Crash/Near-Crash: Impact of Secondary Tasks and Real-Time Detection of Distracted Driving

Peter Ramzy Zaki Bakhit
Louisiana State University and Agricultural and Mechanical College, peterramzy277@gmail.com

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations

Part of the Transportation Engineering Commons

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_dissertations/4723

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
CRASH/NEAR-CRASH: IMPACT OF SECONDARY TASKS AND REAL-TIME DETECTION OF DISTRACTED DRIVING

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agriculture and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Civil & Environmental Engineering

by

Peter Ramzy Zaki Bakhit
B.Sc, Cairo University, 2012
M.S., Cairo University, 2015
December 2018
TO MY FAMILY
ACKNOWLEDGEMENTS

First and before any acknowledgements, I would like to express my cordial gratitude to my advisor Dr. Sherif Ishak for his guidance and his belief that I will make a good researcher. I could not have reached this successful end without his full support, his thoughtful insights, and his mentorship in both personal and academic lives. Without Dr. Ishak, I could not have been able to publish and present more than twelve research papers at major academic conferences and journals such as Transportation Research Board meetings (TRB). I will always be indebted to him for giving me the opportunity to succeed and become the person who I am now. Thank you Dr. Ishak!

I also want to thank another committee member, Dr. Chester Wilmot for his thoughtful comments, suggestions, and his endless support to reach this point. To Dr. BeiBei Guo, another committee member, thank you for advising and helping whenever I faced any statistical challenge. I feel very grateful for the time and the effort you dedicated for our research discussions. I could not have been completed this dissertation without this extraordinary committee members.

To my father and mother, thank you for the unconditional love, the support and the understanding that helped me endure all the difficulties I faced during my graduate studies. Thank you for wishing me the best and I hope I have made you both proud of me.

To my hidden sources of inspiration — my brothers: Mina and Andrew —, thank you for the motivational support I received from both of you either directly or indirectly. Your commitments and dedications to whatever task you both do in life have showed me how to become a self-motivated and committed person. I feel truly indebted to both of you.

Finally, to my colleagues, thank you for teaching, helping, and sharing your knowledge with me. I feel grateful to all of you and to every single person who helped me reaching this point in my life.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................................................................................ iii

LIST OF TABLES .................................................................................................................. vi

LIST OF FIGURES ............................................................................................................. vii

LIST OF ABBREVIATIONS ................................................................................................. viii

ABSTRACT ............................................................................................................................ ix

1 INTRODUCTION .................................................................................................................. 1
   1.1 General Overview ......................................................................................................... 1
   1.2 Research Motivation ..................................................................................................... 2
   1.3 Research Objectives .................................................................................................... 4
   1.4 Dissertation Outline ................................................................................................... 5

2 LITERATURE REVIEW ....................................................................................................... 8
   2.1 Part 1: Driver Distraction ............................................................................................ 8
   2.2 Part 2: Driver Attention Allocation Process ............................................................... 19
   2.3 Part 3: Distracted Driving Detection ........................................................................... 21
   2.4 Limitations in Previous Studies .................................................................................. 23

3 DATA FORMATION AND GENERAL METHODOLOGY .................................................. 26
   3.1 SHRP2 NDS Overview .............................................................................................. 26
   3.2 Data Format ................................................................................................................ 28
   3.3 Event Detailed Data .................................................................................................... 28
   3.4 Time-Series Data ....................................................................................................... 30
   3.5 Driver Questionnaire ................................................................................................. 34
   3.6 Naturalistic Engagement in Secondary Tasks Dataset (NEST) .................................. 34
   3.7 General Methodology ............................................................................................... 38

4 CRASH/NEAR-CRASH RISK ASSESSMENT OF DISTRACTED DRIVING AND ENGAGEMENT IN SECONDARY TASKS ................................................................. 42
   4.1 Introduction ................................................................................................................ 42
   4.2 Data Description ......................................................................................................... 43
   4.3 Models Development ................................................................................................. 46
   4.4 Discussion .................................................................................................................. 54
   4.5 Summary .................................................................................................................... 56

5 DETECTING DISTRACTED DRIVING VISUAL BEHAVIOR ........................................... 58
   5.1 Introduction and Background .................................................................................... 58
   5.2 Data Description ....................................................................................................... 61
   5.3 Methods .................................................................................................................... 62
   5.4 Results and Discussion ............................................................................................. 65
LIST OF TABLES

Table 2.1. Percentages of secondary tasks involved in distraction related crashes to all crashes .................................................. 16
Table 3.1. SHRP2 Naturalistic driving variables ......................................................................................................................... 33
Table 4.1. List of input variables .................................................................................................................................................. 44
Table 4.2. Secondary tasks classification ......................................................................................................................................... 45
Table 4.3. Bivariate probit model results ......................................................................................................................................... 48
Table 4.4. Association analysis model results ................................................................................................................................. 55
Table 4.5. Secondary tasks ranking ................................................................................................................................................. 56
Table 5.1. Eye glance classification .................................................................................................................................................. 64
Table 5.2. Renewal cycle results ..................................................................................................................................................... 66
Table 5.3. $N_{RC}$ mixed model ANOVA results ............................................................................................................................. 67
Table 5.4. Renewal cycle distribution across different secondary tasks ......................................................................................... 68
Table 5.5. DI Mixed-model ANOVA results ....................................................................................................................................... 71
Table 6.1. List of input variables ..................................................................................................................................................... 77
Table 6.2. ANN results - confusion matrix ....................................................................................................................................... 79
Table 6.3. ANN performance measures ............................................................................................................................................ 79
Table B.1. Variables collected with DAS ........................................................................................................................................ 102
LIST OF FIGURES

Figure 2.1. Driving simulators units. ........................................................................................................ 12
Figure 2.2. Karlsson's distraction activity. .................................................................................................... 13
Figure 2.3. Naturalistic driving study example. ............................................................................................ 18
Figure 3.1. SHRP2 NDS data collection sites. ............................................................................................... 26
Figure 3.2. Data Acquisition System (DAS) — InSight website. ................................................................. 27
Figure 3.3. SHRP2 NDS website (InSight website). ....................................................................................... 27
Figure 3.4. Speed profile sample .................................................................................................................. 31
Figure 3.5. Acceleration profile sample ......................................................................................................... 31
Figure 3.6. Drivers age distribution in SHRP2 NDS database. ................................................................. 34
Figure 3.7. Gender distribution in NEST dataset. ......................................................................................... 35
Figure 3.8. Age distribution in NEST dataset. .............................................................................................. 36
Figure 3.9. Vehicle classification in NEST dataset. ..................................................................................... 36
Figure 3.10. Drivers' annual miles traveled distribution in NEST dataset. ................................................ 37
Figure 3.11. Research framework. ............................................................................................................... 39
Figure 4.1. Odds ratios of different secondary tasks. .................................................................................... 51
Figure 5.1. Glance location distribution over time for NEST safety critical events (colored). .... 63
Figure 5.2. Eye glance locations. ................................................................................................................ 64
Figure 6.1. Methodology ............................................................................................................................... 74
Figure 6.2. ANN model structure. ............................................................................................................... 75
Figure 6.3. ANN results - variable importance chart .................................................................................. 80
Figure B.1. Data acquisition system equipment. ......................................................................................... 101
Figure B.2. Eye movement calibration. ........................................................................................................ 103
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHRP2</td>
<td>Second Strategic Highway Research Program</td>
</tr>
<tr>
<td>NEST</td>
<td>Naturalistic Engagement in Secondary Tasks dataset</td>
</tr>
<tr>
<td>NN</td>
<td>Neural Network</td>
</tr>
<tr>
<td>NHTSA</td>
<td>National Highway Traffic Safety Administration</td>
</tr>
<tr>
<td>VMT</td>
<td>Vehicle Miles Traveled</td>
</tr>
<tr>
<td>SCE(s)</td>
<td>Safety Critical Event(s) (Crash/Near-Crash events)</td>
</tr>
<tr>
<td>NDS</td>
<td>Naturalistic Driving Study</td>
</tr>
<tr>
<td>SHRP2 NDS</td>
<td>Second Strategic Highway Research Program Naturalistic Driving Study dataset</td>
</tr>
<tr>
<td>ANN</td>
<td>Artificial Neural Network</td>
</tr>
<tr>
<td>LSU</td>
<td>Louisiana State University</td>
</tr>
<tr>
<td>FRD</td>
<td>Fields of Relevant Driving</td>
</tr>
<tr>
<td>DAS</td>
<td>Data Acquisition System</td>
</tr>
<tr>
<td>VTTI</td>
<td>Virginia Tech Transportation Institute</td>
</tr>
<tr>
<td>OR</td>
<td>Odds Ratio</td>
</tr>
<tr>
<td>TEORT</td>
<td>Total Eyes Off-Road Time</td>
</tr>
<tr>
<td>FFBP</td>
<td>Feed Forward Backward Propagation</td>
</tr>
<tr>
<td>AUC</td>
<td>Area Under the Curve</td>
</tr>
<tr>
<td>DI</td>
<td>Distraction Index</td>
</tr>
<tr>
<td>N&lt;sub&gt;RC&lt;/sub&gt;</td>
<td>Number of Renewal Cycles per driving event</td>
</tr>
<tr>
<td>ADAS</td>
<td>Advanced Driver Assistance Systems</td>
</tr>
</tbody>
</table>
ABSTRACT

The main goal of this dissertation is to investigate the problem of distracted driving from two different perspectives. First, the identification of possible sources of distraction and their associated crash/near-crash risk. That can assist government officials toward more informed decision-making process, allowing for optimized allocation of available resources to reduce roadway crashes and improve traffic safety. Second, actively counteracting the distracted driving phenomenon by quantitative evaluation of eye glance patterns.

This dissertation research consists of two different parts. The first part provides an in-depth analysis for the increased crash/near-crash risk associated with different secondary task activities using the largest real-world naturalistic driving dataset (SHRP2 Naturalistic Driving Study). Several statistical and data mining techniques are developed to analyze the distracted driving and crash risk. More specifically, two different models were employed to quantify the increased risk associated with each secondary task: a baseline-category logit model, and a rule mining association model. The baseline-category logit model identified the increased risk in terms of odds ratios, while the A-priori association algorithm detected the associated risks in terms of rules. Each rule was then evaluated based on the lift index. The two models succeeded in ranking all the secondary task activities according to the associated increased crash/near-crash risk efficiently.

To actively counteract to the distracted driving phenomenon, a new approach was developed to analyze eye glance patterns and quantify distracted driving behavior under safety and non-Safety Critical Events (SCEs). This approach is then applied to the Naturalistic Engagement in Secondary Tasks (NEST) dataset to investigate how drivers allocate their attention while
driving, especially while distracted. The analysis revealed that distracted driving behavior can be well characterized using two new distraction risk indicators. Additional statistical analyses showed that the two indicators increase significantly for SCE compared to normal driving events. Consequently, an artificial neural network (ANN) model was developed to test the SCEs predictability power when accounting for the two new indicators. The ANN model was able to predict the SCEs with an overall accuracy of 96.1%. This outcome can help build reliable algorithms for in-vehicle driving assistance systems to alert drivers before SCEs.
1 INTRODUCTION

1.1 General Overview

Driving is a central part in people’s daily lives. Recent studies have shown that American driver spends approximately 17,600 minutes behind the wheel each year (1). This is equivalent to seven 40-hour weeks at the office. During that time a driver travels around 10,900 miles and drives more than 290 hours. The AAA foundation safety report estimated the Vehicle Miles Traveled (VMT) in year 2015 to be 2.45 trillion vehicle miles. This number represents a 2.5% increase in the VMT that was estimated in 2014. Various attempts have been made to cope with the increased number of vehicles on roadways and provide safer surface transportation system. However, the number of fatal crashes continues to increase. The National Highway Traffic Safety Administration (NHTSA) estimated the total number of fatal crashes in 2015 to be 32,166 compared to 30,056 fatal crashes in 2014 (2). The fatality rate per million VMT in 2015 was 1.13, compared to 1.08 in 2014 (2). These numbers and facts demonstrate how driving safety represents a fundamental issue in transportation research. Reducing the number of fatal crashes does not only save people’s lives but also helps reduce the economic impact associated with traffic crashes.

According to the open literature, distracted driving and driving inattention are two leading causes of roadway crashes (3; 4). A recent report by NHTSA has indicated that more than 3,477 Americans were killed and 391,000 were injured in distraction-related Safety Critical Events (SCEs) (5). One of the main causes of distraction is engagement in secondary tasks while driving. Despite the complexity associated with the driving task, it is not uncommon to observe drivers perform other secondary tasks while operating a vehicle. Thus, understanding the role of distracted
driving and driving inattention in crash occurrences is important to the development and implementation of crash prevention measures.

According to the Auto-Safety report published in 2013, distracted driving is the number one cause of death among youths in the United States (6). The report shows that the number of teenage drivers (under 20 years old) killed as a result of distracted driving is higher than the number of teenage drivers killed as a result of drunk driving. Nonetheless, teen drivers continue to use cellphones, especially when they drive alone. Moreover, in this report, a survey was conducted to gain some insights into distracted driving behavior. Among 2000 young participants, 71% reported that reading, receiving texts, and emailing are unacceptable. Nevertheless, nearly 45% of them continue to do so. The survey also shows that 32% of these teenagers read texts and emails in the presence of passengers, while 95% do so if no passenger exists. Similarly, 90% of the participants post on social media sites while driving, but only 29% do so in presence of passengers. It seems that advanced technology such as smart phones and vehicle integrated systems plays a major role in increasing the number of distractors nowadays. However, there is also a wide belief that advanced technology including new distraction countermeasure systems could help solve the distracted driving phenomenon.

1.2 Research Motivation

Large bodies of research have shown that distracted driving increases the crash risk significantly. However, the majority of these studies developed their crash risk assessment models from empirical studies of driver behavior derived from simulators, lab/test track, surveys, interviews, and controlled traffic experiments (7-11). While these methods allow traffic safety researchers to determine the increased crash risk reasonably accurate, they are deemed insufficient to measure precisely the degree of increased (or decreased) crash risk due to the following reasons:
1- In controlled traffic experiments, participants do not decide where, when, and how to engage in a secondary task, which is not a real representation of real-world secondary task involvement;

2- The transferability of the outcomes from the driving simulators to real life remains questionable;

3- Secondary task engagement exists in non-safety and safety critical events (crash or near-crash). Therefore, in order to measure whether distraction resulted from different secondary tasks or not, and if these tasks affected the crash/near-crash risk or not, it is important to obtain information regarding exposure (normal driving events) and risk of these secondary tasks.

Therefore, in this dissertation research a large naturalistic driving dataset, collected by the second Strategic Highway Research Program Naturalistic Driving Study (SHRP2 NDS), is exploited to investigate the relationship between the engagement in a secondary task and the crash/near-crash likelihood. This dataset includes the most recent information collected from more than 3,000 drivers recruited in six different states in the United States. This dataset not only contains information about crash and near-crash driving events, but also normal driving events. By far, this is the largest naturalistic driving study dataset collected to date and is considered the best representation of the driving population.

Although recent advancement in technology plays a major role in increasing the number of distractors among drivers, advanced technology can also help to develop distraction countermeasure systems. A distraction countermeasure system is “a system that has a way of monitoring the driver to make inferences about the driver attentional status, and if a distraction criterion is met, it activates some distraction countermeasure” (12). Essentially, distracted driving
impairs driver’s visual, physical and cognitive abilities. Eye movement trackers can provide access to several types of distractions. For instance, previous studies have shown that eye movement is not only sensitive to the visual distraction but auditory secondary tasks as well (13-15). The recent advancements in eye movement tracker technology, in addition to the continuous monitoring of driver visual behavior in SHRP2 NDS dataset, provide ample opportunities to better understand distracted driving and eye glance behavior in a real world environment. However, several questions need to be answered first:

- Which approach is best to analyze driver’s visual behavior during safety and non-safety critical events?
- How to quantify driver’s eye glance behavior?
- How to measure the level of distracted driving in real time?
- How to use the distraction level measure in characterizing safety critical events? And,
- How to develop an effective distraction countermeasure system?

It is worth mentioning that, in this dissertation research, crash and near-crash driving events are referred to as Safety Critical Events (SCEs), whereas normal driving events are referred to as non-SCE.

1.3 Research Objectives

This dissertation research has two main goals to reduce the problem of distracted driving. The first goal is to quantify the impact of different secondary tasks on driving safety so that government officials could make informed decisions regarding the allocation of available resources for reducing distraction related roadway crashes. To achieve this goal, the following objectives are proposed:
1. Examine the relationship between the engagement in a secondary task and crash/near-crash likelihood;

2. Develop and compare several statistical and data mining models to estimate the increased crash/near-crash risk due to involvement in a particular secondary task;

The second goal aims to develop a real-time gaze-based algorithm for detecting driver visual distraction. To achieve this goal, the following objectives are proposed:

1. Create an adequate representation for driver attention allocation patterns in safety and non-safety critical events.

2. Investigate driver attention allocation patterns under different secondary task activities using the Naturalistic Engagement in Secondary Tasks (NEST) dataset.

3. Construct new distraction risk indicators to detect the level of driver visual distraction in real time;

4. Develop a crash prevention model that is capable of reducing distraction-related accidents and identify the environmental, vehicle, and sociodemographic factors affecting the crash/near-crash occurrence.

1.4 Dissertation Outline

This dissertation research has three distinct phases. Each phase will be presented in a separate chapter in addition to three other chapters that summarizes: the distracted driving literature review, the study data and general methodology, and finally the conclusions. The following paragraphs outline the organization of this research and present briefly the main outcomes of each chapter.

In Chapter 2, a comprehensive literature review is conducted to define the distracted driving problem and explain the methods and the findings concerned with distracted driving in
previous studies. This chapter also included a review for driver’s attention allocation process and the different methods associated with such process. At the end, the chapter located the research gaps and identified the research needs to better address the distracted driving problem.

Chapter 3 describes the data used in this research (SHRP2 NDS) and explains the general methodology followed in this dissertation. More Specifically, this chapter: (a) described how the study data was collected and presented, (b) listed the different data sources and formats, (c) identified the research key variables, and finally (4) discussed the general methodology followed in this research. The outcomes of this chapter would expect to help the readers to obtain the knowledge required to better understand the next chapters.

Chapter 4 represents the first phase in this study; determining the relative crash/near-crash risk associated with different distraction sources. In particular, this chapter employed SHRP2 NDS dataset to first confirm the relationship between distracted drivers and crash/near-crash likelihood using a multivariate probit model. Subsequently, two different techniques (one is a statistical based technique, and the other is a data-mining based technique) were implemented to quantify the increased crash/near-crash risk due to involvement in a particular distraction activity. A clustering model was then developed to place the secondary task activities into different crash/near-crash risk levels based on how risky they are. This chapter aims to identify sources of distracted driving and provide the risk associated with each source to either helping safety campaigns or providing transportation officials with the information needed to take informed decisions.

Chapter 5 demonstrates the second phase in this research. In this phase an adequate driver eye tracking approach was introduced to analyze and quantify driver attention allocations patterns. This approach was then used to construct two new distraction risk indicators using the Naturalistic
Engagement in Secondary Tasks (NEST) dataset. Statistical analysis were performed on each of these two indicators to test their significance in differentiating between distraction-related safety and non-safety critical events. In line with the research objectives, this chapter shows (a) how distracted drivers visually behave, (b) how to adequately represent and quantify driver visual behavior, and (c) how to develop robust distraction risk indicators.

Chapter 6 includes the third and the last research phase. In this Chapter, the framework for an advanced driver assistance warning system, that can alert distracted drivers if potential crash or near-crash is about to happen, is presented. In particular, an Artificial Neural Network (ANN) model is developed to predict the distraction-related safety critical events using NEST database. This phase also identified the risk factors that contributed the most to the ANN crash prediction model using various vehicle, roadways, and driver characteristics in addition to the new distraction risk indicators developed in Chapter 5.

Finally, Chapter 7 summarizes and concludes the research findings in each phase and also provides some recommendations for future work.
2 LITERATURE REVIEW

Driving is a daily, complex task that requires a driver’s full attention. Despite the complexity associated with this task, it is not uncommon to observe drivers performing other secondary tasks while operating a vehicle. These secondary tasks might include: reading a newspaper in slow moving traffic, making phone calls in preparation for a meeting that is about to take place, shaving to be ready for work, and discussing important topics with a passenger, among many others. While these tasks might seem trivial, they degrade the driving performance and increase the likelihood of a crash or near-crash event. Moreover, the technological features embedded in vehicles nowadays, in addition to the advanced wireless communication devices, have brought a new level of distraction to the driving environment (16). With that in mind, this section aims to summarize the existing knowledge about distracted driving and its effect on traffic safety. This section is divided into three main parts. Part one defines the distracted driving in addition to the findings and research methodologies used in distracted driving studies. Part two covers the driver attention allocation process topic and the methods used in analyzing this process. Part three discusses how previous studies measure the visual distraction behavior and the major limitations in such studies. Finally, a summary is given at the end of the chapter to locate the research needs according to the current knowledge.

2.1 Part 1: Driver Distraction

2.1.1 Driver Distraction Definition

Until recently there was no unified distracted driving definition among transportation researchers. The first attempt was made in 2001 when Ranney et al. (17) characterized driver distraction as follows:
• Driver distraction may be characterized as any activity that takes a driver’s attention away from the driving task;
• Any distraction from adjusting side mirrors, rolling down a window, or tuning a radio can contribute to a crash;
• Four main distraction types are identified, but more than one type can happen simultaneously:
  o Visual
  o Auditory
  o Physical (adjusting radio)
  o Cognitive (lost in thoughts)

After Ranney’s attempt, Stutts, Reinfurt, Staplin and Rodgmen (18) and Stutts et al. (19) stated that “distraction occurs when a compelling event, activity, object, or person shifts driver attention away from the driving task. Therefore, the presence of a compelling event distinguishes the distracted driving behavior from those who are lost in thoughts”. Accordingly, the only difference between this definition and Ranney’s definition was the exclusion of the cognitive distraction. One year later, Beriness, Simpson and Desmond decided to differentiate between the driver distraction and driver inattention definitions (20). They classified driver distraction as a part of the broader category of driver inattention, however they agreed with Stutts’s definition in which a triggering event exists to distinguish driver distractions.

In 2004, Green defined driver distraction as something that draws a driver’s attention from driving to a different task, object or direction (21). This definition is very similar to how Stutts defined the driver distraction; however, in Green’s definition he mentioned that attention is pulled away instead of being voluntarily shifted. Later in 2005, Tasca tried to come up with his own
definition of driver distraction based on the review of the past studies (22). He stated that driver distraction occurs when there is:

- “A voluntary or involuntary shift in a driver’s attention away from the driving task but not related to impairment (alcohol/drugs/fatigue…etc.)”;

- Shift in a driver’s attention occurs when the driver decides to:
  - perform a secondary task(s), or
  - or focus on a person, event or object that is not related to the driving tasks

- Driver inattention reduces driver situational awareness, which impairs his/her driving abilities and results in any of the following:
  - Crash
  - Near-Crash
  - Corrective action by the driver

Although Tusca succeeded in defining driver distraction appropriately, there is one issue that makes this definition different from the others. Tusca classified distracted driving if and only if the driving inattention results in a crash, near-crash or a corrective action by the driver. In other words, any drivers who did not cause any evasive actions, regardless of whether driver inattention existed or not, would not be classified as distracted driver.

The last attempt to agree on a common driver distraction definition was at the “Distracted Driving” conference that was held in Toronto, Canada in 2005. One of the main objectives in this conference was to agree on a suitable definition for “distracted driving” so that different research results could be compared. According to Hedlund’s study, who summarized the outcomes of this conference, driver distraction occurs when a competing task diverts a driver’s attention from the driving task to something else (23). This diversion may result either from inside or outside the
Hedlund also specified the after-effects of distraction in his definition. He mentioned that the “consequences are not necessarily an observable maneuver, but an increase in risk for untoward situations” (23). In 2006, the conference published the official definition of distracted driving as follows:

“Distraction involves a diversion of attention from driving, because the driver is temporarily focusing on an object, person, task, or event not related to driving, which reduces the driver’s awareness, decision-making, and/or performance, leading to an increased risk of corrective actions, near-crashes, or crashes”

In this research, driver distraction is defined when drivers engage themselves in a secondary task activity. This secondary task might result from inside the vehicle such as reaching for objects, cellphone texting...etc., or from multiple outside sources. Additional details regarding the secondary task activities will be presented later in Chapter 3.

2.1.2 Distracted Driving Findings and Methods in Previous Studies

This section presents an overview of some of the findings and the methodologies used in distracted driving research. In order to assess the impact that a distracted driving has on the crash risk, three different approaches are frequently used, namely; experimental studies (driving simulator studies), interview/survey studies, and observational studies (Naturalistic driving studies). Detailed discussion regarding the previous studies’ results in addition to the advantages and disadvantages of each study approach is provided in the following sections.

2.1.2.1 Experimental Studies (Driving Simulator Studies)

Experimental studies are usually performed in either a driving simulator or in a controlled traffic environment (24; 25). Driving simulators could be a high fidelity driving simulator with high degrees of freedom, or just a simple computer monitor with a chair in front of it as shown in
During the experiments, the participants are asked to perform a specific secondary task at a given time according to particular scenarios designed by the experimenter. The experimenter then collects the required driving data and analyzes it to determine the effect of distracted driving on driving behavior and traffic safety.

Figure 2.1. Driving simulators units.

Driving simulator studies have numerous advantages. First, driving simulator’s environments can be controlled. In other words, the experimenter can study the impact of distracted driving under different road scenarios without waiting for them to occur as in a natural environment. These scenarios might include different road geometry designs (vertical and horizontal alignment), different weather conditions (fog, rain, dry…etc.), different types of road lightening (illuminated or not), or a combination of those conditions among many others. Second, dangerous scenarios can be tested without exposing the participant’s life to any kind of risk. Third, it is more convenient and easier to evaluate the efficiency of new in-vehicle application warning systems in driving simulators than it is in the real world. Finally, high resolution detailed data could be obtained as opposed to other data collection methods.
Several studies have been conducted on distracted driving using driving simulators. Almen et. al, used a driving simulator to test how the reading task will affect driving behavior (26). They asked the participant to verbally report the numbers displayed on a computer screen located at the passenger seat. No significant outcomes resulted from this study as the authors faced some difficulties in creating the distraction activity artificially. In Karlsson’s study, a more complicated distraction activity is used to measure the impact of distracted driving on driving behavior (27). In this study, Karlsson used a sound alert to ask the driver to look at a screen inside the car. The task is then to select the right answer (Yes/No) based on a displayed matrix of arrows as shown in Figure 2.2. If an up arrow existed, the participant should press the “YES” button, otherwise, the “NO” button should be pressed. The experiment was conducted on 30 different participants but no significant results were obtained. The author reported that it was not easy to produce the artificial distraction activity, which affected the research outcomes.

![Figure 2.2. Karlsson's distraction activity.](image)

Later, Zhang, Smith, and Witt conducted an interesting study to observe the difference in driving behavior under a particular secondary task (28). The participants were asked to complete a puzzle shown on a screen but in a different way. More specifically, the participant was required to remember a character string displayed on a screen to match with another character string...
displayed on the current screen. The authors found significant differences in driving behavior under normal and distracted driving scenarios respectively, however, they recommended validation in a real world environment. Donmez et al. used the same procedures followed in Zhang’s experiment but with adding a two-stage warnings strategy as a distraction mitigation strategy (29). The warnings were given in either a colored stripped background displayed on a screen or LEDs installed on the car’s dashboard. The warning strategy was based on the driver’s off-road glance duration. The first warning was given when the driver’s off-road glance duration exceeds 2 seconds, while the second warning was given when the driver’s off road glance duration exceeds 3 seconds. The authors mentioned that the mitigation strategy was effective, and they did not face difficulties in inducing the distraction within the driving simulator.

Recently, Codjoe et al. performed an experiment to examine the distracted driving and the associated crash risks using the driving simulator at Louisiana State University (LSU) (30). Codjoe’s experiment had 67 participants who were tested under three distracting activity types; texting, passenger interaction, and cell phone conversation. The study found that texting and passenger interaction impaired driving performance, while no significant effects were recognized for cell phone conversation. The study was unable to make any statistical findings on the driving performance due to the limited sample size. Various other simulator studies have been conducted in the same manner, a more comprehensive overview of distracted driving experimental studies can be found in Caird’s study (24).

Although the experimental studies were successful in recognizing the degradation in driving performance due to the engagement in a secondary task, they were not helpful in making a valid estimate of the actual crash risk for two main reasons. First, participants do not decide where, when, and how to engage in a secondary task, which is not a real representation of real-
world secondary task involvement. Second, the transferability of the outcomes from the driving simulators to real life remains questionable. Thus, the experimental studies are not considered the best approach to determine the increased crash risk resulting from engagement in different secondary tasks.

2.1.2.2 Interview/Survey Studies

Interview studies are another approach in collecting secondary task information (8; 31-34). In these studies, information is gathered using telephone surveys as in McEvoy and Royal studies (32; 33) or online surveys as in Lansdown and Young studies (31; 34). In Sullman’s et al. study, a sample of 287 New Zealand drivers were asked about their cellphone use while driving and the perceived risk (8). The results showed that the percentages of drivers who never used a cell phone or used a cell phone occasionally while driving were 43% and 43%, respectively. The percentage of drivers who used the cell phone frequently was only 14%. While Sullman and Baas’s study was interested in cell phone use as the sole secondary task, the remaining four studies were concerned with all secondary tasks. In McEvoy’s study, the participants were asked to list all secondary tasks that lasted for 5 minutes or more over their last trip (32). Young et al. conducted another survey asking participants to report all kinds of secondary tasks and how often they engaged in them (34). The results of these studies are summarized and displayed as shown in Table 2.1. Recently, in year 2015, State Farm Insurance Company performed a survey study to examine the teenage drivers’ behaviors and attitudes towards distracted driving (35). One thousand teenagers between 16-19 years of age participated in this survey. The participants were asked to list all the secondary task activities they do while driving. Cell phone/ Smart phone usage, searching for music, interacting with GPS, and talking to passengers are the most common secondary tasks obtained from this survey. State Farm’s study also reported that teenage driver
belief that one of the important factors of a teen’s perception of distracted driving is the environment. The study mentioned that teen drivers prefer to use their cell phone during red light stops. Finally, the study showed that teenage drivers are fully aware of the consequences of the distracted driving behavior and they suggest the implementation of legislative, educational, and technological solutions to face the distracted driving problem.

Table 2.1. Percentages of secondary tasks involved in distraction related crashes to all crashes

<table>
<thead>
<tr>
<th>Secondary task</th>
<th>Telephone survey Driver Engaged (%) (33)</th>
<th>Telephone survey Driver Engaged (%) (32)</th>
<th>Online survey Driver Engaged (%) (31)</th>
<th>Online survey Driver Engaged (%) (34)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drink/Eat</td>
<td>49</td>
<td>18</td>
<td>51</td>
<td>80</td>
</tr>
<tr>
<td>Smoke</td>
<td>10</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clothing and body care</td>
<td>8</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Integrated devices</td>
<td>100</td>
<td>91</td>
<td>94</td>
<td></td>
</tr>
<tr>
<td>Other devices</td>
<td>66</td>
<td>7</td>
<td>1</td>
<td>41</td>
</tr>
<tr>
<td>Passenger-related</td>
<td>81</td>
<td>40</td>
<td>81</td>
<td></td>
</tr>
<tr>
<td>Other tasks</td>
<td>16</td>
<td>25</td>
<td>1</td>
<td>23</td>
</tr>
<tr>
<td>Internal tasks</td>
<td>69</td>
<td>69</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>Outside distraction</td>
<td>55</td>
<td>55</td>
<td>58</td>
<td></td>
</tr>
</tbody>
</table>

Despite the valuable information reported in these studies, these studies’ outcomes have been criticized for two main data-related reasons. First, it is reasonable to question whether or not the study sample is a good representation of the driving population. For example, much of the studies use home telephone numbers or web pages in collecting distracted driving information, which restricts the sample to a particular type of drivers (drivers who have home telephone or internet access). Second, self-reporting bias is a major concern. People prefer to under-report themselves, which has the direct effect on the survey outcomes. This might be because of secondary task engagement is socially unacceptable behavior. Third, the outcomes of survey studies vary significantly and this is due to the different methodologies used in asking questions.
2.1.2.3 Naturalistic Driving Studies (NDS)

Naturalistic Driving Studies or Observational studies are the most realistic approach in gathering distracted driving data (36-39). In these studies, vehicles are equipped with advanced data collection devices to record the normal driving behavior — see Figure 2.3. These data are then extracted from the equipped vehicles and reduced via data reductionist analysts for further research. Naturalistic driving studies vary significantly in terms of a data collection time frame. Most of the existing naturalistic driving studies were performed over a short time period (from only a single-drive up to several week long drives) (39). Only one extreme study, the 100-car project, was performed over a one-year period which has been considered the largest NDS project until year 2012 (37). In this study, the authors analyzed the distracted driving and its relationship to SCEs. For this one-year project, only 82 crash events and 761 near crash events were observed. Based on these SCEs, a stratified non-SCEs database was created. It is worth mentioning that for each non-SCE, data variables were recorded for 6-second durations while the vehicles maintained a speed of 5 mph as a minimum, while data variables were recorded for 30-second durations for each SCE. The SCE’s dataset and the non-SCE’s dataset were then collectively analyzed as a part of the 100-car naturalistic driving study project to identify the relative frequency of different secondary tasks during SCEs and non-SCEs. The results indicated that embedded passenger interaction devices in a vehicle, and manipulating objects are the most common secondary tasks among drivers. In general, the outcomes revealed many useful findings regarding the SCEs. However, due to the short coding time span for non-SCE (only 6 seconds), it was difficult to compare the SCE behavior to that of non-SCE. As a result, the authors recommended collecting additional log data in the future to overcome this problem. Moreover, secondary tasks such as eating/drinking, smoking-related, clothing/body care, integrated devices, passenger-related,
outside distractors, and other in-vehicle devices are listed as the most frequent secondary tasks in the other three studies (36; 38; 39).

Although research approaches derived from NDS data are considered more realistic, there are some difficulties facing these types of studies. These studies aim to gather traffic incident data in addition to normal driving data as well. Due to the rare nature of crash events and the ability to obtain realistic crash/near-crash risk estimates, a high number of equipped vehicles in addition to long observation periods are required. Therefore, in this research the second Strategic Highway Research Program 2 Naturalistic Driving Study (SHRP2 NDS), the largest naturalistic driving data to date, is employed to accomplish the research goals. Additional details regarding SHRP2 naturalistic driving study database are presented later in Chapter 3.

Figure 2.3. Naturalistic driving study example.
2.2 Part 2: Driver Attention Allocation Process

2.2.1 Driver Inattention Definition

Similar to driver distraction, there are various definitions associated with driver inattention. For instance, Lee et al. defined the inattentive driving as “diminished attention to activities critical for safe driving in the absence of a competing activity” (3). Victor et al. defined driver inattention as “improper selection of information, either a lack of selection or the selection of irrelevant information” (40). Driver inattention and driver distraction are always presented in observational studies. In one of the crash studies, driver inattention was defined as occurring “when the driver’s mind has wandered from the driving task for some non-compelling reason” such as when the driver is “focusing on internal thoughts” and not giving attention to the driving task. In the 100-car study, inattentive driving was defined as “any point in time that a driver engages in a secondary task, exhibits symptoms of moderate to severe drowsiness, or looks away from the forward roadway” (37). This definition is adopted for the rest of this dissertation research.

2.2.2 Methods and Findings Concerned with Driver Attention Allocation Process

2.2.2.1 Single Focal Point Approach

Driver attention is not a variable that can be measured directly while driving. Therefore, creating an appropriate method to examine the attention allocation is challenging. Nonetheless, the literature shows some attempts to represent the driver attention allocation process and analyze its relationship with traffic safety. Throughout the literature, some studies examined the attention allocation by using a single focal point approach (41-45). More specifically, these studies concentrated on analyzing the frequency and the duration where the driver shifts his attention from his Fields of Relevant Driving (FRD) to another specific location. The outcomes of these studies
have shown that the amount of time the driver spent focusing on a specific area/object is directly proportional to the importance of this area. The longer the driver’s eye glances are, the more the information the driver is expected to acquire. Since the one focal point approach was not successful in describing the attention allocation process that requires multiple focal points, researchers moved onto the scan path approach to represent the driver’s visual attention efficiently.

2.2.2.2 Scan Path Approach

The scan path approach is simply a method that describes the focal points where the driver diverts his glance into, keeping the focal points sequence in order. This method can provide extra information by describing the whole driver shifting attention allocation process that is required to maintain adequate situational awareness. According to Underwood and Wong studies, these studies have shown that most of the scan paths are either shifting away from the forward area or returning to the forward area (45; 46). Therefore, the forward area is usually presented as the most attractive focal point in the driver attention allocation process. Although this method succeeded in addressing the single focal point approach problems, it has serious drawbacks that might lead to incorrect conclusions. First, the scan path approach did not account for the driver glance duration. In other words, drivers might share the same exact scan path but the time allocated to the focal points within that path is different. For example, if a scan path included two focal points - a forward glance and rearview mirror glance -, this method will aggregate all the drivers who conducted this path regardless of the time that each driver allocates or spends glancing in the forward or rearview mirror. So if a particular driver spends two seconds and one second focusing on forward and rearview mirrors, respectively, this driver will be treated similarly to another driver who repeated the same scan path frequently but with less time intervals (0.2 seconds and 0.3 seconds repeatedly). Thus, this method does not show the real distribution of attention allocation
patterns. Consequently, better approaches are needed to discover the driving behavior especially under distracted driving conditions.

2.2.3 Distracted Driving and Driver Inattention Relationship

There is a wide belief that distracted driving and driver inattention are related to each other. For example, in Victor’s et al. definition, driver distraction is defined as “the inappropriate selection of information to the extent that safety-relevant information is missed. Thus, distraction is here defined as a subset of inattention, referring to all instances when attention is misallocated, but excluding cases when attention is not allocated at all” (40). Stutts et al. also defined distraction as “the presence of a triggering event that distinguishes distraction from other forms of driver inattention” (39). Furthermore, Pettitt et al. stated that “the result of distraction is inattentive driving. However, inattention is not always caused by distraction”, which obviously relates the distraction to driver inattention but does not restrict driver inattention to driver distraction solely (47). Based on these studies, distracted driving and driver inattention are correlated. Accordingly, quantifying the level of inattentive driving while the driver is performing a secondary task activity is important and considered a challenging task. In this research, a new approach will be proposed to analyze driver attention allocation process and quantify the level of distraction associated with each distraction related driving event.

2.3 Part 3: Distracted Driving Detection

Essentially, driving is a visual-physical task. This requires drivers to keep their eyes on the road all the time while s/he is operating the vehicle. However, it is not uncommon to see drivers glancing away for long periods which in turn can have serious consequences for traffic safety. Hence, eye glance variables such as eye glance duration, eye glance frequency, and eye
glance history among many others are important metrics for measuring distracted driving behavior. These metrics record the information related to how often and how long a driver is looking on-road and off-road respectively. These metrics are also very sensitive to the visual demands while driving, which can help in identifying distracted driving if it exists. Several studies have been conducted using these metrics. For instance, Zhang et al., used the average duration of driver off-road glances in a 3-second moving time window to identify distracted drivers (48). Donmez et al., used the current glance characteristics in addition to the average glance duration in a 3-second sliding window for the same purpose (49). Victor used the same methodology as in Zhang’s study but in a 60-second sliding window (50). In these studies, it was found that off-road glance duration is a good measure of visual distraction. With a slightly different approach, Kaelsson introduced a buffer index that starts once the driver looks away from the road and if this buffer reaches a specific value, the driver is regarded as being distracted (27).

Although the above mentioned studies succeeded in examining the eye glance behavior and driver distraction, there are some limitations that need to be addressed. First, the findings presented in these studies all resulted from well-controlled driving simulator experiments. Thus, the validity of the results remains questionable. Second, past studies assumed a linear relationship between visual attention and distraction level. However, this relationship could be explained in a more complex model. For example, Wirewille and Engstrom studies described this relationship in an exponential function which results in acceptable outcomes but not as it was anticipated (51; 52). Thus advanced, realistic, adequate, and more robust techniques can benefit the detection of driver visual distraction. Nowadays, naturalistic driving datasets provide continuous eye movement recordings in real world environment. Therefore, there are no longer restrictions in collecting eye glance data in a real environment. These datasets provide ample opportunities to
not only analyze the process attention allocation process but to detect driver distraction in real time and quantify the level of driver distraction as well. However, this is a very challenging task.

2.4 Limitations in Previous Studies

Although previous studies succeeded in determining the impact of different devices and systems on driving behavior, these attempts are insufficient (7-11). These attempts used data collected from different sources such as: labs, test tracks, simulators, surveys, and crash databases, to study the distracted driving behavior. However, these types of data have some limitations. First, in crash databases, crash information is always collected by police officers in a post hoc interview (53). Detailed driving data that have been recorded prior to crashes and near crashes are hard to capture from crash databases. However, collecting pre-crash information is essential to help transportation safety researchers establish the relationship between drivers’ behavior and SCE occurrence. Second, since distracted driving is not a socially acceptable behavior, drivers do not admit committing such act when they are questioned. Having that said, surveys and interview studies are also inadequate sources of distraction data because these studies usually suffer from self-reporting bias (54-56). Moreover, in simulator and test track studies, limited number of distracting activity types is usually tested. Studies that test a wide variety of distracting activities on same drivers are limited and this is due to data collection complexity. Sample size is also a major concern in each of these kinds of studies (24). Therefore, to better understand the crash risk, the distracted driving and its effect on the driving performance need to be studied in a larger context of the driving environment. Naturalistic driving studies, such as SHRP 2 NDS, can help fill the gaps between experimental studies and crash studies by collecting the required data to estimate the crash risk as in crash database studies while still collecting driving behavior and driving performance data. Furthermore, previous studies did not report exactly how crash risk
would increase or decrease when a particular secondary task activity took place. Finally, a substantial proportion of research in this area has focused only on cellphone-use related activities and ignored other secondary task activities. Therefore, more research is needed to estimate the increased crash/near-crash risk due to driver involvement in other secondary tasks as well.

In addition, distracted driving and the associated eye glance behavior have long been studied as mentioned previously. Different methodologies have been proposed by different researchers (13-15). However, more comprehensive approaches are needed to better describe the attention allocation process and the associated eye glance distribution. Reliability and validity of many of these eye tracking algorithms are uncertain and extracted from small experimental studies. This is due to the complex data collection requirements needed to obtain appropriate eye glance data in a real world environment. Nowadays, the improvement in eye tracking technology facilitates the data collection process and enables researchers to validate their eye tracking algorithms realistically. Naturalistic driving studies such as SHRP2 NDS, in which driver behavior is monitored during their regular commutes, provide ample opportunity to better understand distracted driving and eye glance behavior in a real world environment.

To fill these gaps, this dissertation research is an attempt to investigate the increased crash risk resulting from different secondary tasks using real world data. This research also aims to develop a new distraction detection system that is able to detect driver’s visual distraction in real time. To this end, the major contributions of this dissertation include: (1) applying new data mining algorithms to quantify the degree of the increased crash risk for different secondary tasks, (2) comparing the outcomes of these data mining crash risk assessment models with the traditional statistical models to identify the riskiest secondary tasks, (3) developing a new approach that is capable of analyzing the driver attention allocation process adequately and detecting the driver
distraction, and (4) formulating a new distraction level index that quantifies the level of driver
distraction in real time manner.
3 DATA FORMATION AND GENERAL METHODOLOGY

3.1 SHRP2 NDS Overview

In this dissertation research, SHRP2 NDS dataset is employed to achieve the research objectives. SHRP2 NDS dataset is considered the largest naturalistic driving study project that has been conducted in the US to date. This dataset contains trips from more than 3000 drivers, aging between 16-80 years old, located in six different states. This includes 239 vehicles from Indiana, 256 from Pennsylvania, 698 from Florida, 719 from New York, 504 from North Carolina, and 676 from Washington (as shown in Figure 3.1) (57).

![Figure 3.1. SHRP2 NDS data collection sites.](image)

Inside each vehicle, a data acquisition system was installed to continuously record daily driving data. Recorded data included: vehicle dynamics, video front and rear views, driver’s face and hands…etc. Figure 3.2 shows the data acquisition systems designed and installed in each participant’s vehicle. Additional details about the data acquisition system is provided at SHRP2 NDS official website — "InSight” website — and Campbell’s report (58; 59). InSight website is
a webpage that was developed to facilitate the use of SHRP2 NDS database for transportation researchers. This website was designed to allow requesting the data online and to share transportation researchers’s thoughts online. Figure 3.3 shows the InSight website user interface.

Figure 3.2. Data Acquisition System (DAS) — InSight website.

Figure 3.3. SHRP2 NDS website (InSight website).
3.2 Data Format

SHRP2 NDS data were collected and processed in house by Virginia Tech Transportation Institute (VTTI). The database was then requested and delivered by VTTI in the following format:

- Event detailed data: A CSV file that summarizes the full contents of the SHRP2 safety and non-safety critical events videos. This file also includes some sociodemographic variables that characterize the subject driver in each driving event;
- Time-series data: CSV files with all data stored in the data acquisition system for each driving event, such as: speed, acceleration, yaw rate,…etc.;
- Driver survey questionnaire: A CSV file that lists the sociodemographic characteristics for each participant, such as: gender, age group, education, marital status …etc. (Appendix C);
- Driving history questionnaire (Appendix D); and
- A data dictionary spreadsheet;

3.3 Event Detailed Data

Event detailed data are a set of variables that record vehicle, roadway, and driving conditions during the event happening time. These data include variables such as: event type, event severity, traffic flow, weather and surface conditions, traffic density, vehicle type, and whether a secondary task existed or not among many other variables. Table 3.1 lists all the event data variables delivered from VTTI and coded in the event data file. The table also displays the definition of each variable, the variable type, and number of categories coded for each variable.
Additional details for the displayed event detailed data variables is provided on the InSight website (59).

3.3.1 **Key Variables in Event Detailed Data**

In this section, only key variables in the event detailed data will be defined. Among the key variables used in this research is the event severity. Event severity is a general term that describes the event’s outcome where the outcome is denoted as crash, near-crash, or baseline. Crash is defined as “Any contact that the subject vehicle has with an object, either moving or fixed, at any speed in which kinetic energy is measurably transferred or dissipated” (58). Near-crash is defined as “Any circumstance that requires a rapid evasive maneuver by the subject vehicle or any other vehicle, pedestrian, cyclist, or animal to avoid a crash”. Whereas baseline represents an event that is neither a crash nor a near-crash. Additional hints and examples are provided on the InSight website (59).

The second key variable is the precipitating event. According to SHRP2 NDS creation report, precipitating event is defined as “The state of environment or action that began the event sequence under analysis” (58). Examples of precipitating events are; running red light, sudden brake of lead vehicle, pedestrian crossing…etc. Previous studies have found that driving behavior beyond the onset of the precipitating event cannot be considered as a normal driving behavior. Thus, all data collected after the precipitating event should be excluded while analyzing driver’s normal behavior.

The third and last key variable in the event detailed data is the secondary task(s). Secondary tasks variables list the type of the distraction activities where the driver is engaged in during the event happening time. Secondary task variables are coded for every 10 seconds interval.
If the subject driver is involved in more than one secondary task during the same 10-sec period, data reductionists will select the most critical activity that directly affect the event.

3.4 Time-Series Data

Time-series data are a set of variables that record the dynamics of the vehicle in high resolution quality (frequency 10 Hz), such as: vehicle’s speed, vehicle’s acceleration, locations, steering wheel position, and yaw rate among others. Figure 3.4 and Figure 3.5 show a sample of speed and acceleration profiles stored in each time-series file. Time-series data also include some driver behavior variables such as: driver’s gaze location, hands on the wheel, and detailed secondary task. These variables are also coded with the same data resolution (10Hz). Length of time-series data recordings depends on the event type. Data reductionists record 30 seconds of data for safety critical events (crashes and near-crashes), while only 6 seconds are recorded for non-safety critical events. Table 3.1 list some of the time-series data variables delivered by VTTI and reported in the time-series data files. Detailed description for the displayed time-series data variables is available on the InSight website (59).
3.4.1 Key Variables in Time-Series Data

Among the time-series data variables delivered by VTTI, two variables are very important for the rest of this dissertation research, namely; detailed secondary task, and driver eye glance.
location (gaze location). The following parts will define and discuss both variables in further details.

Detailed secondary task is a variable that shows the type of the secondary task in which the driver is engaged at the current moment. This variable is coded for every 0.1-sec change in time and included more than 15 secondary task activities. Examples of these activities are: manipulating objects, personal hygiene, talking/listening on hand-held cell phone, and eating/drinking among many others. Additional information about this variable will be presented later in Chapter 5.

Driver eye glance location is another key variable that indicates where the driver gaze is directed. This variable was also coded every 0.1-sec by a VTTI data reductionist who reviewed the video data on a frame by frame basis. The review included the following set of locations: “cell phone”, “center stack”, “instrument cluster”, “interior object”, “passenger”, “left forward”, “left mirror”, “left window”, “rearview mirror”, “right forward”, “right mirror”, “right window”, “eye closed”, “no video”, and “forward”. Detailed discussion on the eye glance variable is also provided later in Chapter 5.
Table 3.1. SHRP2 Naturalistic driving variables

<table>
<thead>
<tr>
<th>Var. name</th>
<th>Var. type</th>
<th>#of Categories</th>
<th>Description/Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Event data</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seatbelt use</td>
<td>Categorical</td>
<td>4</td>
<td>Lap/shoulder belt, Lap only, Shoulder only, None used</td>
</tr>
<tr>
<td>Number of passengers</td>
<td>Categorical</td>
<td>4</td>
<td>0, 1, 2, 3+</td>
</tr>
<tr>
<td>Weather</td>
<td>Categorical</td>
<td>6</td>
<td>Fog, Mist or Light Rain, Heavy Rain, Snowing, other, No adverse condition</td>
</tr>
<tr>
<td>Driver behavior1,2,3</td>
<td>Categorical</td>
<td>19</td>
<td>“Driving behaviors made by the driver during the event. Behaviors may be apparent at times other than the time of the precipitating factor, such as aggressive driving at an earlier moment which led to retaliatory behavior later. Subsequent inappropriate or illegal behaviors are labeled DriverBehavior2 and DriverBehavior3.”</td>
</tr>
<tr>
<td>Driver impairments</td>
<td>Categorical</td>
<td>7</td>
<td>Drowsy, Drugs, other illicit drugs, Impaired due to previous injury...etc.</td>
</tr>
<tr>
<td>Secondary task</td>
<td>Categorical</td>
<td>11</td>
<td>“Observable driver engagement in any of the listed secondary tasks during the 10s of the event (if available): Cell phone interaction, Adjust/Monitor embedded device, Passenger interaction Reaching for object…etc.”</td>
</tr>
<tr>
<td>Road surface condition</td>
<td>Categorical</td>
<td>4</td>
<td>Dry, Icy, Snowy, Wet</td>
</tr>
<tr>
<td>Traffic flow</td>
<td>Categorical</td>
<td>4</td>
<td>Divided, Undivided, One-way traffic, No lanes</td>
</tr>
<tr>
<td># of Travel lanes</td>
<td>Categorical</td>
<td>9</td>
<td>0,1,2,3,…, 8+</td>
</tr>
<tr>
<td>Traffic density</td>
<td>Categorical</td>
<td>6</td>
<td>LOS A, B, C, D, E,F</td>
</tr>
<tr>
<td>No traffic control</td>
<td>Binary</td>
<td>0/1</td>
<td>indicates whether the participant was not subject to or influenced by any traffic controls during 10s of the event</td>
</tr>
<tr>
<td>Stop Sign</td>
<td>Binary</td>
<td>0/1</td>
<td>indicates whether the participant was subject to or influenced by at least one stop sign or not</td>
</tr>
<tr>
<td>Traffic signal</td>
<td>Binary</td>
<td>0/1</td>
<td>indicates whether the participant was subject to or influenced by at least one traffic signal or not</td>
</tr>
<tr>
<td>Merge sign</td>
<td>Binary</td>
<td>0/1</td>
<td>indicates whether the participant was subject to or influenced by at least one merge sign or not</td>
</tr>
<tr>
<td>Road alignment</td>
<td>Categorical</td>
<td>3</td>
<td>Curve left, Curve right, Straight</td>
</tr>
<tr>
<td>Road grade</td>
<td>Categorical</td>
<td>4</td>
<td>Grade up, Grade down, Hillcrest, Level</td>
</tr>
<tr>
<td>Locality</td>
<td>Categorical</td>
<td>12</td>
<td>Business industrial, Church, Construction zone, Urban, School, Interstate…etc.</td>
</tr>
<tr>
<td>Lighting</td>
<td>Categorical</td>
<td>5</td>
<td>Daylight, Dawn, Darkness lighted, Darkness not lighted, Dusk</td>
</tr>
<tr>
<td><strong>Driver Sociodemographic Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>Categorical</td>
<td>16</td>
<td>16-19, 20-24, 25-29, 30-34,…, 85-89</td>
</tr>
<tr>
<td>Gender</td>
<td>Binary</td>
<td>0/1</td>
<td>Male/Female</td>
</tr>
<tr>
<td>Annual miles</td>
<td>Categorical</td>
<td>7</td>
<td>&lt;5k, 5k-10k, 10k-15k, …,&gt;30k</td>
</tr>
<tr>
<td>#of violations</td>
<td>Categorical</td>
<td>3</td>
<td>0,1,2+</td>
</tr>
<tr>
<td>VMIScore</td>
<td>Categorical</td>
<td>3</td>
<td>Impairment level found from Visualizing Missing Information test results; none, mild, serious</td>
</tr>
<tr>
<td>UFOVScore</td>
<td>Categorical</td>
<td>3</td>
<td>Impairment level found from UFOV test results; none, mild, serious</td>
</tr>
<tr>
<td>Clock Drawing Score</td>
<td>Categorical</td>
<td>6</td>
<td>Perfect, Minor visuospatial errors, Inaccurate time, good visuospatial, Inaccurate time, minor visuospatial errors, Moderate visuospatial errors, Severe visuospatial errors, No reasonable representation of a clock</td>
</tr>
<tr>
<td><strong>Event Time-series</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eye Glance Location</td>
<td>Categorical</td>
<td>10</td>
<td>Forward, Left windshield, Right Windshield, Left mirror, Right mirror, cell phone, passenger…etc.</td>
</tr>
<tr>
<td>Hands on Wheel</td>
<td>Categorical</td>
<td>6</td>
<td>No Hands on, one Hand on one off, Both hands on, …etc.</td>
</tr>
<tr>
<td>Vehicle Speed</td>
<td>Continuous</td>
<td></td>
<td>Vehicle speed at 0.1 sec. frequency</td>
</tr>
<tr>
<td>Vehicle Acceleration</td>
<td>Continuous</td>
<td></td>
<td>Vehicle acceleration in z, y and z directions</td>
</tr>
<tr>
<td>Detailed Secondary task</td>
<td>Categorical</td>
<td>14</td>
<td>Manipulating objects, Talking/Singing to passenger, Reaching for object,…etc.</td>
</tr>
</tbody>
</table>
3.5 Driver Questionnaire

To link the drivers’ characteristics with the driving data obtained from SHRP2 NDS, driver questionnaire data was provided by VTTI (Appendix C, D). Driver questionnaire data include some sociodemographic and driving history variables. Examples of these variables are: driver’s age, gender, average annual miles, and years of driving among many others. Figure 3.6 shows the driver’s age distribution in SHRP2 NDS database. The rest of driver questionnaire variables are listed in Table 3.1.

![Figure 3.6. Drivers age distribution in SHRP2 NDS database.](image)

3.6 Naturalistic Engagement in Secondary Tasks Dataset (NEST)

NEST is a dataset that is mainly concerned with distraction-related driving events. NEST dataset is developed as a subset of the SHRP2 NDS database. The NEST dataset provides detailed information about 400 distraction-related SCEs along with a balancing sample of 800 non-SCEs.
These driving events are collected from 204 drivers with multiple vehicle types, various age groups, and different annual miles traveled. Figure 3.6, Figure 3.7, Figure 3.8, and Figure 3.9 display the participants’ gender, age, vehicle type, and the annual miles traveled distributions included in the NEST dataset respectively. It should be noted that in this study, crash and near-crash driving events are referred to as Safety Critical Events (SCEs), whereas non-crash or normal driving events are referred to as non-Safety Critical Events (non-SCEs). Each SCE is combined with a balancing sample of 3 to 4 non-SCEs. For each driving event (SCE or non-SCE), two types of data are collected: time series and event summary.

Figure 3.7. Gender distribution in NEST dataset.
Figure 3.8. Age distribution in NEST dataset.

Figure 3.9. Vehicle classification in NEST dataset.
The time-series data includes sixteen different variables describing vehicle performance, glance behavior, and secondary task engagement while driving. The time-series data in NEST dataset are collected for longer observation periods compared to SHRP2 NDS time-series data (20-sec for non-SCE and 30-sec for SCEs compared to 6-sec for non-SCE and 30-sec for SCEs in SHRP2 NDS respectively). Every SCE is divided into two parts, 20 seconds and 10 seconds, split by a time marker at a point representing the onset of a precipitating factor. A precipitating factor is defined as the triggering factor of a SCE (defined in section 3.3.1). The glance behavior variable included in this dataset indicates the locations where the driver gaze is directed. This variable was coded for each 0.1-sec by a VTTI data reductionist who reviewed the video data on a frame by frame basis. This variable included the same glance locations defined in SHRP2 NDS dataset and defined in section 3.4. On the other hand, the event summary data describe the vehicle and environmental conditions during an event, such as traffic density, flow, intersection influence, and road surface condition among many other variables.
3.7 General Methodology

This research aims to reduce the problem of distracted driving from two different perspectives. The first perspective is to identify possible sources of distraction so that government officials can make informed decisions regarding the allocation of available resources to reduce distraction related roadway crashes. The second perspective is to actively counteract driver distraction phenomenon by other stimuli presented to the driver. Considering the fact that drivers might be already distracted, the idea in the second perspective is to alert drivers to divert their attention back again to the road. To achieve that, a robust distraction detection system should be developed and validated using real world driving data. Figure 3.11 shows the research framework that will be followed in the rest of this dissertation to accomplish the research goals.
This dissertation research has three distinct phases (the three grey boxes shown in Figure 3.11). The objective in the first phase is to assess the impact that a particular secondary task has on the crash/near-crash risk. This will be achieved with the following steps:

1. A multivariate model will be constructed to examine the correlation between the engagement in a secondary task and the crash/near-crash likelihood. Since the occurrence of both distraction and crash may depend on various explanatory variables including driver’s demographics characteristics, vehicle characteristics, and roadway characteristics, the multivariate approach is chosen to link these two variables.
to those explanatory variables. In particular, a bivariate Probit model will be constructed to identify the factors affecting these two responses and also capture the correlation between them.

2. Based on the outcomes in step 1, two different models will be developed to quantify the increased crash or near-crash risk that results from the different secondary tasks. The two models will estimate the relative risk of the secondary tasks from two different perspectives: a traditional statistical modeling perspective, and a data mining modeling perspective. The models’ results will be presented and the merits of each model will be displayed in Chapter 4.

3. Since some secondary tasks might have similar crash risk impact, similar secondary tasks will be grouped together, based on a clustering algorithm, to identify the high crash risk secondary tasks.

These outcomes can help drivers understand the relative risk associated with the various secondary task activities so that they can adjust their behavior or consider other alternatives. It can also help legislators initiate laws that reduce the crashes resulting particularly from distracted driving. Finally, it can help government officials make informed decisions regarding the allocation available resources to reduce roadway crashes and improve traffic safety.

On the other hand, developing real-time algorithms that are intended to detect driver distraction is challenging. These algorithms are usually designed based on real-time measurements recorded during commute driving. Since SHRP 2 NDS data provide a mean to access measures registered in real-time while driving (for example, eye gaze movements, hands on the wheel…etc.), this research will use these data in an attempt to detect driver distraction and develop new distraction risk indicators that can help in establishing an effective distraction countermeasure system. This research objective will be presented and discussed in details in Chapter 5.
The last research phase (last grey box in Figure 3.11) in this dissertation aims to develop a crash/near-crash prevention model. In regards to this aim, an artificial intelligent model will be constructed to predict distraction-related SCEs given the vehicle, event, sociodemographic, and the new measures extracted from the previous phase (phase 2). The model will be trained, tested and validated using the NEST dataset. The proposed model will then be evaluated and the predictability power of each input variable will be presented in Chapter 6.
4 CRASH/NEAR-CRASH RISK ASSESSMENT OF DISTRACTED DRIVING AND ENGAGEMENT IN SECONDARY TASKS

4.1 Introduction

Driving is a daily complex task that requires a driver’s full attention. Despite the complexity associated with this task, it is not uncommon to observe drivers perform other secondary tasks while operating a vehicle. These secondary tasks might include reading newspaper at slow moving traffic, shaving to be ready for work, and discussing important topics with a passenger, among many others. While these tasks might seem trivial, they degrade the driving performance and increase the likelihood of a crash or near-crash event. Moreover, the technological features embedded in vehicles nowadays, in addition to the advanced wireless communication devices, have brought a new level of distraction to the driving environment. Thus, it is important to estimate the relative crash/near-crash risk for better understanding of the effect of different types of secondary tasks on the driving performance and traffic safety.

The literature review shows evidence of a relationship between engagement in a secondary task and crash likelihood (61). However, there are some limitations that need to be addressed. First, most of the previous statistical models did not take into account the correlation between interrelated variables, such as engagement in a secondary task and the crash likelihood. More specifically, in previous statistical models, the multiple dependent variables are modeled separately, each as a function of a set of independent variables, and therefore, the correlations among the dependent variables were ignored. Second, recent data mining techniques have captured researchers’ attention as they outperform traditional modeling techniques. In this context, the objectives in this chapter are to (a) construct a statistical model that considers the correlation between engagement in a secondary task and the crash/near-crash occurrence (bivariate probit...
(b) estimate valid crash risk measures for different types of secondary tasks using the largest and most representative naturalistic driving dataset (baseline-category logits model); and (c) offer a new methodology for investigating the relationship between the different secondary tasks and the crash/near-crash risk using a new data-mining technique (association rule mining model).

The rest of this chapter is organized as follows. Section 4.2 describes the data used in this chapter. Section 4.3 presents model development, detailing the bivariate probit model and results, and the two alternative models used to quantify the increased crash-risk estimate using the baseline category logits model, and the association rule mining model, respectively. Section 4.4 discusses the models’ outcomes and their implications. Finally, conclusions are provided in Section 4.5.

4.2 Data Description

To achieve the chapter objectives, only the distraction-safety related variables in SHRP2 NDS database are considered in the rest of this chapter analysis. These variables are either related to driver’s engagement in a secondary task or crash/near-crash likelihood, and are selected based on previous distraction-safety-related studies (37; 62; 63). Table 4.1 lists all the different variables used in this chapter.
Prior to the models’ development, the SHRP2 NDS dataset was reduced to remove any biases that might have affected the crash risk estimates. First, a crash event that did not involve injuries or property damage was excluded from the final dataset. Second, driving events were filtered out to remove any events associated with observable driver impairment. Finally, driving records with missing driver information were excluded; leaving 905 crashes, 2,558 near-crashes, and 18,544 baseline events as a final dataset. It should be noted that the “Secondary Task” variable is the key variable in the rest of the chapter. This variable shows the type of the secondary task in which the driver was engaged prior to the crash/near-crash time or during the selected normal driving event. The type of the secondary task was manually coded by reviewing video by VTTI according to the SHRP 2 data dictionary. If no secondary task existed, the variable showed
the “No Secondary Task” outcome. In this chapter, the secondary task activities were classified according to Stutts et al.’s study and as shown in Table 4.2 (39).

Table 4.2. Secondary tasks classification.

<table>
<thead>
<tr>
<th>Secondary task type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eating/drinking</td>
<td>Eating with utensils, Eating without utensils, Drinking with lid and straw, Drinking from an open container…and so forth</td>
</tr>
<tr>
<td>Smoking</td>
<td>Smoking cigar/cigarette, Lighting cigar/cigarette, Extinguishing cigar/cigarette</td>
</tr>
<tr>
<td>Passenger interaction</td>
<td>Passenger in adjacent seat - interaction, Passenger in rear seat - interaction</td>
</tr>
<tr>
<td>Manipulating objects</td>
<td>Object dropped by driver, Object in vehicle, other, and so forth</td>
</tr>
<tr>
<td>Reaching for objects</td>
<td>Reaching for food-related or drink-related item, Reaching for cigar/cigarette, Reaching for personal body-related item, and so forth</td>
</tr>
<tr>
<td>Vehicle integral devices</td>
<td>Adjusting/monitoring climate control, Adjusting/monitoring radio, Inserting/retrieving CD (or similar), and so forth</td>
</tr>
<tr>
<td>Personal hygiene</td>
<td>Combing/brushing/fixing hair, Applying make-up, Shaving, Brushing/Flossing teeth, and so forth</td>
</tr>
<tr>
<td>Outside distractors</td>
<td>Looking at previous crash or incident, Distracted by construction, Looking at pedestrian, and so forth</td>
</tr>
<tr>
<td>Other Secondary Tasks</td>
<td>Other non-specific internal eye glance, Other known secondary task, Unknown type (secondary task present)</td>
</tr>
<tr>
<td>Dancing</td>
<td></td>
</tr>
<tr>
<td>Reading</td>
<td></td>
</tr>
<tr>
<td>Writing</td>
<td></td>
</tr>
<tr>
<td>Pet in vehicle</td>
<td></td>
</tr>
<tr>
<td>Cell phone, talking/listening hand-held</td>
<td></td>
</tr>
<tr>
<td>Cell phone, talking/listening hands-free</td>
<td></td>
</tr>
<tr>
<td>Cell phone, texting</td>
<td></td>
</tr>
<tr>
<td>Cell phone, dialing hand-held</td>
<td></td>
</tr>
<tr>
<td>Cell phone, locating/reaching/answering</td>
<td></td>
</tr>
<tr>
<td>Cell phone, other</td>
<td></td>
</tr>
<tr>
<td>No secondary tasks</td>
<td></td>
</tr>
</tbody>
</table>

According to the NHTSA, distracted driving is responsible for 30% of all crashes (64). In general, distraction has many sources. Engagement in a secondary task while driving is one of the major sources of distraction. Thus, in this research, distracted driving is defined as driver engagement in a secondary task. In most of the previous studies, the responsibility of distracted driving as a main cause of accidents was measured by descriptive statistics (for example, mean, standard deviation, chi squared test, etc.) (37; 62). Despite the significant correlation between distracted driving and crash likelihood, these descriptive statistical analyses cannot clearly identify the association among multiple factors in complex relationships. Therefore, this chapter proposes
a new methodology to identify the correlation among responses that are made simultaneously using a discrete choice model.

4.3 Models Development

4.3.1 Bivariate Probit Model

One of our objectives is to predict the crash/near-crash likelihood given that the driver is distracted, that is, engaged in a secondary task. Since the occurrence of distraction and crash may both depend on various explanatory variables including the driver’s demographics characteristics, vehicle characteristics, and roadway characteristics, the multivariate approach was chosen to link these two variables to the explanatory variables. In particular, a bivariate probit model is constructed to identify the factors affecting these two responses and also capture the correlation between them. Let $y_1$ be the distraction index with $y_1 = 1$ if the driver is distracted and $y_1 = 0$ otherwise, $y_2$ be the safety-critical index with $y_2 = 1$ if crash/near-crash occurs and $y_2 = 0$ otherwise. The bivariate probit model with a latent variable formulation takes the following form:

$$z_1 = \beta_1 X_1 + \epsilon_1, \quad y_1 = 1 \text{ if } z_1 \geq 0, \quad y_1 = 0 \text{ otherwise} \quad \text{Eq.1}$$

$$z_2 = \alpha z_1 + \beta_2 X_2 + \epsilon_2, \quad y_2 = 1 \text{ if } z_2 \geq 0, \quad y_2 = 0 \text{ otherwise}$$

Where, $z_1$ is the latent variable that indicates whether the driver is distracted ($y_1=1$ if $z_1 \geq 0$) or not ($y_1=0$ if $z_1 < 0$),

$X_1$ is the vector of explanatory variables for the first response,

$z_2$ is the latent variable that indicates whether the driver was involved in a safety critical event ($y_2=1$ if $z_2 \geq 0$) or not ($y_2=0$ if $z_2 < 0$),

$X_2$ is the vector of explanatory variables for the second response,
\[ \beta_1, \alpha, \beta_2 \] are the parameters to be estimated,

\[ \epsilon_1, \epsilon_2 \] are two random errors that follow a normal distribution with mean 0 and variance 1.

If the two responses are interrelated, the coefficient \( \alpha \) should be significantly different from 0. By implementing this model, the interrelationship between the distracted driving and the crash/near-crash likelihood could be investigated but from a different perspective.

4.3.1.1 Distracted Driving – Crash/Near-Crash Involvement

In this model, SAS® software was employed to investigate the association between distracted driving and the SCE involvement (crash/near-crash). The two response variables are modeled as a function of a set of independent variables, as described in Equation 1. In the first equation to model distraction, the set of independent variables included driver’s age, gender, marital status, working status, driver training, education, years of driving, relation to junction, and locality, as shown in Table 4.1. These independent variables were selected in accordance with previous studies, which examined the driver’s willingness to be engaged in a secondary tasks based on different personal and traffic flow factors (65-68). Table 4.3 displays the outcomes of the constructed bivariate probit model (only significant variables are shown). In the first equation, it was found that drivers between the ages of 16 and 34 years are more likely to be engaged in a secondary task while driving. The results also showed that the tendency of drivers of either full-time or part-time working status to be distracted while driving is higher than that of non-working drivers. This result is logical, as full-time and part-time drivers are more involved in the driving task. Moreover, Table 4.3 indicates that drivers are more likely to be engaged in a secondary task when they have passengers on board. The table also depicts that drivers prefer to be engaged in a secondary task while they are at intersections. This might suggest that there is a potential relationship between traffic density and secondary task engagement.
Table 4.3. Bivariate probit model results.

<table>
<thead>
<tr>
<th>Model</th>
<th>Coefficient</th>
<th>t-statistics</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First Model</strong>: Engaged in a secondary task (distracted/ not distracted)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AgeGroup 16-19</td>
<td>0.546</td>
<td>6.11</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>AgeGroup 20-24</td>
<td>0.653</td>
<td>7.39</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>AgeGroup 25-29</td>
<td>0.563</td>
<td>4.83</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>AgeGroup 30-34</td>
<td>0.317</td>
<td>2.22</td>
<td>0.0265</td>
</tr>
<tr>
<td>FullTime</td>
<td>0.239</td>
<td>3.54</td>
<td>0.0063</td>
</tr>
<tr>
<td>PartTime</td>
<td>0.129</td>
<td>1.96</td>
<td>0.0492</td>
</tr>
<tr>
<td>Intersection_Related</td>
<td>0.295</td>
<td>2.19</td>
<td>0.0287</td>
</tr>
<tr>
<td>Presence of passengers</td>
<td>0.419</td>
<td>6.17</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td><strong>Second Model</strong>: Involved in crash/near-crash event or not</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Engaged*</td>
<td>1.335</td>
<td>14.78</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>AgeGroup 16-19</td>
<td>1.622</td>
<td>9.258</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>AgeGroup 20-24</td>
<td>1.17</td>
<td>10.529</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>AgeGroup 25-29</td>
<td>0.293</td>
<td>2.49</td>
<td>0.009</td>
</tr>
<tr>
<td>Parking_Related</td>
<td>0.673</td>
<td>4.23</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Intersection</td>
<td>0.26</td>
<td>1.96</td>
<td>0.0492</td>
</tr>
<tr>
<td>Intersection Influence</td>
<td>0.629</td>
<td>6.35</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Road gradient (Grade up)</td>
<td>-0.318</td>
<td>-2.02</td>
<td>0.037</td>
</tr>
</tbody>
</table>

*engaged in secondary task

Unlike the first equation, the second equation to model crash/near-crash uses all the event and roadway characteristics shown in Table 4.1, in addition to driver age and gender as independent variables. The model showed that engagement in a secondary task is significantly correlated to the crash/near-crash likelihood. The positive coefficient implies that secondary task engagement increases the probability of crash/near-crash occurrence. Hence, there is strong statistical evidence of the impact of distracted driving on travel safety. The results also showed that parking and intersection locations are most prone to crash/near-crash occurrences. It is worth mentioning that “Intersection” variable in Table 4.1 refers to whether the event happened at an intersection location or not, while “Intersection Influence” refers to whether the intersection has an impact of the event outcome or not. To summarize, the bivariate probit model found that distracted driving, driver’s age, and intersection influence are the most significant predictors of
crash/near-crash likelihood. In the next section, further analysis will be conducted to quantify the increased crash/near-crash risk that results from different types of secondary tasks.

4.3.2 Secondary Tasks Risk Assessment

In this section, the increased crash/near-crash risk that results from the different secondary tasks is investigated. For this purpose, two different models were developed: a multinomial logit model, and an association analysis model. The two models attempted to quantify the increased crash/near-crash risk from two different perspectives. The multinomial logit model is a traditional statistical technique that is based on probability theory, whereas association analysis is a new powerful data-mining technique that reveals patterns in big data such as SHRP2 NDS data. The model results will be presented and the merits of each model will be discussed.

4.3.2.1 Multinomial Logit Model (Baseline-Category Logit Model)

The multinomial logit model is a statistical technique that is employed when the response variable has more than two categories. In this model, the response variable is the event severity (normal, near-crash, or crash event), whereas the explanatory variables are the secondary tasks listed in Figure 4.1. The baseline-category logit model pairs each response category with a reference response category. As the SHRP2 NDS dataset provides the distribution of secondary tasks in crash/near-crash as well as non-crash events, the increased crash/near-crash risk could be recognized and quantified. When the “normal” category is the baseline, the baseline-category logits are,

\[ \log\left(\frac{\pi_j}{\pi_f}\right), \quad \text{where } j = \text{Near-crash, Crash, and } J = \text{Normal} \quad \text{Eq.2} \]

where, \( \pi_j \) is the probability of the \( j \)th category. The baseline-category logits model with a set of predictors’ variables \( X \) (secondary tasks in our case) is defined as
\[
\log \left( \frac{\pi_j}{\pi_j} \right) = \beta_j X, \quad j = 1, \ldots, J - 1
\]

This model has \( J-1 \) equations with separate parameters for each. The effects vary with the category paired with the reference category. If \( J=2 \), this model simplifies to an ordinary logistic regression. The main advantage of the baseline-category logits model is the simultaneous fit of all the equations together. This advantage produces parameter estimates with smaller standard errors compared to fitting each equation separately using an ordinary logistic regression. In this regard, a PROC QLIM statement was recalled in SAS platform to achieve the modeling requirements.

According to previous studies, Odds Ratio (OR) is frequently used to estimate the relative risk of the secondary tasks while driving (4; 37; 38). ORs in a baseline-category logits model are defined as in a binary logistic model, except that they describe conditional odds. For instance, Figure 4.1 shows that the OR for a driver engaged in cell phone texting is 3.358 for near-crash. This means that the odds that drivers who are engaged in cell phone texting will be involved in a near-crash event rather than a normal driving event are about 3.35 times the odds for drivers who are not engaged in cell phone texting, adjusting for all the other secondary tasks. Similarly, the OR for manipulating objects is 2.262 for a crash event. Hence, we may say that the drivers who manipulate objects while driving have the odds of being involved in a crash event vs. normal event that is about 2.262 times the odds for those who are not engaged in manipulating an object, adjusting for all the other secondary tasks. Figure 4.1 displays the ORs of all secondary tasks in crash and near-crash events.
Figure 4.1. Odds ratios of different secondary tasks.

It should be noted that if the OR for a particular secondary task is less than 1.00 (dashed line), then the secondary task has no harmful effect on traffic safety. Accordingly, passenger interaction, eating/drinking, and dancing show a protective effect rather than a risk effect. However, Figure 4.1 indicates that the remaining secondary tasks are all within the risk range (OR > 1). For near-crash events, reading while driving showed the highest risk with an OR of 8.736, followed by cell phone dialing handheld, and manipulating objects with ORs of 4.598 and 3.863, respectively. Cell phone texting, cell phone other, and cell phone answering/locating/reaching follow, with ORs of 3.358, 3.64, and 3.00 respectively.

4.3.2.2 Association Analysis Model (A Priori Algorithm)

Data-mining techniques have been receiving increased attention from transportation researchers. These techniques have shown successful implementation in addressing safety problems compared with traditional statistical analyses (69-74). Detection of association rules is one of the powerful tools in data-mining techniques. It is considered the most frequent tool
employed in web mining within the retail industry (also known as market basket analysis). However, this method has a wide variety of useful applications such as in transportation safety. The goal of association analysis is to find rules in the form of conditions (antecedents) and results (consequents). The rules are developed based on the a priori algorithm. More details about the mechanism of the a priori algorithm can be found in Agrawal et al.’s study (75). Each developed rule is then evaluated using three performance measures: support, confidence, and lift. For instance, if an extracted rule states that “if (Var1 = x), → (Var2 = y), 30%, 80%”, it means that if variable 1 is equal to x, then the probability (Prob) that variable 2 will be equal to y is 80%, and the joint event (Var1 = x, Var2 = y) occurs in 30% of the observations. Accordingly, support, the first percentage in the rule, is defined as the probability of antecedent and consequent

\[
\text{Support (S)} = \text{Prob(antecedent and consequent)}
\]

Whereas, the confidence, the second percentage in the rule, is the conditional probability of consequent given antecedent and is denoted by;

\[
\text{Confidence (C)} = \text{Prob(consequent|antecedent)} = \frac{\text{Prob(antecedent and consequent)}}{\text{Prob(antecedent)}}
\]

Lift is a performance measure that was presented later by Brin et al.’s study (76). Lift displays the ratio of confidence for the rule to the marginal probability of having the consequent. To illustrate, suppose that 10% of the entire population buys a product X, then a rule that predicts whether people will buy product X with 20% confidence will have a lift of 20/10=2. If another rule tells you that people will buy X with 11% confidence, then the rule has a lift close to 1.00, meaning that having antecedent(s) makes little difference in the probability of having consequent. Therefore, lift is a measure of how helpful the rule is. Rules with lift index different from 1.00 are more interesting. Equation 5 shows the mathematical expression for the lift measure,
Lift (L) = \[ \frac{\text{Prob}(\text{consequent} \mid \text{antecedent})}{\text{Prob}(\text{consequent})} = \frac{\text{Prob}(\text{antecedent} \text{ and consequent})}{\text{Prob}(\text{antecedent}) \times \text{Prob}(\text{consequent})} \]

To conclude, support is a measure of frequency, confidence is the measure of belief, and lift is the measure of the improvement brought by the rule. In the marketing industry, sellers are more interested in finding rules with high support levels, high confidence indexes and lifts greater than 1.00. In transportation safety, crashes/near-crashes (Safety Critical Events SCE) have much lower frequencies than non-SCEs. As the main objective is to find the association between the SCEs and the associated secondary tasks, support of the rules could be quite low. Therefore, a lift performance measure is used in rules evaluation. In this chapter, we were more interested in finding rules connecting the secondary task activities (cell phone texting, eating, writing, manipulating objects, etc.) with the event severity. In essence, the study dataset was transformed into a tabular format, in which the columns represented indicator variables for the secondary task activities and the target variable was an SCE or non-SCE. Before interpreting the results, it is important to mention that the minimum support level specified for the proposed model was set at 0.1%. Regardless of the lift value, this means that no rule would have been extracted if it had a support level lower than 0.1%. This low value was selected because of the interest in extracting information related to rare events (crashes/near-crashes). Table 4.4 shows the rules extracted for secondary tasks and the SCE outcome. The rules are ranked based on the lift index. For better understanding the risk associated with different secondary tasks, rules should be compared with each other. For instance, Table 4.4 includes the following two rules:

“(Cellphone Texting= 1), → (Event= SCE), 2.56%, 27.64%”,

“(Vehicle embedded devices = 1), → (Event= SCE), 3.46%, 13.44%”
This means that the risk associated with cell phone texting is higher than that of operating vehicle-embedded devices. In other words, the probability of observing a SCE given that a driver is engaged in cell phone texting, (27.64%), is higher than that of vehicle-embedded devices (13.44%). Following the same criteria, all secondary tasks could be ranked based on how risky they are. The results indicated that reaching for objects, manipulating objects, reading, and other cell phone interaction activities are the highest risk secondary task activities. However, passenger interaction, eating/drinking, and dancing do not indicate a risk factor for SCE occurrence.

### 4.4 Discussion

This chapter provides quantitative insight into the risk associated with crash/near-crash events when drivers are engaged in secondary task activities. One of the most striking results in Figure 4.1 is the magnitude of the ORs (risk estimate). The figure shows that some activities can increase the crash or near-crash risk by four- to eightfold (such as reaching for objects, reading, and cell phone dialing handheld). These activities are considered high-risk distractors as they not only require multiple steps to be completed but also longer eyes-off-road time (such as, reaching for objects and reading). Surprisingly, other secondary tasks such as passenger interaction showed unexpected impacts. Cooper at al. found that passenger interaction increases the crash risk while studying teenagers drivers (77). However, this study found that the presence of a passenger has a protective effect rather than a risk effect. This could be explained as the presence of a passenger on board being equivalent to having more eyes on the road, which could reduce the crash or near-crash probability. This result is consistent with the findings in Geyer and Regland’s study (78).
Table 4.4. Association analysis model results.

<table>
<thead>
<tr>
<th>Consequent</th>
<th>Antecedent</th>
<th>Support%</th>
<th>Confidence%</th>
<th>Lift</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVENTSEVERITY = SCE</td>
<td>Reaching for object</td>
<td>1.41</td>
<td>40.79</td>
<td>3.35</td>
</tr>
<tr>
<td>EVENTSEVERITY = SCE</td>
<td>Manipulating objects</td>
<td>5.78</td>
<td>39.77</td>
<td>3.27</td>
</tr>
<tr>
<td>EVENTSEVERITY = SCE</td>
<td>Reading</td>
<td>0.11</td>
<td>39.13</td>
<td>3.21</td>
</tr>
<tr>
<td>EVENTSEVERITY = SCE</td>
<td>Cell phone other</td>
<td>0.31</td>
<td>38.81</td>
<td>3.19</td>
</tr>
<tr>
<td>EVENTSEVERITY = SCE</td>
<td>Cell phone Locating/reaching/answering</td>
<td>0.81</td>
<td>30.06</td>
<td>2.47</td>
</tr>
<tr>
<td>EVENTSEVERITY = SCE</td>
<td>Cell phone Dialing hand-held</td>
<td>0.19</td>
<td>30.00</td>
<td>2.46</td>
</tr>
<tr>
<td>EVENTSEVERITY = SCE</td>
<td>Cell phone Texting</td>
<td>2.56</td>
<td>27.64</td>
<td>2.27</td>
</tr>
<tr>
<td>EVENTSEVERITY = SCE</td>
<td>Pet in vehicle</td>
<td>0.19</td>
<td>26.83</td>
<td>2.20</td>
</tr>
<tr>
<td>EVENTSEVERITY = SCE</td>
<td>Cell phone Browsing</td>
<td>0.93</td>
<td>23.12</td>
<td>1.90</td>
</tr>
<tr>
<td>EVENTSEVERITY = SCE</td>
<td>Personal Hygiene</td>
<td>4.01</td>
<td>15.45</td>
<td>1.27</td>
</tr>
<tr>
<td>EVENTSEVERITY = SCE</td>
<td>Talking/singing</td>
<td>7.73</td>
<td>13.72</td>
<td>1.13</td>
</tr>
<tr>
<td>EVENTSEVERITY = SCE</td>
<td>Vehicle embedded devices</td>
<td>3.46</td>
<td>13.44</td>
<td>1.10</td>
</tr>
<tr>
<td>EVENTSEVERITY = SCE</td>
<td>External distractor</td>
<td>10.94</td>
<td>12.72</td>
<td>1.04</td>
</tr>
<tr>
<td>EVENTSEVERITY = SCE</td>
<td>Cell phone Talking/listening, hand-held</td>
<td>3.25</td>
<td>12.02</td>
<td>0.99</td>
</tr>
<tr>
<td>EVENTSEVERITY = SCE</td>
<td>Eating/Drinking/Smoking</td>
<td>4.28</td>
<td>10.34</td>
<td>0.85</td>
</tr>
<tr>
<td>EVENTSEVERITY = SCE</td>
<td>Passenger interaction</td>
<td>14.94</td>
<td>9.22</td>
<td>0.76</td>
</tr>
<tr>
<td>EVENTSEVERITY = SCE</td>
<td>Other known secondary task</td>
<td>3.72</td>
<td>9.00</td>
<td>0.74</td>
</tr>
<tr>
<td>EVENTSEVERITY = SCE</td>
<td>Dancing</td>
<td>1.11</td>
<td>7.11</td>
<td>0.58</td>
</tr>
</tbody>
</table>

As shown in Figure 4.1 and Table 4.4, some secondary tasks have a similar risk impact in relation to ORs or lift values. Thus, it is preferable to group these secondary tasks together. As a result, the k-means clustering algorithm was employed to group secondary tasks with similar risk effects together. k-means is a common unsupervised-learning clustering technique, which partitions n observations of unlabeled data into k clusters in which each observation belongs to the cluster with the nearest mean. Three clustering models were developed using SPSS Modeler with a predetermined number of clusters (k = 4). The clustering models used either the OR, obtained from the baseline-category logits model, or the lift index, obtained from the association model, as a clustering-based variable. The results of the clustering analysis are shown in Table 4.5. Although the highest impact secondary tasks are similar in both models, each model has its own advantages and disadvantages. To have considered more variables in baseline-category logits
models would have exposed the developed model to multicollinearity. This problem can increase the variance of the estimated parameters and hence, lead to higher uncertainties for the extracted point estimates. However, multicollinearity does not represent a problem for the association analysis model. In the association analysis model, no particular variable is defined as a response variable. Consequently, all rules that describe the association between the SCE/non-SCE event attributes can be extracted. In this analysis, only the one-product association rules were requested.

Table 4.5. Secondary tasks ranking.

<table>
<thead>
<tr>
<th>Multinomial Logits Model</th>
<th>A priori association model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Crash</strong> (k=4)</td>
<td><strong>Near-Crash</strong> (k=4)</td>
</tr>
<tr>
<td>Reaching for object</td>
<td>Reading</td>
</tr>
<tr>
<td>Cell phone, Dialing HH</td>
<td>Cell phone, Dialing HH</td>
</tr>
<tr>
<td>Reading</td>
<td>Manipulating object</td>
</tr>
<tr>
<td>Cell phone, Texting</td>
<td>Cell phone, Texting</td>
</tr>
<tr>
<td>Cell phone, others</td>
<td>Cell phone, others</td>
</tr>
<tr>
<td>Manipulating object</td>
<td>Reaching for object</td>
</tr>
<tr>
<td>Pet in vehicle</td>
<td>Cell phone, Ans/Reach</td>
</tr>
<tr>
<td>Vehicle embedded devices</td>
<td>Cell phone. Browsing</td>
</tr>
<tr>
<td>Cell phone, Browsing</td>
<td>Pet in vehicle</td>
</tr>
<tr>
<td>Personal Hygiene</td>
<td>Personal Hygiene</td>
</tr>
<tr>
<td>Cell phone, Ans/Reach</td>
<td>Talking/Singing</td>
</tr>
<tr>
<td>Outside distractor</td>
<td>Outside distractor</td>
</tr>
<tr>
<td>Talking/Singing</td>
<td>Vehicle embedded devices</td>
</tr>
</tbody>
</table>

- Color gradient indicates the k-clusters.

Rules were then filtered to present only the rules connecting the secondary task activities with the event severity. To the authors’ knowledge, this is the first study to employ and adjust the a priori algorithm settings in a distracted driving analysis.

4.5 Summary

This chapter analyzed the increased crash and near-crash risk associated with multiple secondary tasks using a variety of statistical and data-mining models. First, a bivariate model was constructed using the SHRP2 NDS data to examine the relationship between distracted driving and
SCE likelihood from different perspectives. The model indicated that distracted driving is a major contributor to an SCE occurrence. Subsequently, two different models were employed to quantify the increased risk associated with each secondary task: a baseline-category logits model, and a rule mining association model. The baseline-category logits model identified the increased risk in terms of ORs, while the a priori association algorithm detected the associated risks in terms of rules. Each rule was then evaluated based on the lift index. The two models succeeded in ranking all the secondary task activities according to the associated increased crash/near-crash risk efficiently. Both models revealed that reading while driving and reaching for objects are the highest crash risk among all secondary tasks. Furthermore, the k-means algorithm was implemented to cluster secondary tasks with similar risk impacts. Based on the results, a table was constructed to identify the k-means groups and the riskiest secondary tasks within each group. This chapter’s outcomes could help drivers understand the relative risks associated with the various secondary task activities so that they can adjust their behavior or consider alternatives. The outcomes could also help legislators initiate laws that reduce the crashes resulting specifically from distracted driving. Finally, it could help government officials make informed decisions about the allocation of available resources to reduce roadway crashes and improve traffic safety.
5 DETECTING DISTRACTED DRIVING VISUAL BEHAVIOR

Distracted driving behavior and driver inattention are two leading causes of roadway crashes. The state-of-the-art safety research has made several attempts to understand and quantify distracted driving and driver inattention. While each attempt had its limitation, there was a consensus on the relevance of eye glance behavior as a promising parameter in understanding distracted driving. Several studies have been performed to quantify various aspects of driver visual behavior. While the majority of the existing studies have focused on glance duration and glance frequency as the central parameters of interest, few studies incorporated the effect of eccentricity (how far the driver eye glance from the forward area) in their analysis. Combining eye-movement temporal and spatial metrics together may yield a better understanding of distracted driving behavior. In this context, this chapter aims to (1) investigate driver attention allocation patterns under real world SCEs and non-SCEs, (2) analyze distracted drivers visual behavior, and (3) construct robust distraction risk indicators.

5.1 Introduction and Background

Distracted driving is commonly defined as the diversion of attention away from crucial activities to maintaining driver safety toward other competing activities, leading to insufficient or no attention to the main driving task (4). Such diversion impairs driver’s visual, cognitive, physical, and auditory abilities and deteriorates the driving performance. There has been extensive research effort to investigate distracted driving behavior and its impacts on driving performance indicators. These indicators include reaction times (79; 80), variability of vehicle longitudinal and lateral position (79; 81; 82), and steering wheel reversal rate (82; 83), to name a few. Even though many of these indicators can help successfully detect distracted driving behavior to a reasonable
extent, other metrics (such as eye-movement metrics) could perform better in measuring and
detecting driver distraction.

Eye-movement metrics are the most promising diagnostic metrics for measuring driver
distraction (84-86). Glance duration is considered the most employed variable in quantifying
driver visual behavior. For example, Green studied the visual demand of a vehicle navigation
system in terms of glance per task and mean glance duration (87). Green recommended using
eyes-off-road time and on-road-task time as primary safety measures. In another study, Victor
examined the relationship between the visual behavior of drivers involved in different secondary
tasks and crash likelihood using an incomplete version of SHRP2 NDS database (85). Victor
found that the most sensitive glance metric is a linear combination of three parameters; Off3to1
(off-road glances from 3-sec until 1-sec prior to crash), mean off-road glance duration, mean
uncertainty (calculated based on driving uncertainty model). Although Victor introduced a new
strong glance behavior measure, his study was restricted to bumper-to-bumper crashes and to only
3 seconds before the crash/near-crash time. Another study by Fitch investigated the visual
distraction associated with cell phone use based on two features: total-eyes-off-road time and
percent of total-eyes-off-road time (TEORT) (88). These measures have been also used in other
studies such as (62; 89). Fitch concluded that cellphone tasks that require visual and physical
attention increase the TEORT significantly. Using a slightly different approach, Karlsson
introduced a time-based buffer index as a distracted driving measure (27). The core idea is that a
time window of 2 seconds runs backward when the driver begins to look away from the road. If
the 2-sec time window ends, the driver is regarded as being distracted. A similar approach was
developed in Fletcher and Zelinisky’s study; however, they used a counter (forward 2-sec time
window) instead of a timer as in Karlsson’s study (90). These 2-sec time windows measures
performed relatively well when they were applied on distracted driving data (91). However, the
pre-identified time window threshold (2 seconds) remains a major concern. Several other studies examined driver’s glance time and frequency to measure driver visual distraction (92; 93).

Although glance duration feature has shown a strong correlation with distracted driving behavior, several studies focused solely on glance location as a distracted driving identifier (52; 94-96). It has been demonstrated that eccentric glances relative to the speedometer level and onward impair driving as well as event detection (97). In a different study, Klauer presented some descriptive analysis for the crash risk associated with drivers gaze locations (37). In Klauer’s study, the eye glance locations were divided into four different zones; based on the visual angle from center road forward. Even though, the chi-square analysis showed statistical significant differences in the event type at these zones, Klauer’s did not use these zones to construct a distraction indicator measure. Later, Liang used the 100-car study to test impact of the three principal characteristics of eye patterns — duration, history, and location — on distraction detection using simple linear mathematical relationships (98). These relationships were able to detect imminent SCEs with $R^2$ values between 0.13 and 0.88. Liang’s study concluded that more complex models, and naturalistic driving studies with longer observational periods, could better explain this relationship.

In summary, several studies have been performed to quantify various aspects of driver visual behavior. While the majority of the existing studies have focused on glance duration and glance frequency as the central parameters of interest, few studies incorporated the effect of eccentricity in their analysis. Combining eye-movement temporal and spatial metrics together may yield a better understanding of distracted driving behavior. Therefore, this chapter extends the literature in a number of ways. First, this chapter will study the relationship between distracted driving and driver visual attention patterns in real world driving environment using the Naturalistic Engagement in Secondary Tasks (NEST) dataset. According to these patterns and the three
principle characteristics of eye patterns (i.e., duration, frequency, and eccentricity), two new distraction indicator measures will be developed: number of renewal cycles per event ($N_{RC}$), and Distraction Index ($DI$) measures. This distraction index will then be statistically analyzed to test its ability to distinguish SCEs from non-SCEs.

5.2 Data Description

In this chapter, the NEST dataset is used to achieve the chapter objectives. As mentioned earlier in chapter 2, NEST dataset is developed as a subset from the SHRP2 NDS data that focus primarily on distracted driving. The NEST dataset provides detailed information about 400 distraction-related SCEs along with a balancing sample of 800 non-SCEs. Each SCE is combined with a balancing sample of 3 to 4 non-SCE. For each driving event (SCE or non-SCE), two types of data are collected: time-series and event summary.

The time series data includes 16 different variables describing vehicle performance, glance behavior, and secondary task engagement while driving. The time series data are collected for longer observation periods compared to SHRP2 NDS events (20-sec for non-SCEs and 30-sec for SCEs compared to 6-sec for non-SCEs and 30-sec for SCEs in SHRP2 NDS). Every SCE is divided into two parts, 20 seconds and 10 seconds, split by a time marker at a point representing the onset of a precipitating factor. A precipitating factor is defined as the triggering factor of a SCE (such as; running a red light) (58). The glance behavior variable included in NEST dataset indicates the locations where the driver gaze is directed. This variable was coded for each 0.1-sec by a VTTI data reductionist who reviewed the video data on a frame by frame basis. The review included the following set of locations: “cell phone”, “center stack”, “instrument cluster”, “interior object”, “passenger”, “left forward”, “left mirror”, “left window”, “rearview mirror”, “right forward”, “right mirror”, “right window”, “eye closed”, “no video”, and “forward”. Finally, the
secondary task engagement variable records the type of task or activity underway, such as manipulating objects, reading, adjusting embedded devices...etc. On the other hand, the event summary data describe the vehicle and environmental conditions during an event, such as traffic density, flow, intersection influence, and road surface condition among many other variables.

5.3 Methods

This paper implements a renewal cycle concept to understand the driver visual behavior under different circumstances. The renewal cycle approach is inspired by visual attention and cognition studies performed in psychology research (99). A renewal cycle is defined as the driver’s eye shifting process from a reference focal point to another focal point(s) before returning back to the reference focal point. Since the forward area is the most attractive area at which drivers look more comfortably and naturally (as supported by the high frequency of glancing to the forward area in Figure 5.1), it is treated as a fixed reference point during the generation of the renewal cycles. Accordingly, a renewal cycle in this study starts once a driver’s eye glances towards the forward area, followed by visiting other focal point(s), and ending by returning back to the forward area. While recording the different focal points during a renewal cycle, the glance durations at each focal point are also recorded. This approach will not only study distraction by analyzing driver off-road glances but also consider the driver on-road glances over time. In other words, the renewal cycle approach will describe the driver visual behavior as a complete chain process, consisting of on and off-road glances, instead of only studying the off road glances. Extracting such chain processes could help detect driver visual attention patterns associated with certain secondary tasks. Thus, with the high resolution detailed secondary task variable provided in the NEST database, the application of such approach is expected to provide deeper insights into how drivers allocate their attention while driving and performing certain types of tasks.
To extract the renewal cycles’ information from the NEST dataset, four main steps are followed. The data are first processed and combined using R® studio. Then, the time-series eye glance data are allocated to four points, namely; A, B, C, and D. Each point is defined based on the radial gaze angle from the forward roadway, as shown in Figure 5.2 and Table 5.1. This classification is adopted by Klauer’s study during analyzing the eye glance behavior in 100-Car study (37). Based on the eye glance (attention) allocation to the four points, the renewal cycles are generated such that each cycle starts and ends at point A. This is repeated until all eye glance data associated with all events are converted into renewal cycles.

![Figure 5.1. Glance location distribution over time for NEST safety critical events (colored).](image)

Since the main focus of this chapter is to study driver visual behavior in distraction-related driving events, the eye glance behavior within each renewal cycle is grouped into two categories: drivers paying attention to the road (looking at the forward area “Point A”), and looking away from the center road forward (Points B, C, and D). Considering these two categories, descriptive statistics are obtained to explain the glance behavior associated with the different renewal cycles.
in the data. Finally, the renewal cycles are then further analyzed to develop a distraction level index (DI) to quantify drivers’ visual distraction.

Table 5.1. Eye glance classification.

<table>
<thead>
<tr>
<th>Points/Regions</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average visual angle</td>
<td>0°</td>
<td>&lt; 20°</td>
<td>20° - 40°</td>
<td>&gt; 40°</td>
</tr>
<tr>
<td>Glance locations</td>
<td>Forward</td>
<td>Left forward, right forward, instrument panel</td>
<td>Center mirror, radio/HVAC, left mirror</td>
<td>Left window, right window, right mirror, passengers, hand-held device, object/other, and eyes closed</td>
</tr>
</tbody>
</table>

Figure 5.2. Eye glance locations.

5.3.1 Data Processing

The driving events data in NEST dataset contain different behaviors. Therefore, in order to study the driver attention allocation process during distracted driving events, all types of behavior other than distracted driving should be excluded from the analysis. To do so, several filters are applied on the data to exclude events with alcohol or drug impairment, events taking place in parking lots, and events with missing eye glance data.
According to Engström’s study, driver state changes from proactive to reactive once a precipitating factor takes place (100). Since the present study focuses on analyzing the driver normal behavior prior to SCEs, the events time-series data are truncated to exclude any information coded after the occurrence of precipitating factors.

5.4 Results and Discussion

5.4.1 Descriptive Analysis of Renewal Cycles

A total number of 3497 renewal cycles is extracted from all driving events (SCE and non-SCEs) in the NEST data. As shown in Table 5.2, very few renewal cycles (5 in SCEs and 6 in non-SCEs) have drivers eye glance moving across all focal points, while the majority of the renewal cycles (955 in SCEs and 2291 in non-SCEs) have only two focal points. These numbers indicate that drivers prefer to frequently pay attention back to the forward area so that they can update their information about the traffic ahead to maintain situational awareness. Since the percentages of three-glance and four-glance renewal cycles are very low, these cycles are treated as extreme cases for the rest of our analysis.

When looking at the glance durations, Table 5.2 shows that drivers tend to spend 2-3s looking forward and 1s looking elsewhere when they are performing two-focal-point cycles. As the number of focal points increases in the renewal cycles, drivers tend to spend less time looking forward and more time looking elsewhere. This confirms that drivers get distracted with different levels measured by the number of focal points within the renewal cycles and the associated glance duration at each focal point. It is hypothesized that the number of renewal cycles within each event is also an important measure for the level of distraction and hence the level of risk associated with glance behavior. This is supported by looking at the average per-event number of renewal cycles.
for SCEs ($\bar{N}_{RC} = 6$) compared to that associated with non-SCEs ($\bar{N}_{RC} = 3$) as shown in Table 5.2. The $N_{RC}$ is further analyzed through statistical analysis in the following section.

Table 5.2. Renewal cycle results.

<table>
<thead>
<tr>
<th></th>
<th>Number of glances</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>Total</td>
</tr>
<tr>
<td>(a) SCE ($\bar{N}_{RC} = 6$)</td>
<td>955(89.4%)</td>
<td>108(10.1%)</td>
<td>5 (0.47%)</td>
<td>1068</td>
</tr>
<tr>
<td>Frequency</td>
<td>2.37</td>
<td>2.25</td>
<td>1.68</td>
<td>2.36</td>
</tr>
<tr>
<td>Duration of forward glance(s)</td>
<td>3.13</td>
<td>2.8</td>
<td>1.57</td>
<td>3.1</td>
</tr>
<tr>
<td>Mean*</td>
<td>20.7</td>
<td>19.0</td>
<td>3.5</td>
<td>20.7</td>
</tr>
<tr>
<td>Standard deviation*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration of non-forward glance(s)</td>
<td>1.10</td>
<td>2.38</td>
<td>2.96</td>
<td>1.54</td>
</tr>
<tr>
<td>Mean*</td>
<td>0.97</td>
<td>2.21</td>
<td>1.91</td>
<td>1.23</td>
</tr>
<tr>
<td>Standard deviation*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum*</td>
<td>10.4</td>
<td>13.5</td>
<td>6.2</td>
<td>13.5</td>
</tr>
<tr>
<td>Mean duration of renewal cycle</td>
<td>3.47</td>
<td>4.67</td>
<td>4.64</td>
<td>3.6</td>
</tr>
<tr>
<td>(b) Non-SCE ($\bar{N}_{RC} = 3$)</td>
<td>2291 (94.3%)</td>
<td>132 (5.43%)</td>
<td>6 (0.25%)</td>
<td>2429</td>
</tr>
<tr>
<td>Frequency</td>
<td>2.47</td>
<td>2.349</td>
<td>1.9</td>
<td>2.461</td>
</tr>
<tr>
<td>Duration of forward glance(s)</td>
<td>2.746</td>
<td>2.623</td>
<td>1.93</td>
<td>2.738</td>
</tr>
<tr>
<td>Mean*</td>
<td>15.2</td>
<td>14.1</td>
<td>4.5</td>
<td>15.2</td>
</tr>
<tr>
<td>Standard deviation*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration of non-forward glance(s)</td>
<td>0.883</td>
<td>1.665</td>
<td>2.017</td>
<td>0.933</td>
</tr>
<tr>
<td>Mean*</td>
<td>0.591</td>
<td>1.003</td>
<td>1.537</td>
<td>0.668</td>
</tr>
<tr>
<td>Standard deviation*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum*</td>
<td>8.6</td>
<td>8.2</td>
<td>5.6</td>
<td>8.6</td>
</tr>
<tr>
<td>Mean duration of renewal cycle</td>
<td>3.321</td>
<td>4.014</td>
<td>4.117</td>
<td>3.39</td>
</tr>
</tbody>
</table>

*All measurements are in seconds

5.4.1.1 Mixed-Effects Model ($N_{RC}$)

The $N_{RC}$ is then calculated for each driving event in the filtered dataset. To evaluate the performance of the $N_{RC}$ measure to distinguish between the two different event types (SCE/non-SCE), a mixed-effects model is used to test whether the mean $N_{RC}$ value associated with SCEs is significantly different from that associated with non-SCEs, while accounting for heterogeneity in the driver population. In the mixed-effects model, the event effect we are interested in is treated as a fixed effect, while the driver effect is treated as a random effect. Hence, the estimated fixed effect for event controls for the variability between individuals (drivers). SAS PROC MIXED is
used to fit the mixed-effects model where the event type is coded as the fixed effect and drivers as a random effect. It was found that there is a strong statistical evidence ($F$-value$_{1, 694} = 146.93$, $p$-value < 0.0001) that the mean $N_{RC}$ value associated with SCEs is significantly higher than that associated with non-SCEs (see Table 5.3). This indicates that $N_{RC}$ could be a promising indicator for characterization of SCEs.

Table 5.3. $N_{RC}$ mixed model ANOVA results.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Num DF</th>
<th>Den DF</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crash</td>
<td>1</td>
<td>694</td>
<td>146.93</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

5.4.2 Renewal Cycles and Secondary Tasks

This section provides a detailed analysis on the extracted renewal cycles to understand drivers’ glance behavior and level of distraction associated with the different types of secondary tasks. Since not enough data are found for the three- or four-focal-point renewal cycles, they are assumed to be extreme cases and only the two-focal-point cycles are analyzed. Table 5.4 illustrates the percentages of the two-glance renewal cycles for the different focal points (B, C, or D). The percentages are calculated for the different event types (SCE and non-SCE) while engaging in different types of secondary tasks. The table shows that manipulating objects not integral to the vehicle driving task such as cell phones, mp3 players, or others has the highest frequency of renewal cycles with glancing to point D (which represents the highest level of distraction). This indicates that manipulating objects is a demanding secondary task that could lead to a significant visual distraction. Looking at the distribution of renewal cycles, talking/listening on a handheld cellphone does not lead to a high level of visual distraction (1.11% and 1.76% renewal cycles with glancing to point D). This implies that distraction associated with this type of secondary tasks is
mostly cognitive or physical rather than being visual. The distribution of the total number of renewal cycles across the different focal points reveals interesting facts: (a) SCEs are dominated by renewal cycles with D focal points (456), which indicates that drivers are more often significantly visually distracted before getting involved in a crash/near-crash event; and (b) non-SCEs have comparable numbers of renewal cycles for the different focal points, which indicates that drivers are more keen to keep their situational awareness (since glancing to B and C could be associated with updating information about the driving environment).

Table 5.4. Renewal cycle distribution across different secondary tasks.

<table>
<thead>
<tr>
<th>Secondary Tasks</th>
<th>SCEs</th>
<th>Non-SCEs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&quot;B&quot;</td>
<td>&quot;C&quot;</td>
</tr>
<tr>
<td>Driver's Eye Location</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manipulating objects</td>
<td>2.15</td>
<td>4.15</td>
</tr>
<tr>
<td>Talking/Singing to passenger</td>
<td>24.03</td>
<td>20.00</td>
</tr>
<tr>
<td>Holding objects</td>
<td>9.01</td>
<td>8.68</td>
</tr>
<tr>
<td>Personal Hygiene</td>
<td>3.00</td>
<td>3.40</td>
</tr>
<tr>
<td>Talking/Listening on handheld cell phone</td>
<td>10.73</td>
<td>5.28</td>
</tr>
<tr>
<td>External distractor</td>
<td>17.17</td>
<td>23.40</td>
</tr>
<tr>
<td>Reaching for objects</td>
<td>1.72</td>
<td>1.89</td>
</tr>
<tr>
<td>Talking/Singing to self</td>
<td>13.30</td>
<td>8.30</td>
</tr>
<tr>
<td>Others</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Searching for objects (internal objects)</td>
<td>8.15</td>
<td>5.66</td>
</tr>
<tr>
<td>Eating/Drinking</td>
<td>5.58</td>
<td>1.51</td>
</tr>
<tr>
<td>Adjusting/Monitoring embedded devices</td>
<td>0.86</td>
<td>2.26</td>
</tr>
<tr>
<td>Adjusting/monitoring center stack controls</td>
<td>3.86</td>
<td>15.47</td>
</tr>
<tr>
<td>Dancing</td>
<td>0.43</td>
<td></td>
</tr>
<tr>
<td>No secondary task</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (3426 RC)</td>
<td>233</td>
<td>266</td>
</tr>
</tbody>
</table>

5.4.3 Distraction Level Index (DI)

Visual distraction is measured by the frequency of non-forward glances which are defined as looking outside the Field of Relevant Driving (FRD). In this study, off-road glances are defined by focal points B, C, and D. Durations of off-road glances also give an indication about the level of that visual distraction. The longer the drivers look away, the more distracted they become. A
large body of the previous research attempted to estimate levels of distraction as a function of the non-forward glance duration and frequency regardless of the specific glance locations (37; 49). This means that all glances away from the FRD have been assumed to have the same level of distraction. This assumption ignores the fact that some non-forward glances could be meant to gather information about the driving environment (e.g. points B and C in this study). To overcome this limitation, other studies defined the FRD such that they include points B and C (101). This approach however ignores the possibility of drivers being distracted even when looking at those focal points. Therefore, the current study proposes a new distraction level index that uses a rule to assign weights to driver glance behavior. These weights are estimated based on the exact locations drivers are looking at.

To calculate the weight associated with each non-forward focal point, the eccentricity function $E(\alpha)$ designed by Lamble is employed (97). This function penalizes non-forward glances depending on the gaze angle $\alpha$ of each focal point. This is calculated as:

$$E(\alpha) = 6.5758 - \frac{1}{(0.06 \times \alpha + 0.152)}$$

Gaze angle values between $0^\circ$~$20^\circ$, $20^\circ$~$40^\circ$, and $>40^\circ$ are assigned to non-forward glance locations “B”, “C”, and “D”, respectively. These values are used based on Klauer’s study that grouped eye-glance locations together based on the visual angle measured from the center forward (37). In this study, the eccentricity function $E(\alpha)$ is calculated for every one-degree increment within each gaze angle range. Then, the average $E(\alpha)$ values for the different gaze-angle ranges are used as glance behavior weights for the different focal points. This weight can be referred to as eccentricity penalty factor ($\varepsilon$). Based on that, the penalty values obtained for focal points B, C, and D, are 0.2, 1.12, and 2.58, respectively. These values imply that higher gaze angles could lead to higher levels of distraction. This is why glancing to point D is penalized the highest compared to glancing to
points B and C. Whereas, glancing to point B is accompanied with a minor level of distraction, hence is assigned the lowest penalty value. It is clear that the penalty assigned to in-vehicle glances is twice to ten times higher than the penalties assigned to glancing to points B and C, which is reasonable since the latter points might not be associated with any distraction, as mentioned earlier. Using the eccentricity penalties, a robust function is developed to measure $DI$. This function accounts for three main factors including eye glance history, duration, and eccentricity, as shown below.

$$DI = \sum_{i=1}^{N} \frac{O_i}{C_i} * \varepsilon_i$$

From the renewal cycle perspective, eye glance history is measured by the number of renewal cycles $N$ generated for an entire event (SCE or non-SCE). Glance duration is measured as the amount of time a driver spends looking at a non-forward focal point ($O_i$) relative to the length ($C_i$) of renewal cycle $i$. Finally, the eye glance eccentricity associated with each renewal cycle is measured by the eccentricity factor $\varepsilon_i$. In the following section, this function is applied and investigated statistically.

5.4.3.1 Mixed-Effects Model ($DI$)

The $DI$ function is calculated for each driving event in the filtered NEST data, and the same statistical analysis performed in section 5.2.1.1 is then applied but to the $DI$ measure. It was found that there is a strong statistical evidence ($F-value_{1,70}=225.14$, $p-value < 0.0001$) that the mean $DI$ value associated with SCEs is significantly higher than that associated with non-SCEs (as shown in Table 5.5). This implies that $DI$ can be used as an indicator for the risk level associated with drivers glance behavior.
Table 5.5. DI Mixed-model ANOVA results.

<table>
<thead>
<tr>
<th>Effect</th>
<th>Num DF</th>
<th>Den DF</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crash</td>
<td>1</td>
<td>701</td>
<td>225.14</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

According to the mixed-effects models, the two developed distraction risk indicators show a distinguishable property in classifying SCEs and non-SCEs. To test the predictability power of these two indicators, it is recommended to construct a new crash/near-crash prediction model using these indicators in addition to other vehicle and environmental factors. Moreover, these indicators establish a foundation to design a new in-vehicle driver warning system that is capable of alerting distracted drivers. Combining these two indicators with other vehicle performance features would help in better detecting distracted driving and avoiding potential crashes. The following chapter is an attempt to predict distraction-related SCEs using these two indicators in addition to other roadway and vehicle parameters.

5.5 Summary

In this chapter, the relationship between distracted driving and driver inattention was studied using the NEST dataset. A renewal cycle approach was developed to analyze the driver visual behavior under different circumstances. This approach was then adopted to develop a new parameter \( N_{RC} \), and to investigate the relationship between the eye glance behavior and different secondary tasks. The mixed-effects model showed that the \( \bar{N}_{RC} \) measures differ significantly between the two event types (SCE/non-SCE), which indicates that \( N_{RC} \) could be a promising indicator for characterization of SCEs. In addition, a new distraction level index \( DI \) was developed based on the renewal cycle components. The \( DI \) measure was then analyzed based on the event type to gain some insights about its performance. The results demonstrate that SCEs are usually
associated with higher $DI$ values compared non-SCEs. These findings confirm that higher values of $DI$ and $N_{RC}$ measures could have striking implications in predicting SCE. Further analysis will be conducted in chapter 6 to determine the predictability power of these two measures in detecting potential crash/near-crash risks. The findings in this chapter are promising to the quantification of the risk associated with distraction related visual behavior. The developed distraction measure can help quantify levels of visual distraction associated with different types of secondary tasks, and hence, guide policy makers in issuing appropriate laws and regulations.
6 CRASH/NEAR-CRASH PREDICTION MODEL

According to a recent report by AAA Foundation, distracted driving remains a top safety concern with nearly 88% of the drivers believing that distracted driving is on the rise. The report indicates that distracted driving presents even higher risk than aggressive driving and driving under influence. Clearly, advanced technology such as smart phones and vehicle integrated systems plays a major role in increasing the number of driving distractors in recent years. Nonetheless, there is also a wide belief that advanced technology including new distraction countermeasure systems could help solve the distracted driving phenomenon. This chapter aims to identify risk factors and predict distraction-related Safety Critical Events (SCEs) using various vehicle, roadways, and driver characteristics in addition to two new distraction risk indicators. In this context, an Artificial Neural Network (ANN) model is developed to predict the distraction-related SCEs using the Naturalistic Engagement in Secondary Tasks (NEST) dataset. The following sections will present the chapter methodology and analysis in more details.

6.1 Methodology

In this chapter, NEST dataset will be employed to predict SCEs using an ANN model. First, the dataset is processed for the ANN model development. Then, the ANN model is evaluated according to the performance measures presented later in this section. Figure 6.1 shows the detailed research framework followed in this chapter.
6.1.1 Artificial Neural Networks (ANN)

6.1.1.1 Model structure

ANN models attempt to imitate how human brain operates. These models consist of basic units called neurons where these neurons are arranged into layers as shown in Figure 6.2. A typical neural network model involves three main layers: an input layer, where input features are inserted; one or more hidden layer; and an output layer, where outcomes are expected. The neurons within each layer are connected to the neurons in the next layer with varying connection strengths called weights. These weights are then used to propagate the information from the input layer neurons to the hidden layer(s) neurons before a result is eventually obtained at the output layer.
First, the neural network determines the weights randomly which usually lead to incorrect responses. The neural network then learns by training. In the training process the network generates a prediction for each record and then compares the predicted value with the observed value. This process is repeated many times and the results of these comparisons are then used to modify the weights — this type of networks is called Feed Forward Backward Propagation (FFBP) model (102). Once a stopping criterion is met (e.g., mean square error is minimized), the neural network stops the training process and starts the validation and testing processes with the rest of the data.
6.1.1.2 Performance measures

In binary classification problems (0/1), three performance measures are usually used to evaluate the developed model: sensitivity, specificity, and Area Under the Curve (AUC). Sensitivity is defined as the true positive rate which is the ratio between the number of predicted “1”s divided by the observed number of “1”s, whereas specificity is defined as true negative rate which is the ratio between the number of predicted and observed “0”s. The model sensitivity and specificity are defined for a particular cut-off value that determines whether the output will be “0” or “1”. The AUC observes the trend of the model sensitivity versus 1-specificity under different cut-off points. Higher AUC values indicate that the model has good performance regardless of the cut-off point values. In this study these three performance measures are used to assess the ANN results.

In this chapter, NEST vehicle, roadway, and driver variables displayed in Table 6.1 are used to construct the ANN model. In the proposed ANN model, the output layer is defined as a binary outcome; showing whether a SCE is about to happen (“1”) or not (“0”). ANN models are widely popular models for detecting patterns, especially in transportation-related applications (102; 103). Not only can these models find the best non-linear functions to fit the data, but they can also help avoid the multicollinearity problem of traditional statistical techniques such as logistic regression. The following section shows how the data are prepared for ANN model construction.
Table 6.1. List of input variables.

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Variable type</th>
<th>#of Categories</th>
<th>Description/Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Event Summary</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seatbelt use</td>
<td>Categorical</td>
<td>4</td>
<td>Lap/shoulder belt, Lap only, Shoulder only, None used</td>
</tr>
<tr>
<td>Number of passengers</td>
<td>Categorical</td>
<td>4</td>
<td>0, 1, 2, 3+</td>
</tr>
<tr>
<td>Weather</td>
<td>Categorical</td>
<td>6</td>
<td>Fog, Mist or Light Rain, Heavy Rain, Snowing, other, No adverse condition</td>
</tr>
<tr>
<td>Driver behavior1,2,3</td>
<td>Categorical</td>
<td>19</td>
<td>“Driving behaviors made by the driver during the event. Behaviors may be apparent at times other than the time of the precipitating factor, such as aggressive driving at an earlier moment which led to retaliatory behavior later. Subsequent inappropriate or illegal behaviors are labeled DriverBehavior2 and DriverBehavior3.”</td>
</tr>
<tr>
<td>Driver impairments</td>
<td>Categorical</td>
<td>7</td>
<td>Drowsy, Drugs, other illicit drugs, Impaired due to previous injury…etc</td>
</tr>
<tr>
<td>Secondary task**</td>
<td>Categorical</td>
<td>11</td>
<td>“Observable driver engagement in any of the listed secondary tasks during the 10s of the event (if available): Cell phone interaction, Adjust/Monitor embedded device, Passenger interaction Reaching for object …etc”</td>
</tr>
<tr>
<td>Road surface condition**</td>
<td>Categorical</td>
<td>4</td>
<td>Dry, Icy, Snowy, Wet</td>
</tr>
<tr>
<td>Traffic flow**</td>
<td>Categorical</td>
<td>4</td>
<td>Divided, Undivided, One-way traffic, No lanes</td>
</tr>
<tr>
<td># of Travel lanes**</td>
<td>Categorical</td>
<td>9</td>
<td>0,1,2,3,…, 8+</td>
</tr>
<tr>
<td>Traffic density**</td>
<td>Categorical</td>
<td>6</td>
<td>LOS A, B, C, D, E,F</td>
</tr>
<tr>
<td>No traffic control**</td>
<td>Binary</td>
<td>0/1</td>
<td>indicates whether the participant was not subject to or influenced by any traffic controls during 10s of the event</td>
</tr>
<tr>
<td>Stop Sign**</td>
<td>Binary</td>
<td>0/1</td>
<td>indicates whether the participant was subject to or influenced by at least one stop sign or not</td>
</tr>
<tr>
<td>Traffic signal**</td>
<td>Binary</td>
<td>0/1</td>
<td>indicates whether the participant was subject to or influenced by at least one traffic signal or not</td>
</tr>
<tr>
<td>Merge sign**</td>
<td>Binary</td>
<td>0/1</td>
<td>indicates whether the participant was subject to or influenced by at least one merge sign or not</td>
</tr>
<tr>
<td>Road alignment**</td>
<td>Categorical</td>
<td>3</td>
<td>Curve left, Curve right, Straight</td>
</tr>
<tr>
<td>Road grade**</td>
<td>Categorical</td>
<td>4</td>
<td>Grade up, Grade down, Hillcrest, Level</td>
</tr>
<tr>
<td>Locality**</td>
<td>Categorical</td>
<td>12</td>
<td>Business industrial, Church, Construction zone, Urban, School, Interstate…etc</td>
</tr>
<tr>
<td>Lighting**</td>
<td>Categorical</td>
<td>5</td>
<td>Daylight, Dawn, Darkness lighted, Darkness not lighted, Dusk</td>
</tr>
<tr>
<td><strong>Driver Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>Categorical</td>
<td>16</td>
<td>16-19, 20-24, 25-29, 30-34,…, 85-89</td>
</tr>
<tr>
<td>Sex</td>
<td>Binary</td>
<td>0/1</td>
<td>Male/Female</td>
</tr>
<tr>
<td>Annual miles</td>
<td>Categorical</td>
<td>7</td>
<td>&lt;5k, 5k-10k, 10k-15k, …,&gt;30k</td>
</tr>
<tr>
<td>#of violations</td>
<td>Categorical</td>
<td>3</td>
<td>0,1,2+</td>
</tr>
<tr>
<td>VMIScore</td>
<td>Categorical</td>
<td>3</td>
<td>Impairment level found from Visualizing Missing Information test results; none, mild, serious</td>
</tr>
<tr>
<td>UFOVScore</td>
<td>Categorical</td>
<td>3</td>
<td>Impairment level found from UFOV test results; none, mild, serious</td>
</tr>
<tr>
<td>Clock Drawing Score</td>
<td>Categorical</td>
<td>6</td>
<td>Perfect, Minor visuospatial errors, Inaccurate time, good visuospatial, Inaccurate time, minor visuospatial errors, Moderate visuospatial errors, Severe visuospatial errors, No reasonable representation of a clock</td>
</tr>
<tr>
<td>( N_{RC} )</td>
<td>Continuous</td>
<td></td>
<td>Number of renewal cycles per event</td>
</tr>
<tr>
<td>DI</td>
<td>Continuous</td>
<td></td>
<td>Distraction level index</td>
</tr>
</tbody>
</table>
6.1.1.1 Data processing

According to the data description section, NEST dataset contains different variables that happened before and after the precipitating event. Since driver behavior after the precipitating event changes from proactive to reactive status (i.e. driving behavior is categorically different from normal behavior), all the information provided beyond this time point is excluded (100). In other words, this proposed ANN model includes only the variables that take place for the first two 10-sec intervals. For example, if traffic density is collected three times —traffic density1, traffic density2, and traffic density3, then only traffic density1 and traffic density2 are included in the analysis. Moreover, only driving events are considered in the analysis; events that took place in parking lots are excluded. Finally, driving events associated with drug or alcohol impairments are removed from the rest of the analysis. The final dataset that was used in this chapter contained 683 non-SCE and 285 SCEs.

6.2 Results and Discussion

As mentioned earlier in the methodology section, the FFBP is employed to develop the ANN model. Different model structures were tested to select the number of hidden layers, and the number of neurons within these hidden layers that lead to the best model performance. Using SPSS modeler, the best model performance was achieved with one hidden layer and four neurons; this is in addition to the 50 neurons to represent the input variables and one neuron for the output layer. In this ANN model, the output layer was defined as a binary outcome, showing whether a SCE is about to happen (“1”) or not (“0”). Seventy percent of the data are used to train the ANN model, whereas the rest of data are divided equally for the validation and the testing processes. Table 6.2 shows the results of the ANN model in terms of confusion matrices.
A confusion matrix is a matrix that shows the percentage of correctly and incorrectly predicted events (102). According to Table 6.2, the ANN succeeded in predicting SCEs with an average AUC equals to 0.91. Table 6.2 also shows that the model sensitivity — the ratio between the number of predicted SCEs divided by the observed number of SCEs — ranges between 72.7 and 84.17%, whereas the model specificity — the ratio between the number of predicted non-SCEs divided by the observed number of non-SCEs — ranges from 86.6 to 98.25%. Additional performance measures are also provided for the developed ANN classifier in Table 6.3.

<table>
<thead>
<tr>
<th>Observed</th>
<th>Predicted</th>
<th>Area Under the Curve (AUC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;0&quot;</td>
<td>98.25%</td>
<td>1.75%</td>
</tr>
<tr>
<td>&quot;1&quot;</td>
<td>15.83%</td>
<td>84.17%</td>
</tr>
<tr>
<td>Observed</td>
<td>&quot;0&quot;</td>
<td>86.60%</td>
</tr>
<tr>
<td>&quot;1&quot;</td>
<td>27.30%</td>
<td>72.70%</td>
</tr>
<tr>
<td>Observed</td>
<td>&quot;0&quot;</td>
<td>93.70%</td>
</tr>
<tr>
<td>&quot;1&quot;</td>
<td>19.00%</td>
<td>81.00%</td>
</tr>
</tbody>
</table>

Table 6.3. ANN performance measures

<table>
<thead>
<tr>
<th>Measure</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall (%)</td>
<td>79.28</td>
</tr>
<tr>
<td>False Positive Rate (%)</td>
<td>18.05</td>
</tr>
<tr>
<td>Precision (%)</td>
<td>81.94</td>
</tr>
</tbody>
</table>

Since identifying the factors that affected the model the most is the secondary goal in this chapter, the predictor importance chart was used to achieve this goal. The predictor importance chart indicates the relative importance of each variable on a scale from 0 to 1, and the sum of the values for all variables is equal to 1. It should be noted that the predictor importance relates the importance of each variable in making a prediction, not whether the prediction is accurate or not. When looking at the input variables that contributed the most to the predictability power of the
ANN model in Figure 6.3, it is clear that traffic density preceding a SCE ranks first among all input variables. This is followed by $N_{RC}$, driver behavior (defined in Table 3.1), $DI$, and traffic flow.

![Variable Importance Chart](image)

Figure 6.3. ANN results - variable importance chart.

Compared to traditional statistical models (98), this analysis shows that an artificial intelligent model can make better prediction of distraction-related SCEs with an overall accuracy equals to 96.1% — compared to 83% in Liang’s study (98). The developed model also identifies the two driver behavior variables, the $N_{RC}$, and $DI$, as two of the top four variables that impacted the ANN model predictability power. This means that inclusion of driver behavior variables can improve SCE prediction. With the availability of advanced data collection equipment, such as: remote eye tracking sensors, it is possible to track driver behavior and warn inattentive drivers when potential risks arise.

Reducing the number of crashes/near-crashes that are caused by distracted driving is a traffic safety challenge that must be tackled. In response to this challenge, several crash prevention models have been developed (104-106); however few are designed for the distraction-related crash/near-crash events. Generally, traffic density is usually reported as a major contributing factor in crash prevention models (107). The higher the traffic volume on the road, the higher the
likelihood of crashes. In line with these studies, this study showed that traffic density ranks first among other risk indicators in predicting distraction-related SCEs. This result is also consistent with the recent State Farm survey study that identified red light stops and slow moving traffic as the most suitable conditions for secondary task involvement (35).

The findings in this chapter do not only update the current knowledge regarding distracted driving and traffic safety relationship, but also helps institutions such as policy makers and automobile manufacturers address the driver distraction problem. Approaches such as the one presented in this chapter can serve as a foundation for an advanced driver warning systems that can alert drivers if potential crash/near-crash risks increase.

6.3 Summary

In this chapter, the NEST dataset was employed to achieve two main objectives: (1) predict distraction-related SCEs; (2) identify the factors that contributed the most in the prediction process. At first, the NEST data was processed in a thorough manner to exclude any confounding factors that might impact the results. Second, an Artificial Neural Network (ANN) model was developed to predict SCEs based on vehicle, environment, and driver behavior variables. The ANN model was then trained, validated and tested using 70%, 15%, and 15% of the data, respectively. The results show that the ANN was able to predict distraction-related SCEs with an overall accuracy equal to 96.1%. Moreover, the ANN model extracted the importance variable chart that displays the importance of each variable in the prediction process. It was found that variables related to driving behavior were amongst the most important predictive variables. This finding demonstrates that the inclusion of driver behavior variables in addition to other vehicle and roadway variables can improve crash/near-crash prediction and reduce the false alarms. Findings in this chapter can help automobile manufacturers design driver warning assistance systems. The presented ANN
model can serve as a foundation for an advanced driver warning system that can alert drivers of the potential increase in crash risk.
7 SUMMARY AND CONCLUSIONS

The ultimate goal in this dissertation research was to develop tool(s) that help minimize distraction-related crashes and near-crashes events (i.e. safety critical events). This goal was achieved through three distinct research phases. In phase one, in-depth analysis was conducted to estimate the increased crash and near-crash risk associated with different distraction activities. In phase two, real-time visual driver distraction detection algorithm was developed, and new distraction risk indicators were constructed. These new distraction risk indicators in addition to other vehicle, driver and environmental variables were then used in phase three to predict distraction-related crash and near-crash events. The following sections will summarize and conclude the findings in each phase separately.

7.1 Phase One: Crash/Near-Crash Risk Assessment

This research phase provided an in-depth analysis for the increased crash/near-crash risk associated with different secondary tasks, tasks that are not related to the driving task, using the largest and the most comprehensive real-world naturalistic driving database (SHRP2 Naturalistic Driving Study). Several statistical and data mining techniques were developed to analyze the distracted driving and crash risk relationship. First, a bivariate probit model was constructed to investigate the relationship between the engagement in a secondary task and Safety-Critical Events (SCEs) likelihood. Subsequently, two different techniques were implemented to quantify the increased crash/near-crash risk due to involvement in a particular secondary task. The first technique used the baseline-category logits model to estimate the increased crash risk in terms of conditional odds ratios. The second technique used the Apriori association rule mining algorithm...
to reveal the risk associated with each secondary task in terms of support, confidence and lift indexes.

In the bivariate probit model, SAS software was employed to achieve two goals (a) investigate the association between distracted driving (i.e. engagement in a secondary task) and the SCE involvement, and (b) identify the factors that affect the engagement in a secondary task and the SCE likelihood. The bivariate probit model revealed that driver’s age, employment status, and intersections locations are the most significant predictors for driver’s willingness to be engaged in a secondary task. It was also found that driver’s engagement in a secondary task, driver’s age, and intersection influence are the most significant predictors of crash/near-crash likelihood.

Given the fact that distracted driving and crash/near-crash likelihood are significantly correlated, two subsequent models were developed to quantify the increased crash/near-crash risk associated with different secondary tasks: baseline-category logits, and association analysis model. In the baseline-category logits model, the relative crash/near-crash risk of different secondary tasks was computed in terms of odds ratios. The odds ratio was found to be directly proportional to the crash/near-crash risk associated with each secondary task activity. Using SHRP2 NDS database, it was found that reaching for objects, reading, and cell phone dialing handheld activities could increase the crash and near-crash risk by four- to eight-fold. ORs for the other secondary tasks are also computed and displayed in Figure 4.1.

On the other hand, the association analysis model (Apriori algorithm) quantified the increased crash/near-crash risk in form of association rules — e.g. (Cellphone Texting= 1), → (Event= SCE), 2.56%, 27.64%, 2.27”. Each developed rule was then evaluated using three performance measures: support, confidence, and lift; where support is a measure of frequency,
confidence is the measure of belief, and lift is the measure of the improvement brought by the rule (how useful the rule is?). The rules were then ranked based on the lift index in order to rank the different secondary tasks based on the risk estimates. The results indicated that reaching for objects, manipulating objects, reading, and other cell phone interaction activities are the riskiest secondary task activities. All rules connecting secondary task activities and crash/near-crash likelihood were also extracted and displayed in Table 4.4. Far to the author knowledge, this is the first study to employ and adjust the A-priori algorithm settings in distracted driving analysis.

Since some secondary tasks have similar crash/near-crash risk impact in terms of ORs or lift values, K-means clustering algorithm was employed to group secondary tasks with similar risk effects together. Table 4.5 displayed the risk categories and placed each secondary task activity in its appropriate risk category.

It is essential to understand the impact of distracted driving in the larger context of naturalistic driving to provide useful suggestions for countermeasures. The outcomes of this phase can be adopted and implemented at different sectors (automobile industry, decision makers, safety campaigns, etc.) to address distracted driving behavior. For instance, the automobile industry needs to reduce the in-vehicle features that require visual and physical interaction. This, in turn, will increase driver focus and decrease the eyes-off road time. One of the possible recommendations is to lock out all the complex in-vehicle features while the vehicle is in motion. Moreover, as cell phone interaction activities are also one of the main internal distraction sources, it would be preferable to develop a new cell phone mode that prohibits all complex features while the vehicle is in motion (similar to airplane mode). Additionally, drivers should be aware of all the relative risks that are associated with the various secondary task activities so that they can adjust their behavior or consider alternatives. Safety campaigns that convey the message “all
distractions are bad’’ are unrealistic and ineffective. Identifying the most serious secondary tasks can help safety campaigns to achieve their goals effectively. Finally, policymakers and legislative institutions should devise their acts (texting bans, handheld cell phone bans, etc.) based on NDS information, not on unrealistic experiments.

7.2 Phase Two: Distraction Detection System

This research phase aimed to (a) develop a real-time gaze-based algorithm for measuring driver visual distraction, and (b) construct new distraction risk indicators. First, a novel approach was introduced to adequately represent driver attention allocation patterns in safety and non-safety critical events. The proposed approach applied a renewal cycle concept that is inspired by psychological research. A renewal cycle was defined as the driver’s eye shifting process from a center forward area to another focal point(s) before returning back to the center forward area. During this process, the time spent at each focal point, the process’s sequence, and the entire renewal cycle time were all recorded. Second, the renewal cycle approach was implemented to investigate driver attention allocation patterns under different secondary task activities using the Naturalistic Engagement in Secondary Tasks (NEST) data set. NEST dataset is a subset from SHRP2 naturalistic driving study that focused primarily on distracted driving. With the high resolution of the detailed secondary task variable provided in the NEST database, the application of such approach (renewal cycle) is expected to provide deeper insights into how drivers allocate their attention while driving and performing certain types of tasks.

To obtain meaningful results using the renewal cycle approach, the eye glance variable in NEST database was reduced from 15 locations to 4 locations (A, B, C, and D) based on the radial gaze angle from center forward area, where A represents the center forward area. This classification was adopted from the 100-Car study as reported in Liang’s study (98).
A total number of 3497 renewal cycles were extracted from all driving events (SCE and non-SCEs) in the NEST database. The majority of the extracted renewal cycles (955 in SCEs and 2291 in non-SCEs) had only two focal points. These numbers indicated that drivers prefer to frequently pay attention back to the forward area so that they can update their information about the traffic ahead to maintain situational awareness. It was also found that the average per-event number of renewal cycles for SCEs is \(N_{RC}=6\) compared to those associated with non SCEs \(N_{RC}=3\). Therefore, the number of renewal cycles within each event was hypothesized to be an important measure for the level of distraction and hence the level of risk associated with glance behavior. A mixed-effects ANOVA model was then constructed in SAS platform to test whether the \(N_{RC}\) values associated with SCEs are significantly different from those associated with non-SCEs, while accounting for heterogeneity in the driver population. It was found that there is a strong statistical evidence \((F-value_{1, 694}= 146.93, p-value < 0.0001)\) that the \(N_{RC}\) values associated with SCEs is significantly higher than that associated with non-SCEs. This indicated that \(N_{RC}\) could be a promising indicator for characterization of SCEs.

Based on the renewal approach, a new distraction level index \((DI)\) was developed considering the eye glance history, duration, and eccentricity. The newly developed \(DI\) was function of: the total number of renewal cycles per event, the off-road glance duration within each renewal cycle \(O_i\), the renewal cycle length \(C_i\), and an eccentricity penalty factor that penalizes non-forward glances depending on the gaze angle \(\alpha\) of each focal point \(\varepsilon_i\). This function was calculated for each driving event in NEST dataset, for which another mixed-effects ANOVA model was constructed to test whether \(DI\) values for SCEs are different from those computed for non-SCEs or not. It was found that there is a strong statistical evidence \((F-value_{1,701}= 225.14, p-value < 0.0001)\) that the \(DI\) values associated with SCEs is significantly higher than that associated
with non-SCEs. This implied that $DI$ can also be used as an indicator for the risk level associated with distracted drivers’ glance behavior.

The two developed distraction indicator measures ($N_{RC}$ & $DI$) showed a distinguishable property in classifying SCEs and non-SCEs. These findings confirmed that higher values of $DI$ and $N_{RC}$ measures could have striking implications in predicting distraction-related SCEs. The findings in this phase are promising to the quantification of the risk associated with distraction-related visual behavior. The developed distraction measure can help quantify levels of visual distraction associated with different types of secondary tasks, and hence, guide policy makers in issuing appropriate laws and regulations for car manufacturing industry.

7.3 Phase Three: Crash/Near-Crash Prediction Model

The last phase in this dissertation research aimed to develop a driver assistance warning system that can alert distracted drivers if potential crash or near-crash is about to occur. In this regard, an artificial intelligence model was developed to predict distraction-related SCEs using NEST dataset. Different vehicle, environment and driver characteristics in addition to the two new distraction risk indicators ($N_{RC}$ & $DI$) that were previously developed in phase two were further used to construct an Artificial Neural Network (ANN), and also to identify the factors that contributed the most in the prediction process. R studio and SPSS platforms were then used to train, validate and test the developed model using the data provided in NEST dataset. The results showed that the ANN succeeded in predicting distraction related-SCEs with an overall accuracy equals to 96.1%.

The model also identified the variables that affect the distraction-related SCEs the most and presented them in a variable importance chart. When looking at the variable importance chart, it was found that traffic density preceding a SCE ranks first among all input variables. This was
followed by the number of renewal cycles in each driving event ($N_{RC}$), driver behavior (defined in Table 3.1), Distraction Index ($DI$), and traffic flow. This implies that inclusion of driver behavior variables can improve crash/near-crash prediction.

The finding in this phase demonstrated that the inclusion of driver behavior variables in addition to other vehicle and roadway variables can improve crash/near-crash prediction accuracy. Findings in this phase can also help automobile manufacturers design driver warning assistance systems. The presented ANN model can serve as a foundation for an advanced driver warning system that can alert drivers of the potential increase in crash risk.

### 7.4 Closing Remarks and Recommendations

As long as technology advances, distracted driving will remain a major concern in transportation safety. The advanced wireless communication devices and vehicle increased dashboard instrumentation have brought distraction to another level. However, it is believed that advanced technology can also tackle the distracted driving problem by developing new Advanced Driver Assistance Systems (ADAS) that could help in warning distracted drivers if potential crash risk arises. In this dissertation research, a gaze-based real-time distraction index was developed and validated using real world driving data. Even though, this distraction index measure has proved its power to predict distraction-related crash and near-crash events, it is recommended to incorporate additional parameters that could strengthen the distraction index predictability robustness. Moreover, the availability of vehicle kinematics variables (such as: speed, acceleration, lateral acceleration…etc.) in SHRP2 NDS database, would make it possible to improve the distracted driving detection algorithm by including other vehicle lateral and longitudinal control variables.
Therefore, a follow up study will investigate the performance of some vehicle kinematics variables with the developed distraction index ($DI$). If any observable variation with any vehicle performance variable (such as speed or acceleration) is recognized, a new term will be added to the distraction index equation to increase the $DI$ distraction detection robustness. Even though it might be impossible to replace eye movement related indicators completely with driving related parameters, it would definitely valuable to be able to fall back on this type of data when eye tracking is lost.
REFERENCES


59. Institute, V. T. T. [https://insight.shrp2nds.us/](https://insight.shrp2nds.us/).


87. Green, P. Visual and task demands of driver information systems. 1999.


95. Reimer, B., B. Mehler, I. Reagan, D. Kidd, and J. Dobres. Multi-modal demands of a smartphone used to place calls and enter addresses during highway driving relative to two embedded systems. 2015.


APPENDIX A IRB APPROVAL FORM

ACTION ON PROTOCOL CONTINUATION REQUEST

TO: Sherif Ishak
    Civil Engineering

FROM: Dennis Landin
      Chair, Institutional Review Board

DATE: October 1, 2015

RE: IRB# 3371

TITLE: Distracted Driving and Associated Crash Risks

New Protocol/Modification/Continuation: Continuation

Review type: Full ___ Expedited _X_ Review date: 10/1/2015

Risk Factor: Minimal ____ X _____ Uncertain _________ Greater Than Minimal ________

Approved ___ X ___ Disapproved ___________

Approval Date: 10/1/2015 Approval Expiration Date: 9/30/2016

Re-review frequency: (annual unless otherwise stated)

Number of subjects approved: 4000

LSU Proposal Number (if applicable): 42688

Protocol Matches Scope of Work in Grant proposal: (if applicable) ___

By: Dennis Landin, Chairman

PRINCIPAL INVESTIGATOR: PLEASE READ THE FOLLOWING –
Continuing approval is CONDITIONAL on:

1. Adherence to the approved protocol, familiarity with, and adherence to the ethical standards of the Belmont Report, and LSU's Assurance of Compliance with DHHS regulations for the protection of human subjects*
2. Prior approval of a change in protocol, including revision of the consent documents or an increase in the number of subjects over that approved.
3. Obtaining renewed approval (or submittal of a termination report), prior to the approval expiration date, upon request by the IRB office (irrespective of when the project actually begins); notification of project termination.
4. Retention of documentation of informed consent and study records for at least 3 years after the study ends.
5. Continuing attention to the physical and psychological well-being and informed consent of the individual participants, including notification of any information that might affect consent.
6. A prompt report to the IRB of any adverse event affecting a participant potentially arising from the study.
8. SPECIAL NOTE:

*All investigators and support staff have access to copies of the Belmont Report, LSU's Assurance with DHHS, DHHS (45 CFR 46) and FDA regulations governing use of human subjects, and other relevant documents in print in this office or on our World Wide Web site at http://www.lsu.edu/irb
ACTION ON PROTOCOL CONTINUATION REQUEST

TO: Sherif Ishak
Civil Engineering

FROM: Dennis Landin
Chair, Institutional Review Board

DATE: October 1, 2015

RE: IRB# 3371

TITLE: Distracted Driving and Associated Crash Risks

New Protocol/Modification/Continuation: Continuation

Review type: Full ___ Expedited X ___ Review date: 10/1/2015

Risk Factor: Minimal ___ X ___ Uncertain _________ Greater Than Minimal_______

Approved ___ X ___ Disapproved _______

Approval Date: 10/1/2015 Approval Expiration Date: 09/30/2016

Re-review frequency: (annual unless otherwise stated)

Number of subjects approved: 4000

LSU Proposal Number (if applicable): 42688

Protocol Matches Scope of Work in Grant proposal: (if applicable) ___

By: Dennis Landin, Chairman

PRINCIPAL INVESTIGATOR: PLEASE READ THE FOLLOWING –
Continuing approval is CONDITIONAL on:

1. Adherence to the approved protocol, familiarity with, and adherence to the ethical standards of the Belmont Report, and LSU’s Assurance of Compliance with DHHS regulations for the protection of human subjects*
2. Prior approval of a change in protocol, including revision of the consent documents or an increase in the number of subjects over that approved.
3. Obtaining renewed approval (or submittal of a termination report), prior to the approval expiration date, upon request by the IRB office (irrespective of when the project actually begins); notification of project termination.
4. Retention of documentation of informed consent and study records for at least 3 years after the study ends.
5. Continuing attention to the physical and psychological well-being and informed consent of the individual participants, including notification of new information that might affect consent.
6. A prompt report to the IRB of any adverse event affecting a participant potentially arising from the study.
8. SPECIAL NOTE:

*All investigators and support staff have access to copies of the Belmont Report, LSU’s Assurance with DHHS, DHHS (45 CFR 46) and FDA regulations governing use of human subjects, and other relevant documents in print in this office or on our World Wide Web site at http://www.lsu.edu/irb
APPENDIX B  DATA ACQUISITION SYSTEM (DAS)

a. DAS Head unit assembly  
b. DAS Main system unit  
c. DAS Forward radar assembly  
d. Video cameras views  

Figure B.1. Data acquisition system equipment.
Table B.1. Variables collected with DAS.

<table>
<thead>
<tr>
<th>Location</th>
<th>Sensor</th>
<th>Data to Be Collected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head unit</td>
<td>Multiple cameras/video</td>
<td>Video images of the forward view, center stack view, rear and passenger side view,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>and driver face view, information for machine vision (MV) processes (including lane</td>
</tr>
<tr>
<td></td>
<td></td>
<td>tracking and eyes forward data), as well as periodic, irrevocably blurred still</td>
</tr>
<tr>
<td></td>
<td></td>
<td>photographs of the cabin interior to capture passenger presence</td>
</tr>
<tr>
<td>Main unit</td>
<td>Accelerometer data</td>
<td>In 3 axes:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Forward/reverse (x)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Right/left (y)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Down/up (z)</td>
</tr>
<tr>
<td>Main unit</td>
<td>Rate sensors</td>
<td>Yaw rate</td>
</tr>
<tr>
<td>Head unit</td>
<td>GPS</td>
<td>Latitude, longitude, elevation, time, velocity</td>
</tr>
<tr>
<td>Radar unit</td>
<td>Forward radar</td>
<td>Object ID, range, and range rate</td>
</tr>
<tr>
<td>Main unit</td>
<td>Cell phone module</td>
<td>Health checks, location notification, collision notification, and remote software</td>
</tr>
<tr>
<td></td>
<td></td>
<td>upgrades</td>
</tr>
<tr>
<td>Head unit</td>
<td>Illuminance sensor</td>
<td>Ambient lighting levels</td>
</tr>
<tr>
<td>Head unit</td>
<td>Passive alcohol sensor</td>
<td>Presence of alcohol within the vehicle cabin</td>
</tr>
<tr>
<td>Head unit</td>
<td>Incident push button</td>
<td>In the event of an unusual or interesting traffic safety-related event, allows</td>
</tr>
<tr>
<td></td>
<td></td>
<td>participant to open an audio recording channel for 30 seconds; also “flags” the</td>
</tr>
<tr>
<td></td>
<td></td>
<td>data stream for ease of location during data analysis</td>
</tr>
<tr>
<td>Head unit</td>
<td>Audio</td>
<td>Available only in concert with the incident push button as noted above</td>
</tr>
<tr>
<td></td>
<td>Turn signals (from vehicle network data or</td>
<td>Turn signal actuation, which distinguishes between left and right indicated turns</td>
</tr>
<tr>
<td></td>
<td>directly from the signals themselves)</td>
<td></td>
</tr>
<tr>
<td>Main unit</td>
<td>Vehicle network data</td>
<td>Where available, the use of the accelerometer, brakes, ABS, gear position, steering</td>
</tr>
<tr>
<td></td>
<td></td>
<td>wheel angle, speed, horn, seat belt information, airbag deployment, and many other</td>
</tr>
<tr>
<td></td>
<td></td>
<td>such variables</td>
</tr>
</tbody>
</table>
Figure B.2. Eye movement calibration.
There are 21 questions in this survey

**Demographic Data**

1. **What is your gender:**
   
   Please choose *only one* of the following:
   
   - Female
   - Male

2. **What is your date of birth?**
   
   Please enter a date:

3. **What do you consider your ethnicity to be?**
   
   Please choose *only one* of the following:
   
   - Hispanic or Latino
   - Not Hispanic or Latino

4. **What do you consider your race to be?**
   
   Please choose *all that apply*:
   
   - Black or African American
   - White
   - Asian
   - American Indian or Alaska Native
   - Native Hawaiian or Other Pacific Islander
   - Other:

5. **What is your country of birth?**
   
   Please choose *only one* of the following:
6 What is the highest level of education you have completed?

Please choose only one of the following:

- Some high school
- High school diploma or G.E.D.
- Some education beyond high school but no degree
- College degree
- Some graduate or professional school, but no advanced degree (e.g., J.D.S., M.S. or Ph.D.)
- Advanced degree (e.g., J.D.S., M.S. or Ph.D.)

7 What is your marital status?

Please choose only one of the following:

- single
- married
- divorced
- widow(er)
- unmarried partners

8 What is your current household status?

Please choose only one of the following:

- Two-parent household
- One-parent household
- Other

9 Is your primary household:

Please choose only one of the following:

- Owned
- Rented
10 What is your current work status?

Please choose only one of the following:

- Not working outside the home
- Part-time
- Full-time

11 What is your current job title or profession?

Please write your answer here:

12 What is your family’s annual household income (from all sources and before taxes)?

Please choose only one of the following:

- Under $29,000
- $30,000 to $39,999
- $40,000 to $49,999
- $50,000 to $59,999
- $60,000 to $69,999
- $70,000 to $99,999
- $100,000 to $149,999
- $150,000 +

13 How many people including yourself live in your household?

Please write your answer here:

14 Of these, how many are:

Please write your answer(s) here:

Less than 16 years old?

16 - 20?
<table>
<thead>
<tr>
<th>Age Group</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>21 - 25?</td>
<td></td>
</tr>
<tr>
<td>26 - 30?</td>
<td></td>
</tr>
<tr>
<td>31 - 35?</td>
<td></td>
</tr>
<tr>
<td>36 - 40?</td>
<td></td>
</tr>
<tr>
<td>41 - 45?</td>
<td></td>
</tr>
<tr>
<td>46 - 50?</td>
<td></td>
</tr>
<tr>
<td>51 - 55?</td>
<td></td>
</tr>
<tr>
<td>56 - 60?</td>
<td></td>
</tr>
<tr>
<td>61 - 65?</td>
<td></td>
</tr>
<tr>
<td>65 +</td>
<td></td>
</tr>
</tbody>
</table>
15 How many vehicles are there in your household?

Please write your answer here:

16 What is your ZIP Code?

Please write your answer here:

17 How long have you lived in your neighborhood?

Please choose only one of the following:

⊙ Less than 1 year
⊙ 1 to 5 years
⊙ more than 5 years

18 Approximately how many miles did you drive last year?

Please write your answer here:

19 Do you use the study vehicle for any business purposes (such as pizza delivery or realtor activities)?

Please choose only one of the following:

⊙ Yes
⊙ No

20 If ‘Yes’, please indicate the business purpose.

Only answer this question if the following conditions are met:

° Answer was ‘Yes’ at question ‘19’ (Do you use the study vehicle for any business purposes (such as pizza delivery or realtor activities)?)

Please write your answer here:
21 At what age did you receive your driver's license (i.e., when did you drive legally by yourself)?

Please choose only one of the following:

- < 15
- 15
- 15.5
- 16
- 16.5
- 17
- 17.5
- 18
- > 18

Submit your survey.
Thank you for completing this survey.
APPENDIX D  DRIVING HISTORY QUESTIONNAIRE

DRIVING HISTORY

A note on privacy
This survey is anonymous. The record kept of your survey responses does not contain any identifying information about you unless a specific question in the survey has asked for this. If you have responded to a survey that used an identifying token to allow you to access the survey, you can rest assured that the identifying token is not kept with your responses. It is managed in a separate database, and will only be updated to indicate that you have (or haven’t) completed this survey. There is no way of matching identification tokens with survey responses in this survey.

There are 16 questions in this survey

DH 1

1 Over the past five years, what is your average annual mileage (estimate)?

Please choose only one of the following:

- less than 5,000 miles
- 5,000 - 10,000 miles
- 10,000 - 15,000 miles
- 15,000 - 20,000 miles
- 20,000 - 25,000 miles
- 25,000 - 30,000 miles
- more than 30,000 miles

2 How many years have you been driving?

Please write your answer here:

3 Describe the kind(s) of driver training that you have received prior to getting a license as well as throughout your driving experience.

Please choose all that apply:
Informal driver training offered by a parent, family member or friend
☐ Driver’s Education offered through school
☐ Driver’s Education offered through private company
☐ Operated farm equipment or lawn tractors
☐ Motor sports experience
☐ Post-licensure driver skills training/enhancement
☐ Other:

4 In the past year, how many moving or traffic violations have you had?

Please choose only one of the following:

☐ 0
☐ 1
☐ 2
☐ 3
☐ 4
☐ 5
☐ more than 5

5 What type of violations were involved?

Only answer this question if the following conditions are met:
☐ Answer was greater than or equal to ‘more than 5’ at question 4 (4a) (In the past year, how many moving or traffic violations have you had?)

Please write your answer here:
6 In the past year, how many crashes have you been in?

Please choose only one of the following:

- 0
- 1
- 2
- 3
- 4
- 5
- more than 5

7 Accident 1 Severity (select highest or most severe option applicable)

Only answer this question if the following conditions are met:

* Answer was greater than or equal to 'more than 5' at question 6 [5]’ (In the past year, how many crashes have you been in?)

Please choose only one of the following:

- Injury
- Tow-away (any vehicle)
- Police-reported
- Damage (any), but no police report

8 Were you at fault?

Only answer this question if the following conditions are met:

* Answer was greater than or equal to 'more than 5' at question 6 [5]’ (In the past year, how many crashes have you been in?)

Please choose only one of the following:

- Yes
- No
9 Accident 2 Severity (select highest or most severe option applicable)

Only answer this question if the following conditions are met:
° Answer was greater than or equal to 'more than 5' at question '6 [5]' (In the past year, how many crashes have you been in?)

Please choose only one of the following:

- Injury
- Tow-away (any vehicle)
- Police-reported
- Damage (any), but no police report

10 Were you at fault?

Only answer this question if the following conditions are met:
° Answer was greater than or equal to 'more than 5' at question '6 [5]' (In the past year, how many crashes have you been in?)

Please choose only one of the following:

- Yes
- No

11 Accident 3 Severity (select highest or most severe option applicable)

Only answer this question if the following conditions are met:
° Answer was greater than or equal to 'more than 5' at question '6 [5]' (In the past year, how many crashes have you been in?)

Please choose only one of the following:

- Injury
- Tow-away (any vehicle)
- Police-reported
- Damage (any), but no police report

12 Were you at fault?
Only answer this question if the following conditions are met:

Answer was greater than or equal to 'more than 5' at question '6' [5] (In the past year, how many crashes have you been in?)

Please choose only one of the following:

- Yes
- No

13 Accident 4 Severity (select highest or most severe option applicable)

Only answer this question if the following conditions are met:

Answer was greater than or equal to 'more than 5' at question '6' [5] (In the past year, how many crashes have you been in?)

Please choose only one of the following:

- Injury
- Tow-away (any vehicle)
- Police-reported
- Damage (any), but no police report

14 Were you at fault?

Only answer this question if the following conditions are met:

Answer was greater than or equal to 'more than 5' at question '6' [5] (In the past year, how many crashes have you been in?)

Please choose only one of the following:

- Yes
- No

15 Accident 5 Severity (select highest or most severe option applicable)

Only answer this question if the following conditions are met:

Answer was greater than or equal to 'more than 5' at question '6' [5] (In the past year, how many crashes have you been in?)

Please choose only one of the following:
16 Were you at fault?

Only answer this question if the following conditions are met:
* Answer was greater than or equal to 'more than 5' at question '6 [5]' (In the past year, how many crashes have you been in?)

Please choose **only one** of the following:

- Yes
- No

Submit your survey.
Thank you for completing this survey.
APPENDIX E  PERMISSION TO USE PREVIOUSLY PUBLISHED MATERIAL

Title:  Crash and Near-Crash Risk Assessment of Distracted Driving and Engagement in Secondary Tasks: A Naturalistic Driving Study
Author:  Peter R. Bakhit, BeiBei Guo, Sherif Ishak
Publication:  Transportation Research Record
Publisher:  SAGE Publications
Date:  06/15/2018

If you are a SAGE journal author requesting permission to reuse material from your journal article, please note you may be able to reuse your content without requiring permission from SAGE. Please review SAGE's author re-use and archiving policies at https://us.sagepub.com/en-us/nam/journal-author-archiving-policies-and-re-use for more information.

If your request does not fall within SAGE's reuse guidelines, please proceed with submitting your request by selecting one of the other reuse categories that describes your use. Please note, a fee may be charged for reuse of content requiring permission. Please contact permissions@sagepub.com if you have questions.

BACK  CLOSE WINDOW
VITA

Peter Ramzy Zaki Bakhit was born in Cairo, Egypt, in 1990. Peter received his B.S. degree in civil engineering from Cairo University, Egypt, in 2012. In 2015, he received his M.S. in transportation engineering from Cairo University. Currently, Peter is pursuing his Ph.D. degree in transportation engineering at Louisiana State University.

Due to his high undergraduate grades, Peter was appointed as a Teaching and Research Assistant at Cairo University. He was also offered a full scholarship to purse his master’s degree at Cairo University. In 2015, Peter joined the department of Civil and Environmental Engineering at Louisiana State University as a Research Assistant. His research interests include human factors in transportation engineering, traffic safety, data mining, traffic operation and management, and intelligent transportation systems.

Peter is a student member in each of the Institute of Transportation Engineers and the American Society of Civil Engineering. Peter is also an awardee of the “Dissertation Year Fellowship” that was awarded by the Graduate School of Louisiana State University. His nomination was selected from a highly gifted, highly competitive pool of applicants because of his project’s scholarly potential.