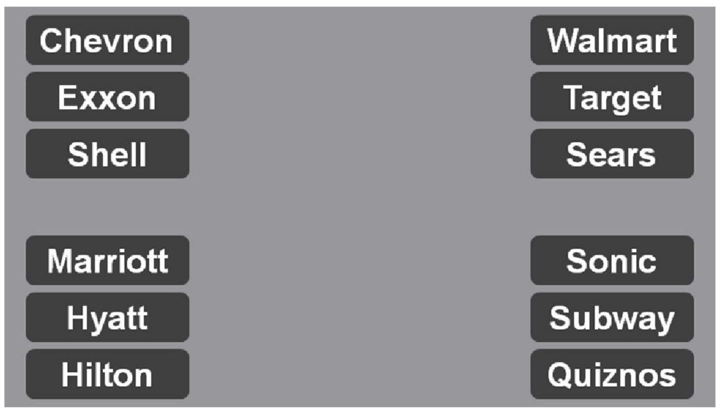
Hotels: Marriott, Hyatt, Hilton

Stores: Walmart, Target, Sears

Foods: Sonic, Subway, Quiznos

Task Instructions

The buttons will appear in one of two layouts. For the first layout, buttons will be organized into categories like this:

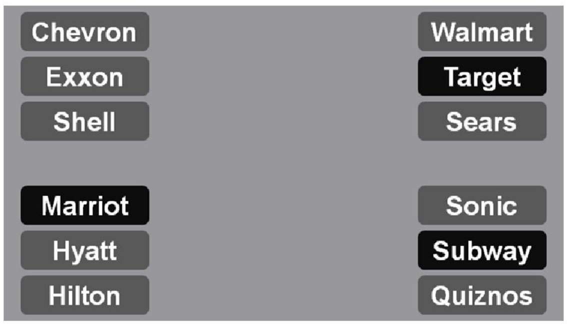


Task Instructions

For the second layout, the buttons will be randomly placed regardless of the categories, like this:

Task Instructions

In both layouts, there will be some screens that highlight buttons like this:

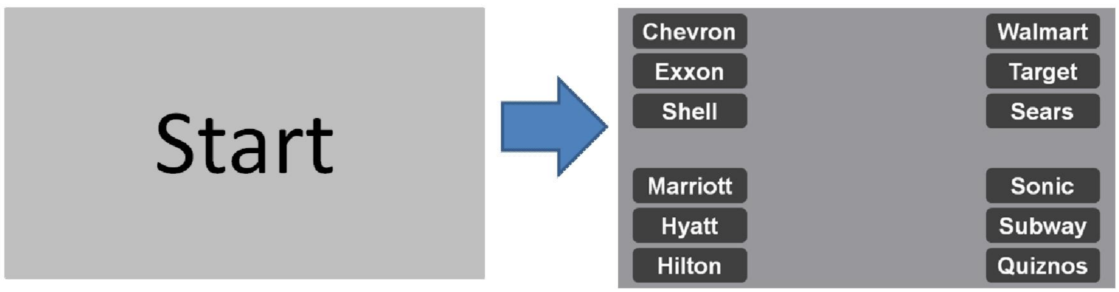


Task Instructions

During the experiment, you will hear a verbal instruction that tells you which button to press.

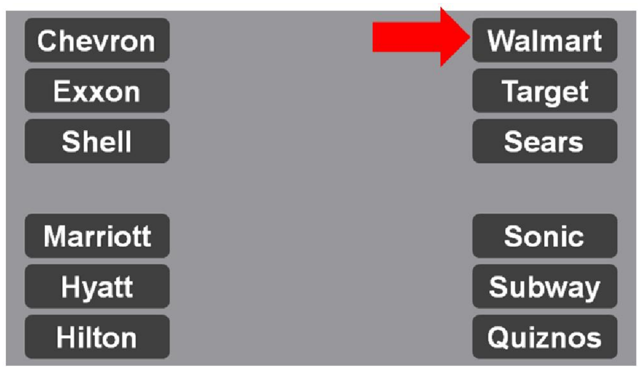
For example you might hear "Walmart."

Touch the start screen to bring up the display:



Task Instructions

Then you should find the correct button and press it as quickly as possible while still driving safely.



The instructions will come at about every 20 seconds.

During your Drive

You will be following a car in front of you.

When the drive starts, you will accelerate to 60 mph.

The cruise control will automatically engage.

You won't have to control your speed or the distance to the car in front of you unless the lead car brakes.

During your Drive

The lead car will occasionally brake and its brake lights will illuminate.

You should tap your brakes as quickly as possible when you see the brake lights.

After you tap your brakes, the cruise control will automatically resume. You will <u>not</u> have to press a "resume" button.

References

3M Commercial Graphics Division. (2010). 3M Visual Attention Service Validation Study. Retrieved from http://solutions.3m.com/3MContentRetrievalAPI/BlobServlet?locale=en_WW&lmd=1272291661000&assetId=1258566024761&assetType=MMM_Image&blobAttribute=ImageFile

Alliance of Automobile Manufacturers. (2003). Statement of Principles on Human-Machine Interfaces (HMI) for Invehicle Information and Communication Systems (Version 3.0). Washington, D.C: Alliance of Automobile Manufacturers. Retrieved from http://www.umich.edu/~driving/safety/guidelines.html

Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?* New York: Oxford University Press.

Bittner, A. C., Simsek, O., Levison, W. H., & Campbell, J. L. (2002). On-road versus simulator data in driver model development - Driver performance model experience. *Transportation Research Record: Journal of the Transportation Research Board, 1803(1)*, 38-44.

Boer, E. R. (1999). Car following from the driver's perspective. *Transportation Research-Part F: traffic psychology and behaviour, 2*(4), 201–206.

Brackstone, M., & McDonald, M. (1999). Car-following: a historical review. *Transportation Research-Part F: traffic psychology and behaviour*, *2*(4), 181–196.

Campbell, J. L., Carney, C., & Kantowitz, B. H. (1998). *Human Factors Design Guidelines for Advanced Traveler Information Systems (ATIS) and Commercial Vehicle Operations (CVO) (No. FHWA-RD-98-057).* McLean, VA: U.S. Dept. of Transportation, Federal Highway Administration.

Card, S. K., Moran, T. P., & Newel, A. (1980). The keystroke-level model for user performance time with interactive systems. *Communication of the ACM*, *23*(7), 396–410.

Chen, H.-Y. W., & Milgram, P. (2011). Determining fixed glance duration for visual occlusion research. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *55*(1), 1904–1908.

Chen, H.-Y. W., & Milgram, P. (2013). A framework for modelling and analysing variability in visual occlusion experiments. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 1884–1888.

Donges, E. (1978). A two-level model of driver steering behavior. *Human Factors*, *20*, 691–707.

Driver Focus-Telematics Working Group of Alliance of Automobile Manufactures. (2002). Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems. *Alliance of Automotive Manufacturers.*

Egeth, H., Jonides, J., & Wall, S. (1972). Parallel processing of multielement displays. *Cognitive Psychology*, *3*(4), 674–698.

Fajen, B. R., & Warren, W. H. (2003). Behavioral dynamics of steering, obstacle avoidance, and route selection. *Journal of Experimental Psychology-Human Perception and Performance*, *29*(2), 343–362.

Fisher, D. L., & Tan, K. C. (1989). Visual displays: The highlighting paradox. *Human Factors*, *31*(1), 17–30.

Fleetwood, M., & Byrne, M. (2006). Modeling the visual search of displays: A revised ACT-R model of icon search based on eye-tracking data. *Human-Computer Interaction*, *21*(2), 153–197.

Godthelp, H. (1986). Vehicle control during curve driving. *Human Factors*, *28*, 211–221.

Godthelp, H., Milgram, P., & Blaauw, G. J. (1984). The development of a time-related measure to describe driving strategy. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *26*(3), 257–268.

Green, P., Levison, W., Paelke, G., & Serafin, C. (1995). Preliminary human factors design guidelines for driver information systems. *NASA* (19980016891). McLean, VA. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Preliminary+Human+Factors+Design+Guidelines+for+Driver+Information+Systems#0

Hankey, J. M., Dingus, T. A., Hanowski, R. J., & Wierwille, W. W. (2000). The development of a design evaluation tool and model of attention demand. *Contract No. DTFH61-96-C-00071). McLean, VA: US Department of Transportation, Federal Highway Administration.*

Harbluk, J. L., Noy, Y. I., Trbovich, P. L., & Eizenman, M. (2007). An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance. *Accident Analysis and Prevention*, *39*(2), 372–379.

Harel, J., Koch, C., & Perona, P. (2007). Graph-based visual saliency. *Advances in Neural Information Processing Systems*, *19*, 545–552.

Hoffman, J. E. (1978). Search through a sequentially presented visual display. *Attention, Perception, & Psychophysics*, *23*(1), 1–11.

Horberry, T., Anderson, J., Regan, M. a, Triggs, T. J., & Brown, J. (2006). Driver distraction: the effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. *Accident; analysis and prevention*, *38*(1), 185–91.

Horrey, W. J., & Lesch, M. F. (2009). Driver-initiated distractions: examining strategic adaptation for in-vehicle task initiation. *Accident Analysis and Prevention*, *41*(1), 115–122.

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.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, *40*(10-12), 1489–506.

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

James, W. (1890). *The Principles of Psychology*. New York: Henry Holt.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, *4*, 219–227.

Laberge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review*, *96*(1), 101–124.

Lee, J., Lee, J. D., & Salvucci, D. D. (2012). Evaluating the distraction potential of connected vehicles. *4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 33–40.

Levison, W. H. (1998). Interactive highway safety design model: Issues related to driver modeling. *Transportation Research Record: Journal of the Transportation Research Board*, *1631*(1), 20–27.

Levison, W. H., Simsek, O., Bittner, A. C., & Hunn, S. J. (2001). Computational techniques used in the driver performance model of the interactive highway safety design model. *Transportation Research Record*, *1779*(01), 17–25.

MapBox. (2013). MAKI ICONS. Retrieved from http://mapbox.com/maki/

National Highway Traffic Safety Administration. (2012). Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices. *Washington, DC: National Highway Traffic Safety Administration (NHTSA), Department of Transportation (DOT).*

National Highway Traffic Safety Administration. (2013). Official US Goverment Website for Distracted Driving. Retrieved from http://www.distraction.gov/content/get-the-facts/facts-and-statistics.html

Niemelä, M., & Saariluoma, P. (2003). Layout attributes and recall. *Behaviour & Information Technology*, *22*(5), 353–363.

Pedhazur, E. J. (1982). *Multiple Regression in Behavioral Research: Explanation and Prediction*. New York: Holt, Rinehard and Winston.

Peters, R. J., Iyer, A., Itti, L., & Koch, C. (2005). Components of bottom-up gaze allocation in natural images. *Vision research*, *45*(18), 2397–416.

Posner, M., Nissen, M., & Ogden, W. (1978). Attended and unattended processing modes: The role of set for spatial location. *Modes of perceiving and processing information*, *137*, 158.

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology*, *32*(1), 3–25.

R Development Core Team. (2011). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Regan, M. A., Young, K. L., & Lee, J. D. (2009). Introduction. In *Driver Distraction: Theory, Effects, and Mitigation* (pp. 3–7). Boca Raton, FL: CRC Press.

Regan, M. A., Young, K. L., Lee, J. D., & Gordon, C. P. (2009). Source of Driver Distraction. In *Driver Distraction: Theory, Effects, and Mitigation* (pp. 249–278). Boca Raton, FL: CRC Press.

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, *7*, 1–22.

Salvucci, D. D. (2001). An integrated model of eye movements and visual encoding. *Cognitive Systems Research*, *1*, 201–220.

Salvucci, D. D. (2005). A multitasking general executive for compound continuous tasks. *Cognitive Science*, *29*(3), 457–492.

Salvucci, D. D. (2006). Modeling driver behavior in a cognitive architecture. *Human Factors*, *48*(2), 362–380.

Salvucci, D. D. (2009). Rapid prototyping and evaluation of in-vehicle interfaces. *Acm Transactions on Computer-Human Interaction*, *16*(2), 9:1–9:33.

Salvucci, D. D., & Beltowska, J. (2008). Effects of memory rehearsal on driver performance: experiment and theoretical account. *Human Factors*, *50*(5), 834–844.

Salvucci, D. D., & Taatgen, N. A. (2008). Threaded cognition: An integrated theory of concurrent multitasking. *Psychological Review*, *115*(1), 101–130.

Schindhelm, R., Gelau, C., Keinath, A., Bengler, K., Kussmann, H., Kompfner, P., … & Maritnetto, M. (2004). *Report on the Review of the Available Guidelines and Standards*. Technischer Bericht IST-1-507674-IP, AIDE-Adaptive Integrated Driver-Vehicle Interface.

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.

Steelman, K. S., McCarley, J. S., & Wickens, C. D. (2011). Modeling the control of attention in visual workspaces. *Human Factors*, *53*(2), 142–153.

Strayer, D. L., & Drews, F. A. (2004). Profiles in driver distraction: Effects of cell phone conversations on younger and older drivers. *Human Factors*, *46*(4), 640–649.

Stutts, J., Feaganes, J., Reinfurt, D., Rodgman, E., Hamlett, C., Gish, K., & Staplin, L. (2005). Driver's exposure to distractions in their natural driving environment. *Accident Analysis and Prevention*, *37*(6), 1093–1101.

Trafton, J. G., Hiatt, L. M., Harrison, A. M., Tamborello, P., Khemlani, S. S., & Schultz, A. C. (2013). ACT-R/E: An embodied cognitive architecture for human-robot interaction. *Journal of Human-Robot Interaction, 2,* 30-54.

Treisman, A. (1985). Preattentive processing in vision. *Computer vision, graphics, and image processing*, *31*(2), 156–177.

Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology*, *40*(2), 201–237.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, *12*(1), 97–136.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychological review*, *95*(1), 15–48.

Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural networks: the official journal of the International Neural Network Society*, *19*(9), 1395–407.

Wickens, C. D., Helleberg, J., Goh, J., Xu, X., & Horrey, B. (2001). *Pilot Task Management: Testing an Attentional Expected Value Model of Visual Scanning*. Savoy, IL: University of Illinois, Aviation Research Lab.

Wickens, C. D., & McCarley, J. S. (2008). *Applied Attention Theory*. Boca Raton: CRC Press.

Wilkie, R., & Wann, J. (2003). Controlling steering and judging heading: Retinal flow, visual direction, and extraretinal information. *Journal of Experimental Psychology-Human Perception and Performance*, *29*(2), 363–378.

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. *Journal of experimental psychology. Human perception and performance*, *15*(3), 419–33.

Wright, R. D., & Ward, L. M. (2008). *Orienting of Attention*. New York: Oxford University Press.