9-1-2021

FACTORS AFFECTING DRIVERS’ OFF-ROAD GLANCE BEHAVIOR WHILE INTERACTING WITH IN-VEHICLE VOICE INTERFACES – INSIGHTS FROM A SECONDARY DATA ANALYSIS

Fangda Zhang
University of Massachusetts Amherst

FACTORS AFFECTING DRIVERS' OFF-ROAD GLANCE BEHAVIOR WHILE INTERACTING WITH IN-VEHICLE VOICE INTERFACES – INSIGHTS FROM A SECONDARY DATA ANALYSIS

A Dissertation Presented

By

FANGDA ZHANG

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

SEPTEMBER 2021

Industrial Engineering and Operations Research

FACTORS AFFECTING DRIVERS' OFF-ROAD GLANCE BEHAVIOR WHILE INTERACTING WITH IN-VEHICLE VOICE INTERFACES – INSIGHTS FROM A SECONDARY DATA ANALYSIS

A Dissertation Presented

By

FANGDA ZHANG

Approved as to style and content by:

____

Shannon C. Roberts, Chair

____

Krista J. Gile, Member

____

Anuj K. Pradhan, Member

____

Sundar Krishnamurty, Department Head
Mechanical & Industrial Engineering

DEDICATION

To my beloved family.

ABSTRACT

FACTORS AFFECTING DRIVERS' OFF-ROAD GLANCE BEHAVIOR WHILE INTERACTING WITH IN-VEHICLE VOICE INTERFACES – INSIGHTS FROM A SECONDARY DATA ANALYSIS

SEPTEMBER 2021

FANGDA ZHANG, B.ENG., SHANGHAI UNIVERSITY

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Shannon C. Roberts

Given the prevalence of in-vehicle technologies and the critical role of visual attention plays in driving safety, this dissertation work aimed to fill the research gap that 1) little was known about the visual demands associated with a driver engaging with in-vehicle voice interfaces; 2) the concurrent effect of interacting with in-vehicle voice interfaces and other commonly discussed individual-level factors has barely been targeted. This research work was a secondary data analysis based on a large-scale field experiment wherein 144 participants had been recruited and driven a test vehicle while performing a series of tasks using voice-based interfaces. Pre- and post-drive questionnaires were employed to collect drivers' individual-level-related data. Participants' visual attention while interacting with voice-based interfaces was characterized by off-road glance behavior and recorded by in-vehicle cameras. Structural equation modeling (SEM) was

leveraged to build a theoretical model connecting participants' individual-level factors to their off-road glance behavior while interacting with in-vehicle voice interfaces.

Results from SEM analysis 1) confirmed that driving is complex as participants' off-road glance behavior was significantly affected by multiple factors; 2) found that participants with higher trust in technologies actually tended to have longer off-road glance behavior as compared to those who trusted technologies less, and this might contradict previous findings and the theoretical basis of trust in a technology; 3) participants' age and gender did have an effect on their off-road glance behavior and the findings were generally in line with relevant research; 4) participants' previous usage of voice interfaces did not predict their off-road glance behavior while their preconceptions about technologies did. Although voice-based interfaces are designed to help reduce drivers' visual attention required for interactions, drivers may still direct their eyes off the road and exhibit risky visual behavior while interacting with them. Besides, individual-level factors can also exert influence on drivers' visual behavior in a way that drivers of certain groups might have riskier behavior when interacting with voice-based interfaces. To promote the general public's adoption of in-vehicle voice interfaces and make the interaction safer, accounting for the psychological and physical factors that are properties of the human component is critical.

TABLE OF CONTENTS

ABSTRACT v

TABLE OF CONTENTS vii

LIST OF TABLES x

LIST OF FIGURES xi

CHAPTER

CHAPTER 1 1

1.1. Motivation 1

1.2. Research Method and Objective 3

1.3. Dissertation Organization 5

CHAPTER 2 7

2.1. Driver Distraction 7

2.1.1. Driver Distraction and Safety 7

2.1.2. General Definition of Driver Distraction 8

2.1.3. Sources of Distraction 10

2.1.4. Types of Distraction 11

2.1.5. Distraction and Driving Tasks 12

2.1.6. Distraction in this Dissertation Work 13

2.2. In-vehicle Technologies & Voice Interface 14

2.3. Driver's Visual Attention & Glance Behavior 16

2.4. Factors Affecting Driver's Glance Behavior 18

2.4.1. Driving Tasks 18

2.4.2. Trust 18

2.4.3. Driving Experience 19

2.4.4. Age 20

2.4.5. Other Individual-level Factors 21

2.5. Current Research and Research Gap 22

CHAPTER 3 24

3.1. General 24

3.2. Participants 26

3.3. Experimental Procedure 27

3.3.1. Study 1 29

3.3.2. Study 2 & 3 31

3.4. Data Reduction and Analysis 33

3.4.1. Data Reduction 33

3.4.2. Data Analysis 35

CHAPTER 4 42

4.1. Measurement Model and Assessment 42

4.2. Structural Model 46

4.3. Model Revision 50

4.3.1. Measurement Model 50

4.3.2. Structural Model 54

CHAPTER 5 57

5.1. Discussion 57

5.2. Limitations and Future Work 64

5.3. Contribution to Human Factors 68

5.4. Conclusion 69

BIBLIOGRAPHY 81

LIST OF TABLES

Table Page

Table 1. Variable definitions, collection methods, scales, and statistics 40

Table 2. Model fit indexes for the initially proposed measurement model 43

Table 3. CFA results, internal consistency and convergent validity of the initially proposed measurement model.. 44

Table 4. Model fit indexes for the initially proposed structural model 46

Table 5. Results of hypothesis testing.. 48

Table 6. Model fit indexes of the revised measurement model 51

Table 7. CFA results, internal consistency and convergent validity of the revised measurement model 53

Table 8. Model fit indexes of the revised structural model 54

LIST OF FIGURES

Figure Page

Figure 1. The proposed research model. 5

Figure 2. Graphical representation of participants' age distribution by gender. 27

Figure 3. Overview of the experimental protocol for the on-road drive in study 1 29

Figure 4. Overview of the experimental protocol for the on-road drive in study 2 & 3... 32

Figure 5. The initial SEM model 39

Figure 6. A simplified schematic showing the results of the structural model. Dotted lines indicate non-significant relationships. Note: *** p < 0.001; ** p < 0.01; * p < 0.001. ... 47

Figure 7. The revised SEM model. Only coefficients of the structural model were demonstrated. Dotted lines indicate non-significant relationships. Note: *** p < 0.001; ** p < 0.01; * p < 0.001. 52

Figure 8. Primary screen of the center console, 2014 Chevrolet Impala 71

Figure 9. Additional small display in the instrument cluster, 2014 Chevrolet Impala 72

Figure 10. Push-to-talk button on the steering wheel, 2014 Chevrolet Impala 72

Figure 11. Layout of the DVI showing the main components of the COMMAND system 74

Figure 12. Close-up of the primary display screen 75

Figure 13. Layout of the DVI 76

Figure 14. Closer view of the display screen in the center cluster 76

Figure 15. Location of the push-to-talk button on the steering wheel 77

CHAPTER 1

INTRODUCTION

1.1. Motivation

Driving is a complex task and relies heavily on visual information (Sivak, 1998). During driving, visual senses are believed to be loaded the most as they continuously gather information from the driving environment (Nabatilan, Aghazadeh, Nimbarte, Harvey, & Chowdhury, 2012). Any distraction tasks that take driver's eyes off the road are likely to post a risk to their safety (Perez & Bertola, 2011; Tivesten & Dozza, 2014). Such distraction can be classified as driver distraction and driver distraction has already been identified as one of the leading causes of crashes (Dingus et al., 2006), and can be defined as the diminished attention of the driver to the driving task (Donmez, Boyle, & Lee, 2006). According to the National Highway Traffic Safety Administration, distracted driving claimed 8 percent of fatal crashes, 15 percent of injury crashes in 2018 (NHTSA, 2020). The distraction is often seen when the driver is engaging in secondary tasks (i.e., not related to driving), such as using a cellphone, interacting with the interface of certain in-vehicle technologies, and conversing with others (Donmez et al., 2006; Nabatilan et al., 2012; N. Zhao, Reimer, Mehler, D'Ambrosio, & Coughlin, 2013). These can all possibly impair the allocation of driver's visual attention (Nabatilan et al., 2012; N. Zhao et al., 2013).

With in-vehicle technologies being massively introduced to modern automobiles, more voice-based in-vehicle interfaces are embedded, and accordingly, additional

interactions between the driver and the interface tend to be inevitable. Although in-vehicle voice interfaces are intended to facilitate driving and reduce distraction brought up by the traditional visual-manual based interfaces, they may at the same time still impose undesired distractions on the driver, resulting in impediment to properly recognize various road and traffic conditions, and hence leading to possible crashes (Donmez et al., 2006; Nabatilan et al., 2012; Peng, Boyle, & Lee, 2014). During this process, driver's visual attention may be taken away from the driving environment. It has been stressed that appropriate allocations of visual attention is an essential element of safe operation of the vehicle (Reimer, Mehler, Wang, & Coughlin, 2012). In the meantime, we believe that driver's glance behavior is a direct reflection of how one's visual attention is located. To study driver's glance behavior is therefore of great importance when the driver spares attention to secondary tasks caused by interacting with in-vehicle technologies such as voice interfaces.

Driving behavior, or more specifically, driver's glance behavior is not only affected by one's engagement in secondary tasks, it is also under the influence of individual-level factors (J. Lee, Mehler, Reimer, & Coughlin, 2016; Nabatilan et al., 2012). For example, past research has identified driving experience as a significant factor affecting driver's visual behavior and driving performance (Nabatilan et al., 2012). Others have shown that self-reported information about the Driver Behavior Questionnaire (DBQ) and sensation seeking exerts influence on driving behavior including glance behavior (J. Lee et al., 2016; N. Zhao et al., 2012). Moreover, individual characteristics such as age and gender might also impact one's driving performance, and

driver's interaction with the in-vehicle interface can in turn affect their post-attitudes toward it (C. Lee, Mehler, Reimer, & Coughlin, 2015).

While a large body of research has focused on either how engaging in various in-vehicle-interface-related secondary tasks can affect visual/glance behavior or the relationship between individual characteristics and one's driving behavior, they are often studied separately. That is, less is known about how they are correlated as a whole to affect driver's glance behavior. Given the ubiquitous appearance of in-vehicle technologies and voice-based interfaces, as well as the critical role of visual attention in driving safety, I proposed this dissertation work aiming to fill the gap that previous studies rarely connected driver's glance behavior to both secondary tasks caused by interactions with in-vehicle voice interfaces and individual-level factors. Based on what has been investigated, what would make this dissertation stand out and step further within this field of research lies in the facts that: 1) unlike lab experiments, it is an on-road study in which participants had drove a real vehicle; 2) participants completed secondary tasks via in-vehicle voice interfaces while driving during the experiment; 3) comprehensive data were collected and analyzed for this study, including from the aspect of driving and individual-level factors.

1.2. Research Method and Objective

The focus of this dissertation was driver's glance behavior while interacting with an in-vehicle voice interface, how it is affected by drivers' individual-level factors, and how they are related. Participants were recruited to complete on-road drives with different secondary tasks, including interactions with an in-vehicle voice interface, and answered a

series of survey questions prior to and after the drive. To serve the research goal, I proposed a theoretical model connecting 6 factors, as can be seen in **Figure 1**.

Specifically, “preconceptions” included: 1) whether one saw her/himself as an early adopter of new technologies; 2) self-reported level of experience with technology; 3) self-reported ability to learn to operate new technologies. “Trust” contained: 1) self-reported overall level of trust in technology; 2) self-reported overall level of trust in established car technologies; 3) self-reported level of trust in new technologies that are being introduced to cars. “Individual characteristics” is represented by participants’ age and gender. “Previous usage” consisted of: 1) frequency of using a voice command interface in any environment; 2) frequency of using a car- or truck-based voice command interface system; 3) frequency of using an electronic navigation system in a car or truck; 4) frequency of driving a car or other motor vehicle. “Glance behavior” included: 1) mean single off-road glance duration; 2) number of glances off-road; 3) total eyes off-road time (TEORT); 4) number of single off-road glances greater than 2.0 seconds. And finally, “post attitudes” was: 1) the likelihood that participants would recommend buying a car with voice control systems; 2) how likely one would consider buying a car with in-vehicle voice interfaces for the next car; 3) how the experience with the in-vehicle voice interface changed one’s trust in new technologies that are being introduced into cars; 4) participants’ overall impression of the vehicle they drove during the experiment.

There were 7 hypotheses involved depicting the relationships between the factors. Under each of the factors were a set of variables contributing toward this factor – those illustrated in the previous graph. By adopting structural equation modeling (SEM), we shall

see how the factors were represented by the corresponding variables, as well as how the model I proposed fitted (Fan et al., 2016; Suhr, 2006). In general, the research objectives of this dissertation were to: 1) identify the main factors that affect driver's glance behavior while interacting with an in-vehicle voice interface; 2) quantitatively verify, and if necessary, modify a theoretical model about driver's glance behavior; 3) uncover the complex relationships between drivers' glance behavior and other commonly studied factors. Eventually, I hoped to acquire useful safety implications about driving with voice-based systems and solutions for improvement from multiple facets.

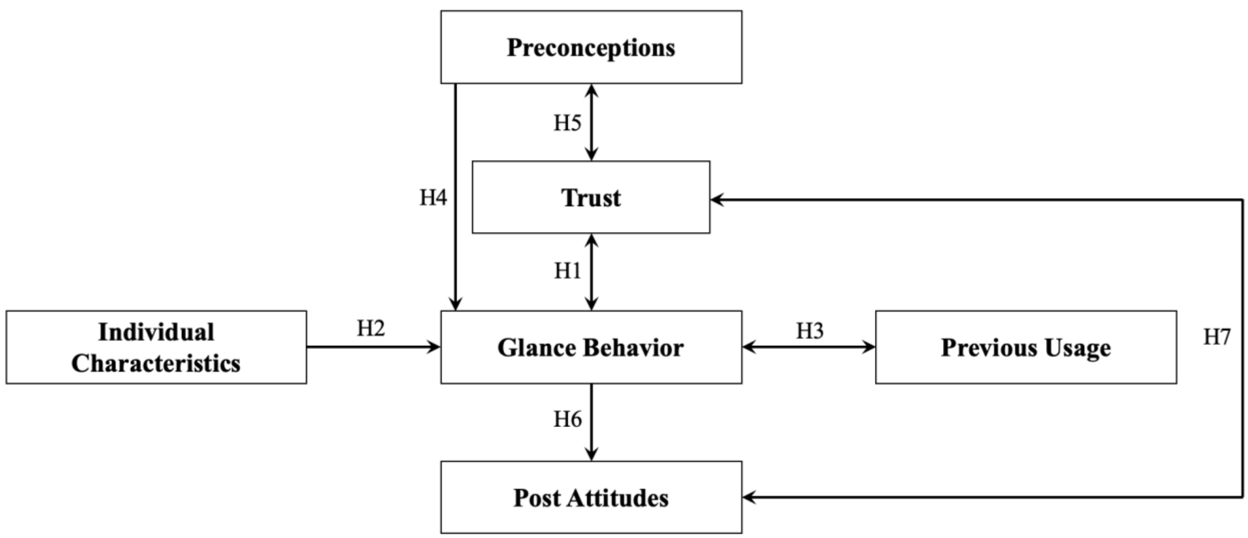

Figure 1. The proposed research model.

1.3. Dissertation Organization

This dissertation is structured in the following way. In Chapter 2, general background about the research questions and literature review are provided. Chapter 3 talks about the research method, in which the experiment and data analysis are illustrated.

Chapter 4 discusses results obtained from this dissertation work. Chapter 5 concludes this dissertation work with the primary findings and their significance and implications, limitations, as well as potential directions for future research.

CHAPTER 2

BACKGROUND & LITERATURE REVIEW

2.1. Driver Distraction

2.1.1. Driver Distraction and Safety

Road traffic injuries are currently estimated to be the ninth leading cause of death across all age groups globally, and are predicted to become the seventh leading cause of death by 2030 (Fernández, Usamentiaga, Carús, & Casado, 2016; World Health Organization, 2015). Closely related to road traffic is driving, which is a complicated task that highly demands driver's attention while driver distraction is believed to diminish driver's attention to the task (Donmez et al., 2006; Horberry, Anderson, Regan, Triggs, & Brown, 2006). There is accumulative evidence showing that driver distraction is one of the leading causes of vehicle crashes and accidents (Dingus et al., 2006; Misokefalou, Papadimitriou, Kopelias, & Eliou, 2016; Papantoniou, Papadimitriou, & Yannis, 2017; Pettitt, Burnett, & Stevens, 2005; Regan, Hallett, & Gordon, 2011; Stutts, Reinfurt, Staplin, & Rodgman, 2001). Based on the record from the National Highway Traffic Safety Administration, distracted driving claimed 8 percent of fatal crashes, 15 percent of injury crashes in the US in 2018; there were 2,841 people killed and an estimated 400,000 people injured in motor vehicle crashes involving distracted drivers (NHTSA, 2020). Collisions caused by distracted driving have already captured the attention from the US Government and professional medical organizations during the past years (Fernández et al., 2016; Llerena et al., 2015).

Since driver distraction is of great safety interest, tons of research has been conducted to examine how the distraction affects driver's driving behavior and the same motivation applied to this dissertation. However, driver distraction and driving behavior are indeed generic ideas and contain multiple facets. For example, quite a few researchers have solely focused on the definition and/or taxonomy of driver distraction (Pettitt et al., 2005; Regan et al., 2011) while another school of research has been investigating the impact of distracting tasks as specific as mobile phone use on driving behavior and crash involvement (Choudhary & Velaga, 2017; World Health Organization, 2011). For now, readers might just need to keep in mind that once driver distraction is clearly defined, tasks that can contribute to it could come from various sources and affect driving behavior in multiple ways (e.g., mean speed, eye movements, hand movements), which is later illustrated in the following sections. Therefore, the state of the art of research in this broad field might hardly be synchronized and summarized within limited paragraphs. In this chapter, accordingly, only research closely related to the focus of this dissertation work is presented, which is basically factors affecting driver's visual attention. Relevant literature review can be found in 2.2, 2.3 and 2.4., although it has been conducted throughout the body of this work.

2.1.2. General Definition of Driver Distraction

There has been a large body of literature discussing driver distraction while the term "driver distraction" itself has also been defined in various ways, and some of them might not be consistent (Peissner, Doebler, & Metze, 2011; Regan et al., 2011). To illustrate, some previous definitions are listed as follows (Regan et al., 2011):

- Driver distraction occurs "whenever a driver is delayed in the recognition of information needed to safely accomplish the driving task, because some event, activity, object, or person [or outside] his vehicle, compelled or tended to induce the driver's shifting of attention away from the driving task" (Treat, 1980).
- "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 ability and/or performance, leading to an increased risk of corrective actions, near-crashes, or crashes" (Hedlund, Simpson, & Mayhew, 2006).
- "Driver distraction is the diversion of attention away from activities critical for safe driving toward a competing activity" (J. D. Lee, Young, & Regan, 2008, pp. 34).
- Driver distraction results "from interference between a driving task and an external stimulation without link with driving" (Hoel, Jaffard, & Van Elslande, 2010).

While approaches like these shaped the idea of "driver distraction" from different angles, they all to some degree reflect several key features in driver distraction: diversion of attention from safe driving; attention diverted to non-driving tasks; adverse effect on driving safety (Regan et al., 2011). It should be noted that the term "driver inattention" is also frequently used, to my best knowledge, there is lack of consensus in the literature about what is meant by "driver distraction" and "driver inattention", and their relationship remains unclear (Fernández et al., 2016; Regan et al., 2011). However, clearly-defined terminologies are desired for research work to ensure the findings are interpretable and comparable. Although driver distraction still remains a poorly and inconsistently defined

concept (World Health Organization, 2011), I decided to follow the definition of driver distraction which states that “driver distraction is the diversion of attention away from activities critical for safe driving toward a competing activity” in this dissertation (J D. Lee et al., 2008). I also differentiate “driver inattention” from “driver distraction” by stating that driver inattention means insufficient or no attention to activities critical for safe driving, and that driver distraction is just one form of driver inattention (Regan et al., 2011).

There is, however, potential issue associated with the definition of driver distraction adopted here – how “critical activities for safe driving” should be defined. Unfortunately, no explicit answers have been found due to the difficulty in defining a prior what the activities might be (Regan et al., 2011). In the present work, I chose to avoid identifying such activities that are critical for safe driving. Instead, the idea of primary and secondary tasks was employed (illustrated in 2.1.5), according to which engaging in “secondary tasks” can be safely viewed as uncritical for safe driving and hence induce driver distraction (J D. Lee et al., 2008; Peissner et al., 2011). In other words, the logic here is simply that driver distraction remains how it is defined while secondary tasks are regarded as the ones that can potentially cause driver distraction as they shall never be critical for safe driving.

2.1.3. Sources of Distraction

Depending on the source of where the distraction comes from, driver distraction is usually divided into two categories: distraction from internal and external of the vehicle (Horberry et al., 2006; World Health Organization, 2011). Internal (in-vehicle) distractions include eating, smoking, talking, using mobile phones as well as using in-vehicle entertainment systems (e.g., use of radio, CDs) (Choudhary & Velaga, 2017; World Health

Organization, 2011). Among all of these distraction sources, mobile phone use is believed to be quite prevailing (Choudhary & Velaga, 2017; Horberry et al., 2006). But the growing number of in-vehicle technologies/systems are of most concern to those involved in road safety (Fernández et al., 2016; World Health Organization, 2011).

Driver distraction is not just related to what is happening inside the vehicle, but also to what is presented outside (Horberry et al., 2006). External (outside-vehicle) distractions might arise when the driver looks at surroundings such as buildings, people or situations outside the vehicle, as well as at billboards and other roadside advertising (World Health Organization, 2011). Researchers have found that new highway developments, a large number of roadside advertisements and street-side vendors and an increased traffic flow all contribute to external distraction to the driver (Horberry et al., 2006).

2.1.4. Types of Distraction

The definition here shall be viewed parallel to the definition in 2.1.3., it is concluded under the broad concept of driver distraction, and yet based on types. In this dissertation, I concluded four types of distraction (Cunningham, Regan, & Imberger, 2017; Papantoniou et al., 2017; World Health Organization, 2011):

- visual distraction is defined as driver distraction triggered by a competing visual activity (e.g., looking at mobile phones);
- auditory distraction is defined as driver distraction triggered by a competing auditory activity (e.g., responding to a ringing mobile phone);

- cognitive distraction is defined as driver distraction triggered by internal thought (e.g., reflecting on a subject of conversation as a result of talking on the phone);
- physical distraction is defined as observable interference with vehicle control by a driver interacting physically with a technology (e.g., leaning over to tune the radio).

It should be noted that multiple types of distraction can occur simultaneously, depending on the specific task the driver is engaging in.

2.1.5. Distraction and Driving Tasks

We've often seen tasks that can impose distraction on drivers, such as using a cell phone, conversing with passengers and tuning the radio, and most of them are not necessarily needed for driving itself. In this regard, two types of driving tasks can be further classified: the primary task and secondary task, which are defined as below (Peissner et al., 2011):

- The primary task is defined as the actual driving task, keeping the vehicle on the road while obeying the traffic regulations and being thoughtful towards other traffic participants. The primary task includes physical actions such as braking and depressing the gas pedal.
- Secondary tasks are not part of the natural driving response, but function to please the comfort and entertainment needs in a car just as selecting music from the music player, receiving and indicating a call, entering some data to the navigation system. Secondary tasks might divert the driver's attention away from the driving task.

Depending on how driver distraction is defined, researchers have employed different sets of terms to describe driving tasks. One most commonly seen is "driving related" and "non-driving related" tasks. However, criticisms could be levelled at the definition for referring to "non-driving related" tasks (Pettitt et al., 2005). Take entering addresses into a navigation system while driving as an example, one might argue that this is indeed driving-related as some drivers prioritize this as part of the driving whereas others are likely to view this as unnecessary to driving. Given that the boundary between driving and non-driving related tasks might be blur, I therefore adopted the naming convention "primary tasks" and "secondary tasks" throughout this work.

2.1.6. Distraction in this Dissertation Work

Following the definition and classification fashions illustrated in the aforementioned content, it can be summarized that driver distraction happens when other competing activities divert the driver's attention from driving. Secondary tasks are believed to be a type of these activities and can often induce driver distraction. Driver distraction is either from external or internal vehicle, with internal distraction causing more safety concerns (Fernández et al., 2016; World Health Organization, 2011). Further, it is believed that when compared to other common in-vehicle secondary tasks such as using a mobile phone, the interaction caused by the increasing number of in-vehicle interfaces/systems plays a bigger part regarding road safety (Fernández et al., 2016). This inspired me and brought us to the general idea of this dissertation work: how driver distraction caused by interacting with in-vehicle interfaces affects driving behavior.

2.2. In-vehicle Technologies & Voice Interface

In-vehicle technologies/systems are becoming increasingly popular and extensively used (Peissner et al., 2011). Due to the demand from drivers for more connectivity and advanced entertainment options while driving, the systems are now widely embedded in modern automobiles via various types of in-vehicle interfaces (Reimer et al., 2014). While they are intended to facilitate driving and optimize driving experience, they can cause additional distraction that is undesired. It has been stressed that safety problems related to driver distraction are expected to escalate as more technologies become available for use in vehicles (Owens, McLaughlin, & Sudweeks, 2011; Stutts et al., 2001). Because the increase in use of those in-vehicle technologies can induce visual, physical and cognitive distraction, which might negatively affect driving performance (Fernández et al., 2016). Past research has already pointed out that the magnitude of the effect is likely to increase if the task requires the driver to glance off the road, for example, the action of manually interacting with in-vehicle interfaces (Garay-Vega et al., 2010).

Driving itself is a complicated task that highly relies on visual information, any task that takes driver's eyes off the road is therefore dangerous (Sivak, 1998). And this is why over the past years, there has been a shift in automotive driver-vehicle interfaces (DVIs) from purely visual-manual interactions to include options for voice-based or voice-assisted interaction (Reimer et al., 2014). It is thought that interacting with a technology by auditory/vocal modality can occur while the driver's eyes are oriented toward the road and her/his hands are on the wheel, thus minimizing the visual attention that needs to be directed to the interface as well as physical actions required to touch any buttons, screens,

or other controllers (Reimer et al., 2014). That is, using voice-based systems might help reduce the demand required for completing the same task by visual-manual systems (Simmons, Caird, & Steel, 2017). Hence, it is not surprising that voice interfaces have become a popular feature in many production vehicles (Reimer et al., 2014).

However, the frequent use of in-vehicle voice interfaces in automobiles has already raised safety concerns. The concerns are often reasonably founded and frequently met with reassurances that voice interfaces are far safer than visual-manual input methods (McWilliams, Reimer, Mehler, Dobres, & McAnulty, 2015). Research focusing on in-vehicle voice interfaces and driver distraction has been directed in the following ways. Some researchers have devoted time to developing an understanding of and assessing the safety, usability, and demand related aspects of voice-interaction in the vehicle (B. Mehler et al., 2014). Others used measures such as stimulus response time, cognitive load, and lane position to assess demands placed on the driver. Other studies, at the same time, focused on voice recognition accuracy, task time, and user preference. Regarding the visual demands associated with the interaction with in-vehicle voice interfaces, only a small number of studies have addressed it in a way that the findings can be replicated in other research (Reimer et al., 2014). While some researchers found that the amount of visual demand for voice controlled activities such as navigation destination entry, as measured by total off-road glance time, exceeded the threshold set by the National Highway Traffic Safety Administration for performing tasks of similar difficulty via visual-manual interaction (Reimer et al., 2014), others stated that drivers' visual demands when performing a given task using a voice-based interface were substantially under the criterion set for visual-manual interfaces (B. Mehler, Reimer, Dobres, & Coughlin, 2015).

Nonetheless, drivers' engagement of visual attention associated with the use of in-vehicle voice interfaces could be substantial and shall not be ignored (Chiang, Brooks, & Weir, 2005; Reimer et al., 2014).

Given that the strongest motivation of the evolution of in-vehicle interfaces from visual-manual based to voice-command based lies in the reduction of the distraction mainly impairing driver's visual attention, plus driving itself is essentially a visual task, less is known regarding exactly how driver's visual attention is allocated when interacting with in-vehicle voice interfaces (McWilliams et al., 2015; B. Mehler et al., 2014). Visual attention, on the other hand, is often measured by eye-movement metrics and more specifically, glance behavior (Bakhit, Osman, Guo, & Ishak, 2019; McLaughlin, Hankey, & Dingus, 2009).

2.3. Driver's Visual Attention & Glance Behavior

The primary input to driving or primary tasks that maintain the vehicle's trajectory is visual, which is by far the most important source of information for the driver (Underwood, 2007). During driving, visual senses are loaded the most as they continuously gather information from the driving environment (Nabatilan et al., 2012). Hence, driving is believed to be a highly visual task. Any distraction tasks that take driver's eyes off the road is much likely to threaten her/his safety (Perez & Bertola, 2011; Tivesten & Dozza, 2014). Visual attention has already been identified as a contributing factor to traffic crashes (Crundall, Shenton, & Underwood, 2004; Konstantopoulos, Chapman, & Crundall, 2010). While interacting with in-vehicle interfaces can impose various distractions on the driver (e.g., visual distraction, cognitive distraction), researchers have found that visual

distraction interferes with driving performance more than cognitive distraction and dominates distraction-related decrements to driving performance when the driver is both visually and cognitively distracted (Fitch, Bartholomew, Hanowski, & Perez, 2015; Liang & Lee, 2010). Thus, for the developers and engineers, driving is a task in which visual aspects must be considered in the design of information system (Oh, Ko, & Ji, 2016; Owsley, 2011).

Due to the highly visual information content employed in driving, monitoring the eyes is considered valuable for making a number of inferences (McLaughlin et al., 2009). And eye-movement metrics are the most promising diagnostic metrics for measuring driver distraction (Bakhit et al., 2019). Glance locations generally include driving-related locations, including the forward roadway scene, in-vehicle interfaces, etc. (McLaughlin et al., 2009). Accordingly, more time looking ahead toward the forward road scene would be ideal, whereas longer and frequent glances off-forward-roadway might cause risk. In this dissertation, glance behavior was quantified in the section that follows considering glance off-forward-roadway, which basically followed the NHTSA guidelines (2013). In the presence of secondary tasks, total glance time, mean single glance time, and number of glances are often considered surrogate measure of safety, which were all utilized in this dissertation work (McLaughlin et al., 2009). A single eye glance off the forward roadway for longer than 2 seconds while interacting with an in-vehicle interface is said to be safety critical (Geitner et al., 2017; National Highway Traffic Safety Administration, 2013), the measure was therefore also included in this work.

2.4. Factors Affecting Driver’s Glance Behavior

2.4.1. Driving Tasks

In accordance with the definitions in this dissertation work, driving tasks consist of primary tasks and secondary tasks while secondary tasks are believed to be the one that can cause driver distraction. And by the distraction, secondary tasks can often impair driver’s visual attention (Fitch et al., 2015; Liang & Lee, 2010). In other words, secondary tasks could be responsible for driver’s glance behavior. Specifically, these tasks have been shown to be texting and calling via a mobile phone (B. Mehler et al., 2016; Tivesten & Dozza, 2014); interacting with in-vehicle interfaces (Horberry et al., 2006; J. Lee, Caven, Haake, & Brown, 2001; Zheng, Shokouhi, Thomsen, Sathyanarayana, & Hansen, 2016); looking at outside vehicle elements such as advertisements and road elements (Misokefalou et al., 2016). It can be concluded from our previous definitions about sources of distraction again that tasks affecting driver’s glance behavior can be from both internal and external of the vehicle, while researchers believe that in-vehicle tasks do impair driver’s performance and in particular, interacting with in-vehicle systems might have the most negative impact (Horberry et al., 2006).

2.4.2. Trust

Trust involves the formation of reliance on another actor, person or device, to fulfill a given role for the trustor. The level of trust in a technology is believed to naturally influence how people interact with it (Geitner et al., 2017; J. D. Lee & See, 2004). Trust has been defined in various ways by past literature, and yet the definitions are believed to all include the element of one’s expectation regarding behaviors or outcomes and

willingness to be vulnerable to the actions performed by another party based on the expectation that the other will perform the important task, without monitoring or controlling that party (J. D. Lee & See, 2004; Mayer, Davis, & Schoorman, 1995).

In the field of human factors in transportation/driving safety, existing work has connected trust to drivers' visual attention, primarily with the presence of drivers' interaction with driving automation systems (Hergeth, Lorenz, Vilimek, & Krems, 2016; Noble, Miles, Perez, Guo, & Klauer, 2021). While certain levels of driving automation are able to assist driving or even perform the main task under designated contexts, drivers are still required to monitor driving. This is why human factors researchers are particularly interested in drivers' visual attention when the driving task is partially or fully performed by the system (Noble et al., 2021). Over trusting or not trusting the system may potentially result in improper driving behavior, including visual behavior. Unlike driving automation systems, most of the in-vehicle technologies are not intended to take the responsibility of driving from drivers, therefore, any interactions with them that take the drivers' eyes off the road could be riskier. Regarding the specific modality that is studied in this work - in-vehicle voice interfaces - very limited work has investigated how trust affects drivers' visual attention, and it has been suggested that more work is desired to explore this aspect (Geitner et al., 2017).

2.4.3. Driving Experience

Inexperienced drivers are believed to be vulnerable to road traffic accidents (Underwood, 2007). Researchers have suggested that the efficiency of visual search strategies is one of the fundamental changes in skill that marks the transition from novice

to experienced driver (Konstantopoulos et al., 2010; Underwood, 2007). That is, experienced drivers shall have “safer” visual behavior relative to those that are still novice. Empirically, differences have been observed in driver’s scanning behavior such that experienced drivers increase their scanning behavior when encountered more complex road conditions, whereas novice drivers do not tend to have this sensitivity (Underwood, 2007). But whether experienced drivers still have “safer” visual behavior while interacting with in-vehicle interfaces, especially voice interfaces, appears to be missing from our research database. Therefore, I decided to take driver’s experience into account in this dissertation.

2.4.4. Age

The percentage of the population in the US and many other countries over 60 is increasing, and thus eye conditions, diseases, and vision impairments associated with aging represent a larger segment of the societal health challenge (Owsley, 2011). Many visual difficulties occur with age, such as decline in acuity and peripheral field loss (Hoffman, McDowd, Atchley, & Dubinsky, 2005). It is therefore important to see how age affects those people’s daily tasks, especially those that require high visual demand. A critical one would be driving.

Findings of age effect on driver’s glance behavior from previous work appear to be slightly mixed. Some literature has shown that age affects driver’s glance behavior measures such as total eyes off-road time and mean glance duration in a way such that they all become longer as age increases (B. Mehler et al., 2014). Some also strengthen this by stating that older drivers tend to have longer total time of gaze concentrated on the distracting stimuli as compared to younger drivers (Misokefalou et al., 2016). Another

group of researchers has concluded that age directly predicts visual impairment (Hoffman et al., 2005). But at the same time, other researchers did not find age significant in affecting driver's gaze concentration with added cognitive demand (Reimer et al., 2012). Therefore, further research is desired to examine the effect of age on driver's glance behavior to be added to either school of the existing body of research (i.e., to provide more empirical evidence).

2.4.5. Other Individual-level Factors

Factors related to the driver, or what I called individual-level factors in this dissertation including age and gender are said to be highly significant in driver's behavior when distracted (Misokefalou et al., 2016). Research on the effect of age and driving experience on driver's glance behavior has been introduced in the precedent sections, I will talk about other individual-level factors here.

While some researchers did not find significant difference between female and male drivers in total gaze concentration time toward distracting stimuli (Misokefalou et al., 2016), others showed that male drivers had higher percentage of long duration glances (> 2s) and total eyes off-road time as compared to female drivers (B. Mehler et al., 2014). Besides gender, driver's preconceptions about in-vehicle technologies/systems and trust in technologies have also been shown to affect driving performance (C. Lee et al., 2015). It should be noted that these variables are extremely rarely studied regarding driver's specific glance behavior when interacting with in-vehicle interfaces, they are therefore of great research interest and novelty.

2.5. Current Research and Research Gap

The main research gap between this dissertation work and past research is the following:

- limited work addressed the evaluation of the visual demands associated with a driver engaging with in-vehicle voice interfaces (B. Mehler et al., 2014);
- individual-level factors such as age, gender and driving experience were barely taken into consideration when studying driver's glance behavior while interacting with in-vehicle voice interfaces, although they had been shown to significantly affect driving behavior including visual behavior;
- to the best of my knowledge, the concurrent effect of interacting with in-vehicle voice interfaces and other individual-level factors on driver's visual behavior has not been targeted.

Therefore, the objective of this dissertation work was to identify the main factors affecting driver's glance behavior when interacting with in-vehicle voice interfaces, and how the factors are related to each other. From the proposed theoretical model (**Figure 1**), the research hypotheses were:

- H1: There is a relationship between driver's trust in technology and glance behavior such that when one has higher trust in technology, she/he has less off-road glance behavior, and vice versa.
- H2: Effect of individual characteristics on glance behavior: drivers of different individual characteristics (i.e., age and gender) will have different glance behavior.

- H3: There is a relationship between driver's "previous usage" and glance behavior such that when one has more frequent "previous usage", she/he has less off-road glance behavior, and vice versa.
- H4: Effect of preconceptions on glance behavior: drivers with more positive preconceptions will have less off-road glance behavior.
- H5: There is a relationship between driver's preconceptions of technology and trust in technology such that when one has more positive preconceptions about technology, she/he has higher trust in technology, and vice versa.
- H6: Effect of glance behavior on post attitudes: driver's glance behavior affects her/his post attitudes toward the voice interface, those who have less off-road glance behavior will have more positive post attitudes.
- H7: There is a relationship between one's trust in technology and her/his post attitudes toward the voice system such that higher trust in technology will be associated with more positive post attitudes, and vice versa.

CHAPTER 3

METHODOLOGY

3.1. General

This section illustrates the experiment and procedure, as well as data collection and analysis. Although the present dissertation work is essentially secondary data analysis and full details regarding the experimental setup can be found elsewhere (B. Mehler, Reimer, Dobres, & Coughlin, 2015; B. Mehler, Reimer, Dobres, McAnulty, & Coughlin, 2015; B. Mehler, Reimer, McAnulty, et al., 2015), I am still introducing the experimental part here, hoping that that would offer the readers a better sense of the original experiment and particularly what kind of data I am focusing on, as well as under which context they were gathered.

To begin with, the general idea guiding the experimental part of the present research work was to have participants drive real vehicles while completing secondary tasks using in-vehicle voice interfaces via a filed study. By collecting their driving data, especially glance data, as well as responses to various questionnaires including demographic information, we would then be able to obtain a rich body of data centered around driver's glance behavior for further investigation.

Broadly speaking, the whole experiment consisted of three independent experiments, sequentially, they were named study 1, study 2 and study 3. This naming fashion will hold and be used throughout this dissertation. All studies were nearly identical besides that: 1) participants of each study were totally different, there were no duplicated

exposures to multiple studies for any individual; 2) the apparatus adopted was different across the three studies, that is, the vehicle and associated interface that participants drove and interacted with varied; 3) the order and randomness of the tasks that participants were required to complete during the drive of study 1 was slightly different from that of study 2 and study 3 due to road construction and optimization of the experimental design. Study 1 utilized a 2014 Chevrolet Impala equipped with the MyLink system (a detailed look at the interface is shown in Appendix A), study 2 was undertaken in a 2014 Mercedes CLA with its COMMAND system while study 3 was conducted using a 2015 Toyota Corolla and its Entune Premium Audio system (both system interfaces are demonstrated in Appendix A). While they did share few dissimilarities and were conducted separately and at different times, the three studies were analyzed as an entity in this dissertation because the research focus was factors affecting driver's glance behavior, rather than potential differences in driving behavior resulting from different vehicles one drove.

As aforementioned, this dissertation work focused on secondary data analysis and was essentially based on a collaborative project with researchers from the AgeLab at Massachusetts Institute of Technology (MIT). The MIT AgeLab took the lead in recruiting participants and conducting experiments. Therefore, some facts and descriptions about the experiments and participants used throughout this section were borrowed from their preliminary reports (B. Mehler, Reimer, Dobres, & Coughlin, 2015; B. Mehler, Reimer, Dobres, McAnulty, et al., 2015; B. Mehler, Reimer, McAnulty, et al., 2015). To note, the direction of this dissertation work was never tackled in any of the MIT AgeLab's reports.

3.2. Participants

There were 48 usable participant cases equally balanced by gender and across four age groups (20-24, 25-39, 40-54 and 55-69 years) in each study. Since there were 3 studies, the total number of valid cases after combination should have equaled 144. The average age was 39.9 years (SD = 15.86 years). However, my further data preprocessing performed on the raw data showed that in study 3, one participant's demographics and other information originally collected via pre- and post-questionnaires went missing whereas his/her driving-related data were properly recorded. Additionally, there was one more missing value from other participants in 3 of the survey questions, and 2 more missing values in 1 of the questions while all the driving data remained complete. Missing data were handled by data imputation via "mice" package in R (Buuren & Groothuis-Oudshoorn, 2010). Data imputation was run 5 times, which accordingly yielded 5 "suggested" values for each missing value and I randomly selected the fourth option for the missing values. As a result, the available cases were still 144.

The age groups adopted corresponded to the age distribution recommended by NHTSA (2013) for the assessment of visual-manual driver distraction for in-vehicle electronic devices, with the exception of not recruiting 18- and 19-year-olds. This recommendation was followed because an essential piece of the experiment was driver's interaction with in-vehicle voice interfaces. The distribution of participants' age and gender is presented in **Figure 2**. All participants needed to meet the following criteria:

- A valid driver's license for more than three years
- Driving on average three or more times per week

- Being in self-reported reasonably good health for their age and meeting a set of health exclusion criteria
- Clearly understanding and speaking English
- No police reported accident in the past year
- Not actively using any medications causing drowsiness
- Not having been a participant in an AgeLab (affiliated to MIT) on-road driving study in the past 6 months

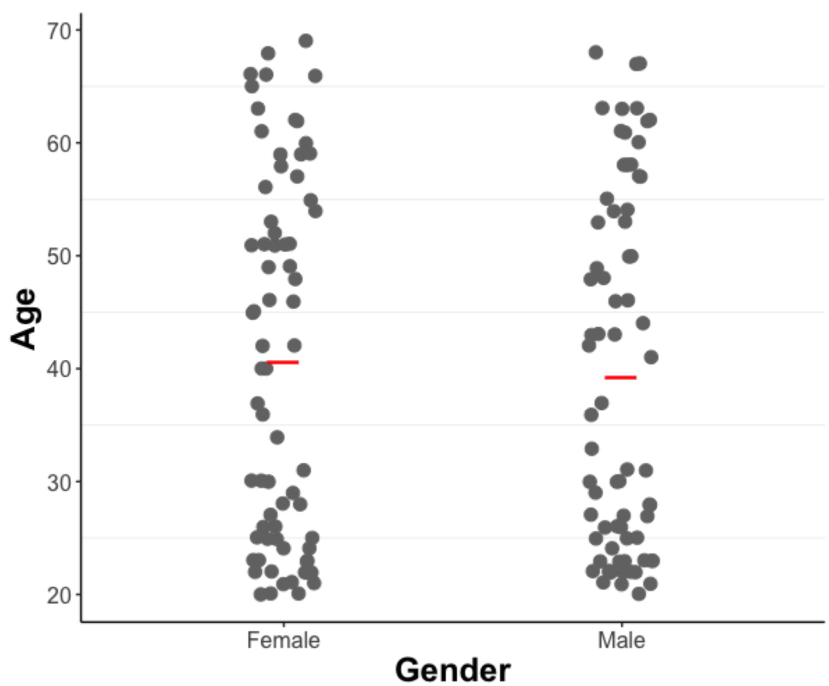

Figure 2. Graphical representation of participants' age distribution by gender.

3.3. Experimental Procedure

Participants were first given pre-experiment questionnaires gathering information about their demographics, driving behavior and history, general perceptions toward

technologies, as well as health and well-being. These questionnaires were filled out in a lab environment before participants had any interactions with the vehicle or in-vehicle interface. Participants were then physically introduced to the vehicle and the corresponding in-vehicle interface. Regarding the tasks that participants would later be performing during the drive, specifically, they were:

- Voice-based interface tasks
  - Full address destination entry
  - Point of interest destination entry
  - Cancel navigation (for each of above)
  - Contact calling (single & multiple phones per contact)
- Visual-manual tasks (radio tuning)
  - Single press preset selection – Radio Easy
  - Specified station manual radio tuning – Radio Hard
- N-back (auditory-vocal-cognitive calibration reference task)
  - 3 demand levels (0, 1, & 2-back)

As shown in **Figure 3** and **Figure 4**, participants received training twice at the MIT parking lot and a rest stop on I-495 depending on the task that they would complete right after. Once participants completed the first training and felt familiar with the vehicle and the system, they would drive onto I-93 North for a period of adaption. Task completion occurred under actual highway conditions on I-495 outside the greater Boston area and on

I-93 south of the interchange with I-495, heading back toward Boston. The highway sections of I-495 utilized consisted of three travel lanes in each direction, are boarded largely by forest, and had posted speed limit of 65 mph. Until now (adaptation after training at the MIT parking lot), all studies followed the same procedures (the actual content of the first training might differ based on how tasks were ordered), but as mentioned previously, the flexibility for the balancing of task order across the protocol was constrained for study 1. Hence, the procedures after adaption of study 1 and study 2 & 3 were different and will be illustrated separately after this point.

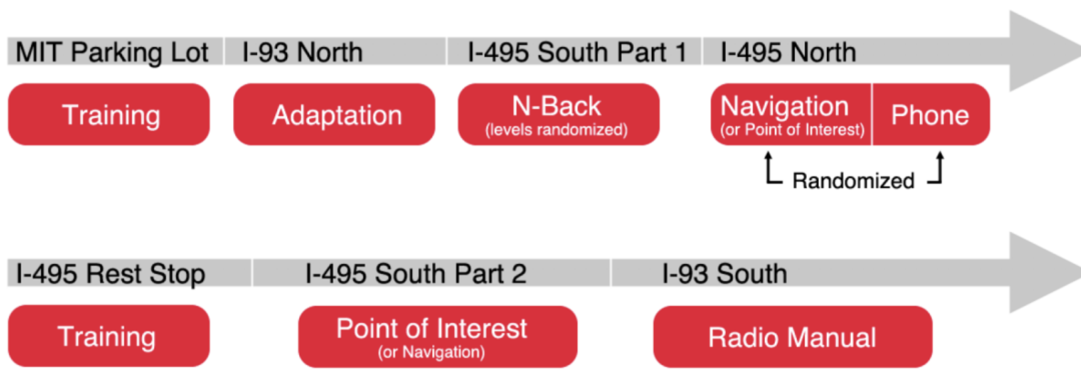

Figure 3. Overview of the experimental protocol for the on-road drive in study 1

3.3.1. Study 1

Figure 3 shows the protocol of the on-road drive for study 1, due to road construction on portions of the I-495 segment of the driving course, the order of tasks was different from that of study 2 & 3. Specifically, the n-back cognitive reference task set was always presented first in the protocol and manual radio tuning was always presented the last task set. The navigation address entry, point of interest selection (POI), and the contact

phone calling task sets were presented in a counterbalanced order across the sample as task sets 2, 3, and 4 with the constraint that the phone tasks were never presented as set 4.

The voice-command based tasks were: full address entry into the navigation system, selection of specified points of interests (POIs), and phone contact calling. Further, reference tasks consisted of visual-manual radio tuning (single button press preset station selection and the more intensive radio reference running task) and the n-back audio-vocal-cognitive calibration reference task at the 0-, 1-, and 2-back levels. Details about the tasks and steps to complete them using the MyLink system in study 1 can be found in Appendix B.

In-vehicle training on the n-back task always occurred in the MIT parking lot along with contact phone calling and either the navigation address entry task or the POI selection task, depending on which one was presented before the I-495 mid-experimental rest stop. Training on the remaining navigation system task (address entry or POI selection) and on the manual radio tuning tasks was provided at the I-495 rest stop.

Questionnaires-based assessment of the participants' experiences with the tasks were obtained at the rest stop for tasks completed up to that point. Experiences related to the remaining tasks were obtained back at MIT in the parked vehicle prior to reentering the research building where additional questionnaire-based evaluations were obtained and final debriefing took place.

3.3.2. Study 2 & 3

As demonstrated in **Figure 4**, the n-back and POI entry tasks were paired together as were the contact phone calling tasks and destination address entry into the navigation system. The tasks were ordered in a way such that half the participants experienced the n-back and POI entry tasks during the first portion of I-495 drive and half on the second, and the ordering of the tasks within a paring was randomized across the sample. The manual radio tuning reference tasks were always presented last on the return route on I-93.

The voice-command based tasks consisted of full address entry into the navigation system, selection of specified points of interests (POIs), and phone contact calling. Reference tasks consisted of visual-manual radio tuning (single button press preset station selection and the more intensive radio reference running task) and the n-back audio-vocal-cognitive calibration reference task at the 0-, 1-, and 2-back levels. The task categories were identical to that in study 1. Details about the tasks and steps to complete them using the COMMAND system (study 2) and Entune Premium Audio (study 3) can be found in Appendix B. However, adjustments were made in study 2 & 3 to reduce the total duration of the experiment based on experiences in study 1 that suggested the length of the study was approaching the comfort threshold of some participants.

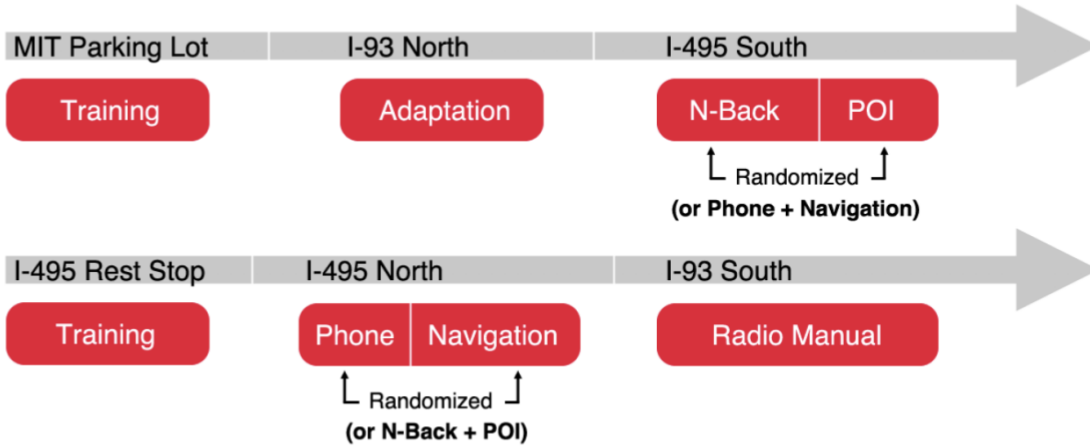

Figure 4. Overview of the experimental protocol for the on-road drive in study 2 & 3

Specifically, contact phone calling was identical in all three studies (i.e., calling 4 specified contacts, 2 "easy" and 2 "hard"). In study 2 & 3, the full address destination entry task consisted of the same first 3 addresses employed in study 1 (2 study specified addresses followed by the participant's home address). The 4th address used in study 1 was dropped to reduce total experimental time and demand on participants. The point of interest (POI) task used in study 1 was reduced from 4 entries to 3 as well to reduce total experimental duration and demand on participants. One of the restaurants used previously in study 1 did not exist in the CLA's (vehicle used in study 2) database, and hence, a restaurant with similar name and location in the alphabetic listing was substituted in study 2 & 3. Also, a change was made in the radio tuning reference task in study 2 & 3 to account for one of the original target radio stations no longer being available. The new start and target stations were selected such that an identical manual tunning distance was involved.

Identical to study 1, the same questionnaires-based assessment of the participants' experiences with the tasks were obtained at the rest stop for tasks completed up to that

point. Experiences related to the remaining tasks were obtained back at MIT parking lot prior to reentering the research building where additional questionnaire-based evaluations were obtained and final debriefing took place.

3.4. Data Reduction and Analysis

3.4.1. Data Reduction

In the phase of data reduction, all three studies followed the same steps. According to the study protocol, as can be seen in **Figure 3** or **Figure 4**, there were totally four task periods during each trial occurring on the I-495 portion. And single task driving reference periods, as the name suggests, refer to the periods of time when the participant was driving freely or, “just driving”. These periods were calculated for 2 minutes prior to a recorded audio message indicating the start of a new task period, and hence there were also four “just driving” periods corresponding to the four tasks on the I-495 portion of the drive. They were just prior to the N-back, destination address entry, contact phone dialing, and POI entry task periods. In the present research work, the focus was specifically on the task of “Navigation” (the full address destination entry of voice-based interface tasks). Like introduced in 3.3.2, participants in study 1 completed four Navigation tasks while in study 2 and study 3, they completed three. The fourth task in study 1 was dropped in study 2 & 3 while the first three tasks in study 1 could be treated the same as the three tasks that participants in study 2 & 3. Accordingly, I chose to focus on these three same Navigation tasks that all participants had completed and average the glance-related metrics across the three tasks for each participant.

Eye glance measures were quantified following ISO standards (ISO 15007-1, 2002; ISO 15007-2, 2001) with a glance to region of interest defined to include the transition time to the object/region. In the manual coding of video images, the timing of glance was labeled from the first video frame illustrating movement to a “new” location of interest to the last video frame prior to movement to a “new” location. Glance data were manually coded based on video of the driver following the taxonomy and procedures outlined in previous research (Reimer, Mehler, Dobres, & Coughlin, 2013). A software that allowed for rapid frame-by-frame review and coding was available, each task period of interest was independently coded by two evaluators. Any discrepancies between the two evaluators (the identification of conflicting glance targets, missed glances, or glance timing that differed by more than 200 ms) were mediated by a third evaluator.

Glance behavior was quantified in the section that follows considering glances off-the-forward-roadway, per the NHTSA guideline (National Highway Traffic Safety Administration, 2013). Therefore, the glance data here were concerned with participants’ off-road glance behavior. To be more specific, four metrics were chosen and utilized as the measures of participants’ visual attention: 1) mean off-road glance duration: the average length of time that participants’ single glance was off the road under a certain Navigation task; 2) number of off-road glances: the total number of participants’ off-road glances under a certain Navigation task; 3) total eyes off-road time (TEORT): the total duration that participants’ glance was off the road under a certain Navigation task; 4) Number of glances greater than 2 seconds: the number of off-road glances greater than 2 seconds for a certain Navigation task. Again, it should be noted that these four measures were originally recorded based on individual tasks. As aforementioned, during the Navigation task period,

there were multiple full address destination entry tasks completed by the participants and three of them were analyzed in the present work while I chose to use average values. Therefore, each of the four metrics were first added up and then divided by the number of tasks (3) participants performed during the Navigation period. All other relevant variables were collected by the pre- and post-questionnaires, and the descriptive statistics can be seen in **Table 1**.

3.4.2. Data Analysis

The main analytical technique employed was SEM. SEM is a very general statistical modeling technique which is widely used in behavioral sciences (Hox & Timo, 1998). It uses various types of models to depict relationships among variables, with the goal of providing a quantitative test of a pre-defined theoretical model by the researcher (Schumacker & Lomax, 2010d; Sheykhfard & Haghighi, 2020). In other words, SEM actually helps test multiple sub-models, often regression, path, and confirmatory factor models under the frame of the entire research model (Schumacker & Lomax, 2010d). It refers to a structure of several models that are linked in a single method (Mohamed & Bromfield, 2017). Eventually, the theoretical model will be tested as to how it fits the sample data, and a more appropriate/better model might be obtained.

There are two major types of variables commonly discussed in SEM: latent variables and observed variables (Schumacker & Lomax, 2010d). Latent variables (constructs or factors) refer to variables that cannot be directly observed or measured from the sample data while observed variables, per the name suggests, are those that can be directly observed. Latent variables are often inferred from or indirectly defined by a set of

observed variables (Schumacker & Lomax, 2010d). Whether or not they are observed, the variables can also be defined as independent variables or dependent variables (Schumacker & Lomax, 2010d). An independent variable is a variable that is not influenced by other variables whereas a dependent variable is influenced by others (Schumacker & Lomax, 2010d).

Closely related to the variables are the primary types of analysis in SEM. Typically, SEM can be viewed as a combination of factor analysis and regression or path analysis (Hox & Timo, 1998). These analyses, again, are performed on the "sub-models" that the researcher pre-defined under the whole structure equation model, and will help test how the depicted relationships and the models hold. For example, if the analysis indicated that a given correlation between two latent variables was not significant, the researcher might want to consider removing this correlation between the variables in the model. Usually, the theoretical model that one initially proposed would consist of two "sub-models": measurement model and structural model. The measurement model is specified to define the relationships between the latent variables and the observed variables while the structural model is specified to depict the relationships among the latent variables (Schumacker & Lomax, 2010b). Accordingly, a school of research suggests that the development and testing of structural equation models follow the steps that the measurement models be established first and then structural models be formed (Schumacker & Lomax, 2010e). I followed the fashion in establishing and testing the proposed research model in this dissertation work.

In the present work, data analysis was performed in R (R Core Team, 2020), and the statistical significance level (α) was set to be 0.05 throughout the study where statistical test was performed. SEM was employed to assess the relationships and model fit of the initially proposed theoretical model. Readers can revisit the diagram shown in **Figure 1** for a brief representation of the initially proposed model. Note that **Figure 1** was drawn to help explicitly introduce the research hypotheses of the dissertation, it is not technically considered the full initial "SEM model" which we will see shortly. To clarify, the SEM technique was performed on the detailed and full SEM model that is demonstrated in **Figure 5**. In **Figure 5**, the latent variables are enclosed by ellipses while the observed variables are placed in rectangles; the arrows connecting them represent my initial specifications of which latent variables might be indirectly inferred from which set(s) of observed variables. The bold arrows connecting the latent variables themselves constitute what is called "structural model", which will be tackled later in this session. A special case here worth mentioning is that while being observed variables, age and gender were considered more like latent variables – there was no further "latent" variable to be inferred from each of them, and I was only interested in how they affected other latent variables. But essentially, they were still observed variables and would be included later in the analysis of structural model. Therefore, they were not included here in the analysis for establishing the measurement model. Once the theoretical research model has been defined, the first step, like mentioned earlier, is to establish the measurement model. The way how the latent variables were initially specified and the observed variables were collected and coded is illustrated in **Table 1**. In addition, summary statistics (mean and standard deviation) of the observed variables are provided in the table.

Factor analysis is often employed to help establish the measurement model, it attempts to determine which sets of observed variables share common variance-covariance characteristics that define theoretical constructs or equivalently, latent variables (Schumacker & Lomax, 2010a). Given the fact that the relationships between the observed variables and latent variables in this study were already pre-defined by me at the beginning, what was left was to “confirm” the proposed observed-latent-variable relationships. Accordingly, confirmatory factor analysis (CFA) was leveraged, with my aim to statistically testing whether or not the sample data collected from the participants confirmed my proposed relationships (Schumacker & Lomax, 2010a). In addition to testing how the model fitted the sample data (by examining a series of model fit indexes such as χ^2/df, the comparative fit index and the Tucker-Lewis index), reliability and validity was also tested for the measurement model. Reliability is concerned with the ability a measure to be consistent and often referred to as internal consistency (Schumacker & Lomax, 2010b). Internal consistency can be viewed as a measure of the extent to which a group of variables (scores) are related to each other to measure the same thing and is often assessed via Cronbach’s alpha and composite reliability (Hassan & Abdel-Aty, 2011; Zhang et al., 2019). Validity, meanwhile, is concerned with how well the group of variables (scores) indicates what they purport to measure, and can be inferred from convergent validity, discriminant validity, and so on (Schumacker & Lomax, 2010b). Convergent validity is often assessed by the Average Variance Extracted (AVE) index (Zhang et al., 2019). After the measurement model is assessed and established, the next step is to form the structural model, which revolves around the relationships between the latent variables. Eventually, the whole model will be tested again as to how it fits the sample data.

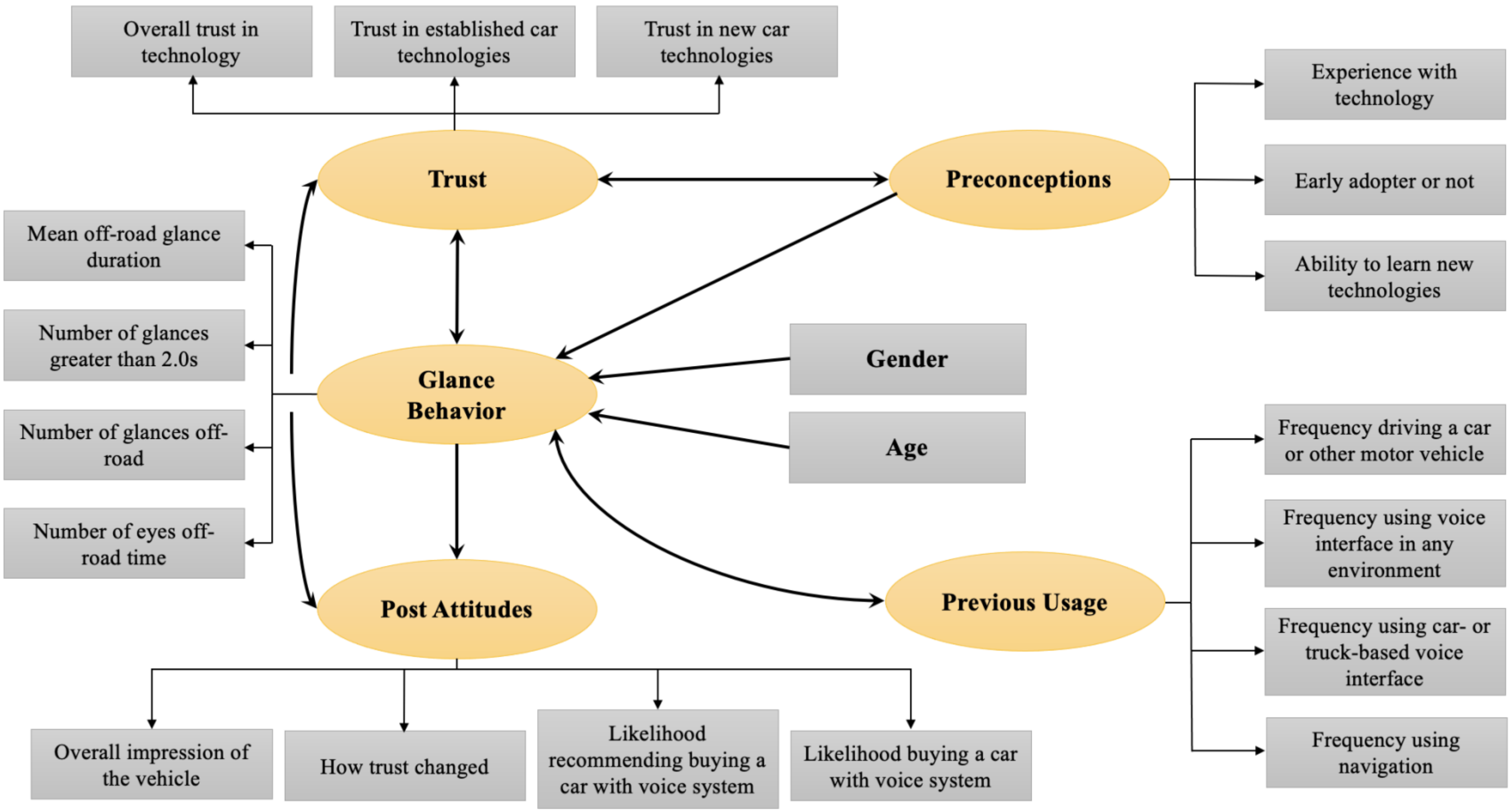

Figure 5. The initial SEM model

Table 1. Variable definitions, collection methods, scales, and statistics

Latent Variables	Items	Observed Variables	Collection Methods	Scale	Mean	Standard Deviation
Preconceptions	P1	Whether or not one saw himself as an early adopter of new technologies	Pre-experiment Survey	1-10 rating scale	7.08	2.08
	P2	Self-reported level of experience with technology	Pre-experiment Survey	1-10 rating scale	8.26	1.44
	P3	Self-reported ability to learn to operate new technologies	Pre-experiment Survey	1-10 rating scale	8.40	1.43
Trust	T1	Self reported overall level of trust in technology	Pre-experiment Survey	1-10 rating scale	7.92	1.39
	T2	Self reported overall level of trust in established technology	Pre-experiment Survey	1-10 rating scale	8.84	1.05
	T3	Self-reported level of trust in new technologies that are being introduced to cars	Pre-experiment Survey	1-10 rating scale	7.58	1.45
Previous Usage	PU1	Frequency of using a voice command interface in any environment	Pre-experiment Survey	1-10 rating scale	3.94	1.55
	PU2	Frequency of using a car- or truck-based voice command interface system	Pre-experiment Survey	1-10 rating scale	5.11	1.40
	PU3	Frequency of using an electronic navigation system in a car or truck	Pre-experiment Survey	1-10 rating scale	3.29	1.18
	PU4	Frequency of driving a car or other motor vehicle	Pre-experiment Survey	1-5 Ordinal	1.24	0.44
Glance Behavior	GB1	Mean single off-road glance duration (s)	In-vehicle Cameras	Numeric	0.74	0.15
	GB2	Number of glances off-road	In-vehicle Cameras	Numeric	24.22	14.13
	GB3	Total eyes off-road time (TEORT) (s)	In-vehicle Cameras	Numeric	18.06	10.80
	GB4	Number of glances greater than 2.0 seconds	In-vehicle Cameras	Numeric	0.29	0.69
Post Attitudes	PA1	The likelihood that participants would recommend buying a car with voice control system	Post-experiment Survey	1-10 rating scale	7.35	2.51

	PA2	How likely one would consider buying a car with in-vehicle voice interface for the next car	Post-experiment Survey	1-10 rating scale	8.31	2.21
	PA3	How the experience with the in-vehicle voice interface changed one's trust in new technologies that are being introduced into cars	Post-experiment Survey	1-10 rating scale	3.64	0.82
	PA4	Participants' overall impression of the vehicle they drove during the experiment	Post-experiment Survey	1-10 rating scale	7.77	2.00
Gender[a]	/	Gender	Pre-experiment Survey	Categorical	/	/
Age[a]	/	Age	Pre-experiment Survey	Numeric	39.77	15.87

[a] Gender and age were measured directly, they were represented exactly as how they were measured. They were classified as "latent variables" for SEM analysis purpose.

CHAPTER 4

RESULTS

4.1. Measurement Model and Assessment

The CFA results and model fit of the measurement model is presented in **Table 2** and **Table 3**. Overall, the results did not indicate an excellent fit of the proposed measurement model to the sample data. Specifically, as listed in **Table 2**, two out of six of the commonly used model fit indexes failed to meet the recommended criteria to be considered “excellent/good fit” (Schreiber, Nora, Stage, Barlow, & King, 2006; Schumacker & Lomax, 2010e). The ones that satisfied the recommended value/threshold were the Chi-square ratio, the comparative fit index (CFI), the Tucker-Lewis index (TLI), and the goodness-of-fit index (GFI). However, these four indexes might not be enough to determine whether a certain model is a good fit to the given data. Researchers have been assessing model fit by simultaneously taking into account the Chi-square value, root mean square error of approximation (RMSEA), the comparative fit index (CFI), the Tucker-Lewis index (TLI), standardized root mean square residual (SRMR), Akaike Information Criterion (AIC), and so on. It is recommended that if the vast majority of the chosen indexes indicate a good fit, then there is probably a good fit (Schreiber et al., 2006).

While **Table 2** shows how the measurement model fitted the data, **Table 3** tells us how well the latent variables were represented by the chosen observed variables. It is generally believed by a school of researchers that to be deemed as “good” internal consistency, Cronbach’s alpha should be greater than 0.7 (Sheykhfard & Haghighi, 2020;

Taber, 2018; Zhang et al., 2019). Toward this end, the latent variable Trust (0.699), Previous Usage (0.626), and Glance Behavior (0.644) might not be seen as with good internal consistency among their posited observed variables. However, one highly cited literature has reviewed comprehensive research work and concluded that Cronbach's alpha is actually reported in various ways and that many studies report Cronbach's alpha over 0.6 as either "acceptable", "sufficient", or "satisfactory" (Taber, 2018). At the same time, any value over 0.7 has been classified more as "good" and "(fairly) high" (Taber, 2018). Therefore, it shall be safe to state that all the latent variables in the initially proposed measurement model possessed, if not strictly "good", "acceptable" or "satisfactory" internal consistency.

Table 2. Model fit indexes for the initially proposed measurement model

Fit Index	Value	Recommended Value	Is the Model a Good Fit?
χ^2/df	1.953	≤ 2 or 3	Yes
CFI	0.963	≥ 0.95	Yes
TLI	0.954	≥ 0.95	Yes
GFI	0.979	> 0.90	Yes
RMSEA	0.082	< 0.05	No
SRMR	0.104	< 0.08	No

What may raise concerns lies in the validity of the measurement model. From **Table 3**, the average variance extracted (AVE) of Previous Usage was only slightly over 0.4 (0.405) while that of Glance Behavior was 0.392. To guarantee convergent validity, some researchers have recommended that the AVE be greater than 0.5 to be considered as adequate (Fornell & Larcker, 1981; Zhang et al., 2019). But according to Fornell & Larcker (1981), AVE itself is actually a more conservative measure than composite reliability, and researchers may conclude whether or not the convergent validity is adequate for a latent

Table 3. CFA results, internal consistency and convergent validity of the initially proposed measurement model

Latent Variable	Item	Estimate	Factor Loading	Cronbach's Alpha	Average Variance Extracted (AVE)	Composite Reliability
Preconceptions	Prec1	1	0.863	0.835	0.700	0.875
	Prec2	0.943	0.814			
	Prec3	0.965	0.833			
Trust	T1	1	0.889	0.699	0.506	0.743
	T2	0.514	0.457			
	T3	0.811	0.721			
Previous Usage	PU1	1	0.921	0.626	0.405	0.641
	PU2	0.812	0.748			
	PU3	0.495	0.456			
	PU4	-0.063	-0.058			
Glance Behavior	GB1	1	0.703	0.644	0.392	0.688
	GB2	32.643	0.245			
	GB3	49.318	0.484			
	GB4	5.717	0.884			
Post Attitudes	PA1	1	0.934	0.750	0.553	0.826
	PA2	0.854	0.798			
	PA3	0.574	0.537			
	PA4	0.690	0.645			

variable on the basis of composite reliability alone. Since the composite reliability of both Previous Usage (0.641) and Glance Behavior (0.88) were above the acceptance level (0.6), I decided to conclude that all the latent variables specified in the original measurement model were with an acceptable internal consistency and convergent validity (Lam, 2012).

Factor loadings listed in **Table 3** were the standardized parameter estimates, and the factor loading of an observed variable basically quantifies the extent to which the variable was related to the latent variable (Bandalos, 2018). Note that all the observed variables had a significant factor loading on their posited latent variables besides PU4, suggesting that participants' frequency of driving a car or other motor vehicle was not significantly related to the construct Previous Usage. Besides, although statistically significant, the factor loading of GB3 was relatively small (0.245). This indicates that the variance explained by GB3 (average number of glances off-road) on Glance Behavior was relatively small.

Altogether, the results discussed above indicate that: 1) the proposed measurement model might not be a good fit to the collected data; 2) Previous Usage might not be well represented by the initially chosen observed variables. Accordingly, measurement model revision is desired, which is usually done prior to building and assessing the structural model (Schumacker & Lomax, 2010b). But in the present work, I still chose to first completely present the SEM results of the initial theoretical model even though I had realized that the measurement model would not be a good fit, aiming to offer a full picture of how the postulated model worked. Model revision and assessment will be covered right after in section 4.3.

4.2. Structural Model

The model fit indexes for the initially proposed structural model are presented in **Table 4**, and they essentially represent how the theoretical model shown in **Figure 5** fitted the sample data. Again, four indexes failed to meet the recommended criterion and the measure χ^2/df exceeded the lower threshold 2 (2.507), although it was still less than 3. The results indicated that the structural model, or the initially proposed full theoretical model did not fit the sample data well. The coefficients and their associated statistical significance results among the latent variables are demonstrated in **Figure 6**. Solid lines represent significant relationships while dotted lines are used for non-significant relationships. These helped statistically quantify the depicted relationships and test my initial hypotheses regarding the original model.

Table 4. Model fit indexes for the initially proposed structural model

Fit Index	Value	Recommended Value	Satisfied?
χ^2/df	2.507	≤ 2 or 3	Yes for 3
CFI	0.916	≥ 0.95	No
TLI	0.920	≥ 0.95	No
GFI	0.923	> 0.90	Yes
RMSEA	0.103	< 0.05	No
SRMR	0.118	< 0.08	No

Three out of the seven hypotheses were supported, they were H2, H5, and H7, the results of hypothesis testing are presented in **Table 5**. Specifically, it was found that participants of different demographics possessed different glance behavior under the Navigation task: as participants' age increased, their off-road glance behavior tended to be longer (β = 0.523, p = 0.000); as compared to female participants, males appeared to have

longer off-road glance behavior (β = 0.230, p = 0.015). H5 was favored because participants who had more positive Preconceptions toward technologies were shown to

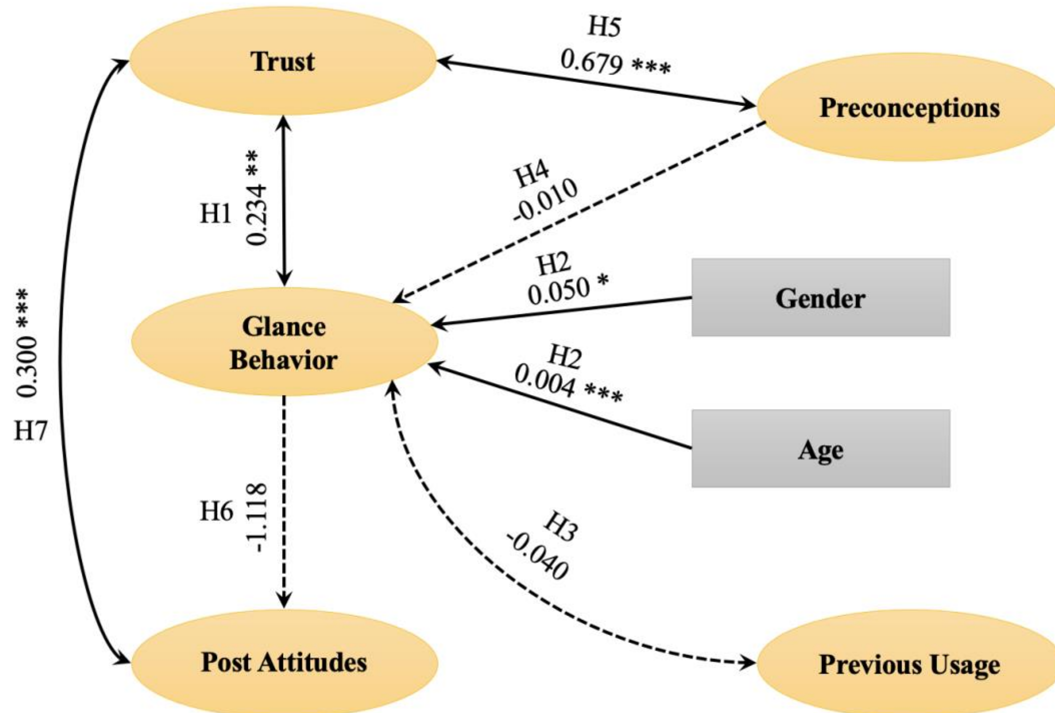

Figure 6. A simplified schematic showing the results of the structural model. Dotted lines indicate non-significant relationships. Note: *** p < 0.001; ** p < 0.01; * p < 0.001.

have higher trust in (vehicle) technologies, and vice versa (*r* = 0.679, p = 0.000). With respect to H7, the results suggested that participants with higher trust in (vehicle) technologies tended to have more positive post attitudes toward the vehicle and the voice interface that they had experienced (*r* = 0.300, p = 0.000). On the other hand, however, H1, H3, H4 and H6 were not supported by the results. The original H1 stated that there would be a negative correlation between Trust and Glance Behavior whereas the correlation coefficient indicated otherwise, although it was statistically significant (*r* = 0.234, p = 0.009). H3 proposed that there would be another negative relationship between

Table 5. Results of hypothesis testing

Hypothesis Number	Description	Statistical Test Result	Supported?
H1	There is a relationship between driver's trust in technology and glance behavior such that when one has higher trust in technology, she/he has less off-road glance behavior, and vice versa.	$r = 0.234$, $p = 0.009$	No
H2	Effect of individual characteristics on glance behavior: drivers of different individual characteristics (i.e., age and gender) will have different glance behavior.	$\beta = 0.523$, $p = 0.000$ (age) $\beta = 0.230$, $p = 0.015$ (gender)	Yes
H3	There is a relationship between driver's "previous usage" and glance behavior such that when one has more frequent "previous usage", she/he has less off-road glance behavior, and vice versa.	$r = -0.040$, $p = 0.662$	No
H4	Effect of preconceptions on glance behavior: drivers with more positive preconceptions will have less off-road glance behavior.	$r = -0.010$, $p = 0.221$	No
H5	There is a relationship between driver's preconceptions of technology and trust in technology such that when one has more positive preconceptions about technology, she/he has higher trust in technology, and vice versa.	$r = 0.679$, $p = 0.000$	Yes
H6	Effect of glance behavior on post attitudes: driver's glance behavior affects her/his post attitudes toward the voice interface, those who have less off-road glance behavior will have more positive post attitudes.	$\beta = -1.118$, $p = 0.075$	No
H7	There is a relationship between one's trust in technology and her/his post attitudes toward the voice system such that higher trust in technology will be associated with more positive post attitudes, and vice versa.	$r = 0.300$, $p = 0.000$	Yes

participants’ Previous Usage and Glance Behavior, while the coefficient suggested so, it was not statistically significant. As to H4, although the coefficient (-0.042) indicated the same effect of Preconceptions on Glance Behavior, it was not significant. H6 stated that Glance Behavior would have a negative effect on Post Attitudes, and again, the coefficient was not with statistical significance.

So far, the measurement and structural model of the initial theoretical research model have been assessed, and three key findings can be summarized: 1) the proposed model (**Figure 5**) connecting participants’ glance behavior while completing the Navigation task using an in-vehicle voice interface and other factors/variables was not a good fit to the data that were collected from the experiments, and this might be attributed to the facts that the latent variable Previous Usage was not adequately represented by the chosen observed variables, and that some of the relationships depicted in the structural model were not properly specified at the beginning; 2) only three of the seven original hypotheses were supported, although the results were obtained from a potentially subpar model; 3) the initially proposed research model, theoretically, might have failed to reflect enough knowledge regarding how participants’ glance behavior was affected by other variables as only age, gender and participants’ preconceptions about technologies were hypothesized to have a direct effect on their off-road glance behavior. These findings are strong indications that model revision for both the measurement model and the structural model is necessary if a better fit is to be achieved.

4.3. Model Revision

4.3.1. Measurement Model

In this section, I will talk about model modification and how a better-fit model was obtained, as well as any new results. It has been common practice to modify an unsatisfied model by eliminating parameters that are not significant, and adding parameters that improve the fit (Hox & Timo, 1998; Schumacker & Lomax, 2010c). For the initial measurement model, PU4 was already shown to not have a significant factor loading on its posited latent variable Previous Usage and was therefore removed in this round of revision. Also, it should be noted that no residual covariance was initially specified among the observed variables in the measurement model (**Figure 5**), this type of variance is often used to help explain the variance of a given observed variable that is not captured by its corresponding latent variable but rather by another observed variable. In the revised model, several such covariances were added. **Figure 7** helps illustrate how the revised measurement model was specified. The curved two-headed arrows connecting the small circles with text starting with "e" represent the residual covariance between the observed variables. Basically, these circles were the error term of their corresponding observed variables, and always existed in the initial and this revised model, although they were not explicitly drawn in **Figure 5** (since no residual covariance was specified and hence of no interest in the initial model). The revised measurement model possessed the model fit indexes as illustrated in **Table 6**. Compared to the model fit indexes of the initial model (**Table 2**), all the fit index values of the revised model have been improved, and five of the

six indexes now satisfied the recommended value/threshold to be considered as good fit. The value of RMSEA was slightly greater than the recommended 0.05 (0.052), and a school

Table 6. Model fit indexes of the revised measurement model

Fit Index	Value	Recommended Value	Satisfied?
χ^2/df	1.384	≤ 2 or 3	Yes
CFI	0.988	≥ 0.95	Yes
TLI	0.984	≥ 0.95	Yes
GFI	0.988	> 0.90	Yes
RMSEA	0.052	≤ 0.05	No (fair)
SRMR	0.068	< 0.08	Yes

of researchers have suggested that the values between 0.05 and 0.08 indicate fair fit and that values between 0.09 and 0.1 indicate mediocre fit (MacCallum, Browne, & Sugawara, 1996). Therefore, it shall be safe to conclude that the revised measurement model was a good fit to the data. The results presented in **Table 7** shows that the latent variables were still with acceptable or satisfactory internal consistency and convergent validity for this revised measurement model (Fornell & Larcker, 1981; Lam, 2012).

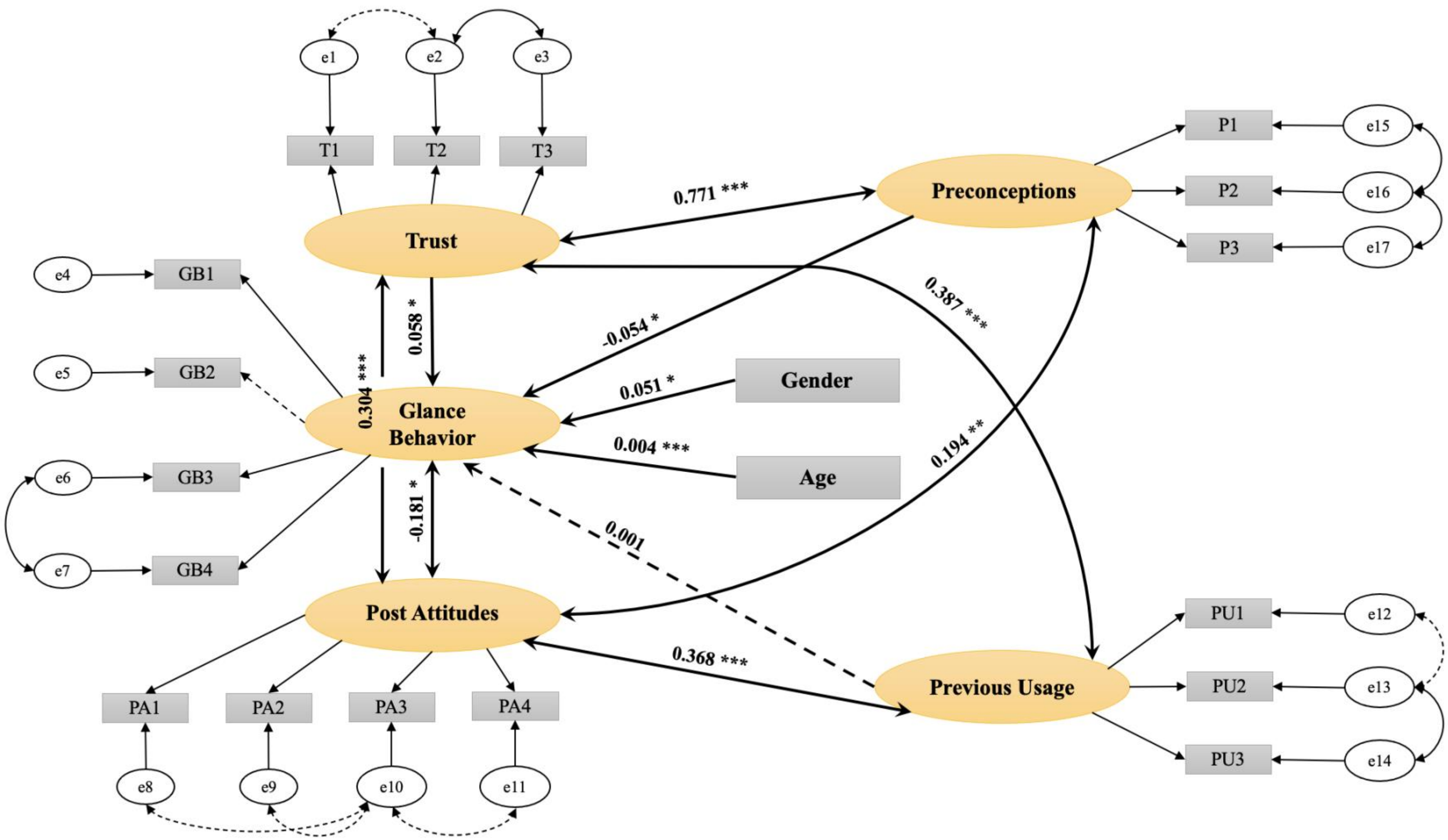

Figure 7. The revised SEM model. Only coefficients of the structural model were demonstrated. Dotted lines indicate non-significant relationships. Note: *** p < 0.001; ** p < 0.01; * p < 0.001.

Table 7. CFA results, internal consistency and convergent validity of the revised measurement model

Latent Variable	Item	Estimate	Factor Loading	Cronbach's Alpha	Average Variance Extracted (AVE)	Composite Reliability
Preconceptions	Prec1	1	0.873	0.835	0.548	0.779
	Prec2	0.643	0.561			
	Prec3	0.863	0.754			
Trust	T1	1	0.897	0.699	0.460	0.693
	T2	0.379	0.340			
	T3	0.755	0.677			
Previous Usage	PU1	1	0.858	0.701	0.415	0.659
	PU2	0.695	0.597			
	PU3	0.453	0.389			
Glance Behavior	GB1	1	0.682	0.644	0.391	0.673
	GB2	17.200	0.125			
	GB3	42.180	0.402			
	GB4	6.541	0.982			
Post Attitudes	PA1	1	0.913	0.750	0.569	0.836
	PA2	0.939	0.857			
	PA3	0.644	0.588			
	PA4	0.661	0.603			

4.3.2. Structural Model

Figure 7 demonstrates the full SEM model for the revised structural model. Relative to the initial structural model (**Figure 6**), the revised structural model incorporated more relationships and contained several modified relationships. For example, Trust and Previous Usage were specified to have a direct effect on participants' off-road glance behavior. The model fit indexes of this model prove that it was a good fit to the experimental data as five of the six indexes met the recommended criteria while RMSEA was only 0.001 (0.051) above the recommended value (**Table 8**) (MacCallum et al., 1996). With this revised model, Glance Behavior was the only predicted latent variable, and 41.8% of its variance was explained.

Table 8. Model fit indexes of the revised structural model

Fit Index	Value	Recommended Value	Satisfied?
χ^2/df	1.371	≤ 2 or 3	Yes
CFI	0.983	≥ 0.95	Yes
TLI	0.982	≥ 0.95	Yes
GFI	0.960	> 0.90	Yes
RMSEA	0.051	< 0.05	No (fair)
SRMR	0.063	< 0.08	Yes

This revised structural model suggests that with the experimental data, drivers' off-road glance behavior was significantly affected by trust in (vehicle) technologies, age, gender, and preconceptions about technologies. Specifically, as compared to female drivers, male drivers appeared to be more likely to have more off-road glance behavior (β = 0.051, p = 0.013), and older drivers too tended to possess more off-road glance behavior (β = 0.004, p = 0.000). Trust in (vehicle) technologies positively affected drivers' glance

behavior in such a way that when drivers had higher trust, their off-road glance behavior increased (β = 0.058, p = 0.016). As for preconceptions about technologies, when drivers had more positive such preconceptions, they were more likely to have less off-road glance behavior (β = -0.054, p = 0.031). Previous usage did not have a significant effect on drivers' off-road glance behavior while the regression coefficient was positive (β = 0.001).

Other correlations are also observed from the revised structural model. Drivers' off-road glance behavior was negatively correlated to their post attitudes toward the vehicle and voice interface (r = -0.181, p = 0.041). Drivers' trust in (vehicle) technologies was positively correlated to their preconceptions about technologies (r = 0.771, p = 0.000), previous usage (r = 0.387, p = 0.000), and post attitudes toward the vehicle and voice interface (r = 0.304, p = 0.000). Moreover, drivers' off-road glance behavior was positively correlated with their previous usage (r = 0.368, p = 0.000) and preconceptions about technologies (r = 0.194, p = 0.010).

It should be noted that the revised model is just a variation and expansion of the initial model. While the revised model was shown to adequately fit the experimental data, it may not be the only model that would fit the data. With this better fitting model, several observations were obtained regarding my original research hypotheses. A direct effect of trust on glance behavior was found while my original hypothesis was about correlation between these two factors (H1), which was unidirectional. The effect of trust on glance behavior was positive: participants who had higher trust in (vehicle) technologies tended to possess longer off-road glance behavior when interacting with the voice-based interface. H2 was still supported: participants with different age and gender showed significantly

different patterns of glance behavior. Unlike the original H3 which proposed a correlation between Previous Usage and Glance Behavior, a direct effect of Previous Usage on Glance Behavior was specified and tested positive, although non-significant. H4 also held in the revised model – Preconceptions exerted significantly negative impact on Glance Behavior. When participants had more positive such preconceptions about technologies, they appeared to have less off-road glance behavior when interacting with the in-vehicle voice interface. A positive correlation between Preconceptions and Trust was discovered, which favored the original H5. When one had more positive preconceptions about technologies, she/he had higher trust in (car) technologies, and vice versa. Initially, H6 hypothesized that there would be a negative effect of Glance Behavior on Post Attitudes, whereas in the revised model a significantly negative correlation between the two was found. This suggested that when participants had less off-road glance behavior during the interaction with the voice-based interface, their post attitudes toward the vehicle and interface (e.g., willingness to purchase a vehicle with voice-based interface) tended to be more positive, and vice versa. The original H7 was also supported – when participants had higher trust in (car) technologies, their post attitudes toward the vehicle and voice interface appeared to be more positive, and vice versa.

CHAPTER 5

DISCUSSION & CONCLUSION

5.1. Discussion

With in-vehicle technologies being introduced to modern automobiles, drivers are now able to communicate with the vehicle for entertainment or other needs in a more convenient and supposedly, safer way, due to the belief and evidence suggesting that as compared to the traditional visual-manual manipulations, interacting with voice-based interfaces shall require less visual attention from the driver (Peissner et al., 2011; Reimer et al., 2014; Simmons et al., 2017). Driving has always been believed to be complex, rely heavily on visual information, and closely related to multiple factors such as individual characteristics (e.g., age and gender), trust in and experience with (vehicle) technologies (Sivak, 1998). Interacting with in-vehicle voice interfaces is not an exception from causing drivers additional distraction, including visual distraction, and driver distraction has been stressed to be one of the leading causes of crashes (Dingus et al., 2006). It is therefore worth investigating how commonly studied factors might affect drivers' visual attention while interacting with in-vehicle voice interfaces, which to the best of my knowledge, was rarely tackled by previous research. Given the complicated nature of driving and drivers' behavior, structural equation modeling (SEM) appears to be able to address the complexity (X. Zhao, Xu, Ma, Li, & Chen, 2019). This dissertation work built upon a large-scale study and utilized SEM to conduct a secondary data analysis, aiming to identify factors affecting drivers' allocation of visual attention while interacting with an in-vehicle voice interface and uncover the complex relationships between it and other factors.

With the best fitting model, the results confirmed that drivers' behavior, specifically off-road glance behavior when interacting with an in-vehicle voice interface, is complex because it was associated with multiple factors. The results regarding how participants' trust in (vehicle) technologies affected their glance behavior while interacting with in-vehicle voice interfaces was generally in line with past research stating that trust in technology may be associated with subsequent driver behavior when engaging with an in-vehicle interface (Geitner et al., 2017). But with regard to specifically how drivers' glance behavior was affected by trust, the positive effect seemed to contradict previous research that stated that higher trust in new car technologies was associated with fewer long duration off-road glances (Geitner et al., 2017).

In the present work, the observed variable number of glances greater than 2.0 seconds was the one influenced mostly and positively by the latent variable Glance Behavior (factor loading 0.982), and the results indicated that higher trust was associated with more (long duration) off-road glance behavior. The results suggest that those with higher trust directed their eye glances more often off the road when performing the navigation task using the voice-based interface. This might seem counterintuitive as one might think when the driver has higher trust in vehicle technologies, she/he will have stronger faith that the technology, or in this work the voice interface, is able to help achieve the task goal and as a result, the driver will not need to pay significantly more visual attention to anywhere else than the road because the voice interface should capture the command and function well. In fact, trust is believed to be the reflection of one's expectation that the trustee will perform an important action as expected or can be relied upon, and the trustor is willing to be vulnerable to the action (J. D. Lee & See, 2004; Mayer

et al., 1995). Take, driver's gaze behavior and trust in automation as an example: it was found that when drivers reported higher automation trust, they monitored the automation system less frequently while driving (Hergeth et al., 2016). Therefore, the relationship between trust and visual attention when the driver is interacting with an in-vehicle interface found in this dissertation may not strongly support either the theoretical bases of trust or some past findings (Geitner et al., 2017; J. D. Lee & See, 2004).

One possible explanation for this phenomenon might be that the latent variable Trust, although proved to possess fair internal consistency and convergent validity, did not directly assess drivers' particular trust in in-vehicle "voice interfaces". Instead, it was constituted by questions asking drivers' general trust in technologies and in-vehicle technologies. As a result, trust utilized for analysis might have failed to build the clear connection between the trustee (participants) and the actual trustor (in-vehicle voice interfaces) and could probably explain why the relationship obtained here was to some degree contrary to the theoretical bases of trust (J. D. Lee & See, 2004) because the trustor targeted for the survey questions was about general technologies while in the field experiment, the trustor that presented was the voice-based interface. The findings are not ungrounded, they are just not favored by the theory or the very few relevant work. Again, in this specific field of how trust affects glance behavior under the interaction with in-vehicle voice interfaces, knowledge and scientific evidence is limited. Whether or not the findings obtained here can safely be generalized remains questionable and thus, examination of additional datasets seems desired (Geitner et al., 2017). Nonetheless, the effect of trust upon off-road glance behavior discovered in this work corroborates the belief that one's trust in a technology naturally influences how she/he interacts with it (Geitner et

al., 2017; J. D. Lee & See, 2004). Moreover, the finding fills the gap and adds to the existing research database a novel piece of knowledge regarding the specific role trust plays in drivers' off-road glance behavior while interacting with voice-based systems.

The effect of age and gender on drivers' glance behavior discovered in this work is generally in line with relevant research that showed that when completing certain tasks using a voice-command interface while driving, older drivers tended to have higher TEORT as compared to younger drivers and that male drivers showed higher TEORT and mean glance duration off the road relative to female drivers (B. Mehler et al., 2014). Findings from a body of literature stating that male drivers often drive less safely than females (Swedler, Bowman, & Baker, 2012) might account for why males in the present work tended to have longer off-road glance behavior. But there also exists relevant work that showed that female drivers had lower driving safety levels as compared to male drivers (X. Zhao et al., 2019). Since off-road glances greater than 2.0 seconds is considered safety-critical (Geitner et al., 2017; National Highway Traffic Safety Administration, 2013) and was one observed variable in this work mostly and positively influenced by the latent variable Glance Behavior, it shall be safe and appropriate to interpret the gender and age effect here as that male drivers and older drivers were driving riskier than female drivers and younger drivers when interacting with an in-vehicle voice interface. This might be considered in support of the school of research concluding that male drivers were less safer than female drivers (Swedler et al., 2012). Also, off-road glances might suggest that the interface was frequently checked for visual feedback by the driver (McWilliams et al., 2015; Reimer et al., 2014), which could be an explanation for the longer off-road glance behavior that older drivers possessed because their ability of multitasking or interacting with an in-

vehicle interface while under the high demanding driving task may not be as good as young drivers.

Unlike participants' preconceptions about technologies (whether one saw her/himself as an early adopter of technologies and one's ability to learn to operate new technologies), participants' previous usage of voice interfaces and navigations did not significantly affect their glance behavior. Considering that when the original experiments were conducted, all the vehicles and voice-based interfaces employed were quite new on the market, the chances that participants had previously used the same vehicles and interfaces were rare. Therefore, participants' interactions with the interfaces might also be seen as with new technologies. It is then not hard to explain why participants with more positive preconceptions (e.g., self-identified as an early adopter of new technologies and have strong ability to learn to operate them) appeared to have less off-road glance behavior when using the voice-based interface to perform certain tasks – they might have more knowledge in new technologies and learn to master them faster. The findings also suggest that even though one used a certain technology frequently in the past, her/his performance on operating a similar but new technology might depend on whether this person is "tech-savvy". This is in line with past literature that stated that when evaluating the demands that real drivers may experience when interacting with a certain system, it is critical to consider the systems that they have already been familiar with and those that they are encountering for the first time independently (B. Mehler et al., 2014). The correlations involving Post Attitudes are suggesting that it was significantly and positively affected by ones' trust in and preconceptions about (vehicle) technologies, as well as previous usage of voice-based interfaces and navigations. This is consistent with previous findings (C. Lee et al., 2015).

For a better deployment of vehicles with voice-based interfaces, one should cultivate potential costumers' trust in voice-based technologies, demonstrate the functionality, arrange as many as possible pre-interactions/trainings with voice-based interfaces with the customers by dealerships, and advertise the feature on a wider range of platforms so that the public would acquire more information about such technology.

The model that I initially proposed (**Figure 5**) was proved to not fit the experimental data well, indicating that the sample (experimental data) covariance matrix and the proposed model implied covariance were not similar (Schumacker & Lomax, 2010e). This was mainly because the latent variable Previous Usage was not adequately specified, the residual covariances among the observed variables were missing, and the relationships that I originally depicted for the structural model were not properly established. It might be learned from the results that PU4, which basically asked participants about their frequency of driving a car or other motor vehicle did not significantly contribute to the latent variable Previous Usage as the other three observed variables did, which asked participants how often they used a voice command interface in any environment, the frequency that they used a car- or truck-based voice command interface, and how often they used an electronic navigation system in a car or truck, respectively. While this might seem naturally follow because PU4 was asked regarding one's car or motor vehicle usage, it shall be safer for me to revise the wording – to further specify "Previous Usage" as "Previous Usage of In-vehicle Voice and Navigation System".

For human factors researchers doing data analysis, their aim is to evaluate the relationships between a number of variables (Wickens, Lee, Liu, & Gordon, 2004b). In the

field of technology adoption and behavioral research, certain variables are frequently measured via surveys and questionnaires, just as what has been shown in this dissertation work. Relatedly, one major concern is their validity (Wickens et al., 2004b). In this work, I grouped the survey questions to various latent variables on the grounds of what has been studied previously along with intuition. It turned out that they eventually all possessed a fair (if not excellent or good) convergent validity. However, they could have been a poor representation of the observed variables collected from the participants as the researchers who conducted the original experiments did not intentionally pre-group the surveys or questionnaires in a clear way that SEM would be the analytical method in future analysis, nor did the researchers borrow the questionnaires or surveys from past literature in which they had been clearly specified and tested to represent certain constructs. Since the present work was a secondary data analysis based on what had already been collected from a large-scale study, my philosophy in conducting this work might have not been perfectly consistent with what the experimenters had proposed. Nevertheless, the idea of representing multiple observed variables in a smaller number of unobserved/latent variables is a key part of SEM analysis. For researchers who plan to use SEM on their research, it's beneficial to use questions that have already been grouped and tested by past literature or to pilot test questions to ensure that the questions and variables that might be used have good internal consistency and validity before formally launching the study.

The original experiments were conducted in 2014 and 2015, it is therefore worth mentioning that with the development and refinement of in-vehicle voice interfaces over the past few years, the systems nowadays shall possess a better performance. That is, with drivers performing similar navigation tasks using a more recent voice interface, we might

expect a general trend of reduction in off-road visual behavior, and yet this still needs more scientific evidence to tell us how in-vehicle voice interfaces have evolved and whether they are now able to assist drivers in a safer way. Regarding participants' different off-road glance behavior while interacting with the voice interfaces, racial disparities might have been responsible – it might have taken more effort from both the interface and the driver to complete a certain navigation task due to different pronouncing habits and/or accents.

5.2. Limitations and Future Work

Limitations of this dissertation work mainly fall into the following aspects. The participants employed were geographically limited as they were mainly recruited from the greater Boston area. The apparatus utilized in each of the three studies was different, which could have affected how drivers interacted with the voice interface. For example, some interfaces may take longer to respond to the driver's command than others, the layout of the interfaces differed such that some might require participants to look away from the roadway to a greater extent and consequently, there might have existed unexplained variance in some of the variables/factors such as the latent variable Glance Behavior owing to the different vehicles and interfaces. This could have become a significant variable affecting drivers' off-road glance behavior while performing tasks (McWilliams et al., 2015). As introduced in Appendix B, when participants were performing the Navigation tasks via the voice-based interface, the exact procedure (e.g., voice command, the prompt delivered by the interface) differed as well, the effect of which might not be precluded. While the adoption of SEM helped concurrently address the complex relationships between drivers' glance behavior and other factors, interacting effects have not been captured. But

according to past research, age interacting with gender was able to impose significant impact on drivers' off-road glance behavior when using voice-command interfaces (B. Mehler et al., 2014). The latent variables specified were with relatively satisfactory internal consistency and convergent validity, which might to some degree have resulted in the variables being defined loosely and consequently, losing some accuracy. As an example, some researchers chose to also examine the values of factor loadings of individual observed variables for convergent validity, if they are statistically significant and exceed 0.6 for the posited latent variable, the convergent validity is then guaranteed (Zhang et al., 2019). Moreover, discriminant validity, which reflects the extent to which the constructs differ from one another, was not considered in the present work (Zhang et al., 2019). Lastly, participants' off-road glance behavior in this dissertation was characterized by the glances that were off the forward roadway while they were interacting with the voice interfaces, but whether the glances were directed toward the in-vehicle display that's associated with the voice command was unknown. In other words, it is not impossible that some of the off-road glance behavior was not due to the participants completing the tasks using the voice interfaces or seeking feedback for the tasks.

Per the findings obtained in this dissertation work, drivers' allocation of visual attention when using an in-vehicle voice interface to perform navigation tasks, as represented by various off-road glance measures, was complex as it was under the effect of multiple factors. Since driving relies substantially on visual information and attention, if purely treating more off-road glance behavior as a detriment to drivers' safety (it shall be safe to do so), the results then suggest that drivers of certain characteristics (e.g., male drivers and older drivers) were driving riskier than others while interacting with the in-

vehicle voice interface to perform certain tasks. The results also confirm that completing certain tasks using voice-based interfaces may not preclude the need to consider possible visual demand from the driver (Bruce Mehler et al., 2014). It is important that systems like in-vehicle voice interfaces divert drivers' attention from the primary driving task as little as possible (McWilliams et al., 2015). Therefore, future design of in-vehicle voice interfaces might consider incorporating additional feedback via audio when the driver is making voice commands to help keep her/his eyes on the road without frequently checking if the system understood what was told and is functioning as expected. For instance, as can be referred in Appendix B, the voice-based interface employed in study 1 would require participants' manual input to confirm the destination address as the last step when navigating while in study 2 and study 3, the voice-based interface would provide participants with auditory prompt and receive verbal confirmation for the choice of destination address. Although how these different modalities affected participants' off-road glance behavior when performing the tasks remains uncertain and might be hard to accurately examine from the given dataset and studies, it is worth exploring. At the same time, instruction or training regarding how to properly use in-vehicle voice interfaces could be offered at the dealership after customers purchased a vehicle with such a feature (C. Lee et al., 2015). Both recommendations are made with the purpose to mitigate any potentially negative effect on driving safety caused by difference in individual characteristics.

To promote the acceptance and adoption of this new modality of interaction with vehicle technologies, effort should be made to ensure that the potential costumers' trust in voice-based interfaces is well established and at a high level. Also, placing customers into situations where they can physically experience voice-based interfaces may broadly

enhance their attitudes toward them and willingness to purchase a vehicle with such features (C. Lee et al., 2015). Designers should consider incorporating individual characteristics such as age and gender into the design process such that the negative effect resulted from the difference in users' characteristics could be mitigated. Furthermore, designers may not ignore the effort that is required for one to interact with in-vehicle voice interfaces, they should try to reduce the complexity and increase the clarity of interaction between drivers and the interfaces (Zhang et al., 2019).

Future research is recommended to be directed in ways to answer these research questions: 1) how the same set of variables proposed here affects drivers' safety-critical-related off-road glance behavior while interacting with an in-vehicle voice interface; 2) whether or not the relationships uncovered here, especially how trust affects glance behavior, can be replicated with more data and scientific evidence and if not, whether or not asking trust questions specifically and directly regarding one's trust in in-vehicle "voice interfaces" would lead us to new findings; 3) when studying the effect of trust on drivers' off-road glance behavior while interacting with in-vehicle voice interfaces, whether their inherent visual behavior while not engaging in any secondary tasks also has an effect; 4) how different layouts of voice-based interfaces (e.g., content in auditory prompt, secondary display) affect driver's glance behavior while interacting with them and if certain designs could mediate the possible negative effect resulted from drivers' characteristics (e.g., if providing additional audio feedback to older drivers would shorten their off-road glance behavior while performing tasks using voice-based interfaces).

5.3. Contribution to Human Factors

The goal of human factors may be defined as making the human interaction with systems one that enhances performance, increases safety, and increases user satisfaction (Wickens, Lee, Liu, & Gordon, 2004a). With the rapid development of technologies and increasing demand from drivers for more connectivity and entertainment options while driving, developers and manufacturers in the automobile industry are making the driver-vehicle interaction more convenient, entertaining, and automated by providing various types of in-vehicle technologies/systems as well as automations (Reimer et al., 2014). But owing to the difference in the potential customers' individual-level factors, whether the new forms of driver-vehicle interaction that are being introduced will receive unanimous acceptance and work as projected might remain uncertain.

While the three goals of human factors illustrated at the beginning of this section are not mutually exclusive, they should be treated equally important and considered simultaneously; no one is superior to the others in the system design process. From the present work, it's clear that some users' safety was compromised due to the interaction with the system, some were less satisfied with the system they had experienced, and some took significantly more effort to perform the task using the system, and all the differences might be attributed to the studied individual-level factors. To reach an optimal balance in a system among performance enhancement, safety benefits, as well as user satisfaction, the human component is undoubtedly critical and should be included in the system design process. This dissertation again stresses the importance of accounting for the psychological and physical factors that are properties of the human component (Wickens et al., 2004a).

5.4. Conclusion

This dissertation work utilized structural equation modeling to explore factors affecting drivers' off-road glance behavior while interacting with an in-vehicle voice interface to perform navigation tasks and was conducted as a secondary data analysis based upon a large-scale study. Drivers' trust in and preconceptions about (vehicle) technologies, as well as gender and age were shown to be the most contributing and significant factors affecting their off-road glance behavior while interacting with an in-vehicle voice interface. The present work favors the general idea that driving is complex. To mitigate the potentially negative effect caused by drivers' characteristics on their visual attention while driving with interaction with voice-based interfaces and boost the deployment and acceptance of in-vehicle voice interfaces, actions and future research are desired in this manufacturer-dealership-driver loop. Given that the ultimate goal of the study of human factors is toward optimizing system design for humas, the present work again highlights the need to account for the factors that are properties of the human (Wickens et al., 2004a).

APPENDIX A

STUDY VEHICLES & INTERFACES

Appendix A introduces the driver vehicle interface (DVI) and the layout of the voice-based interface of the three production level vehicles employed in the studies. Again, since the original experiments were conducted by another group of researchers, the descriptions and graphical representations documented here might be borrowed from their preliminary reports (B. Mehler, Reimer, Dobres, & Coughlin, 2015; B. Mehler, Reimer, Dobres, McAnulty, et al., 2015; B. Mehler, Reimer, McAnulty, et al., 2015).

Study 1

Figure 8 shows the primary screen in the center console, and the display is when one is selecting a phone number to call from the phone contact. An additional small display in support of the main infotainment system is demonstrated in **Figure 9**, and it is located between the tachometer and speedometer. **Figure 10** shows the location of the button which participants would press to perform the voice-based tasks.

The vehicle was instrumented with a custom data acquisition system for time synchronized recording of data from:

- vehicle information via the controller area network (CAN) bus,
- a Garmin 18X Global Positioning system (GPS) unit,
- a MEDAC System/3™ physiological monitoring unit to provide EKG (for heart rate determination) and skin conductance level (SCL) signals,

- video cameras,
- a wide area microphone to capture driver speech and audio from the vehicle's speech system.

Figure 8. Primary screen of the center console, 2014 Chevrolet Impala

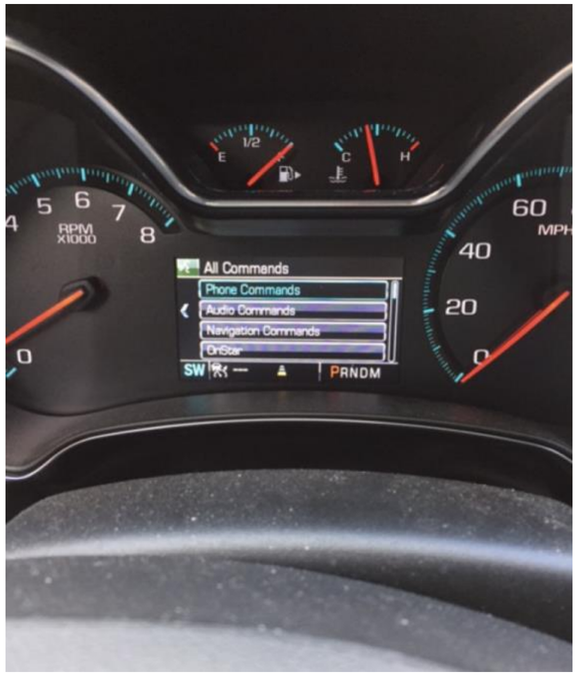

Figure 9. Additional small display in the instrument cluster, 2014 Chevrolet Impala

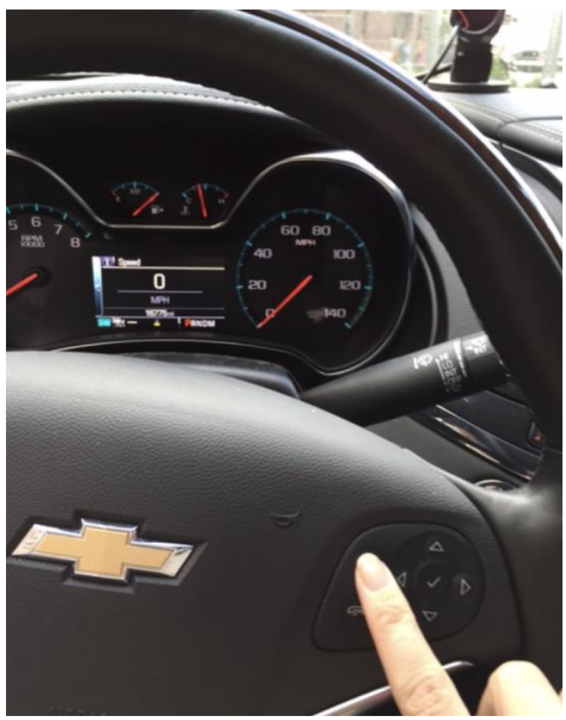

Figure 10. Push-to-talk button on the steering wheel, 2014 Chevrolet Impala

The five video cameras provided views intended to capture the driver's face for primary glance behavior analysis, the driver's interactions with the vehicle's steering wheel

and center console, the forward roadway (narrow and wide-angle images), and a rear roadway view. Data were captured at:

- 10 HZ for the CAN bus and GPS,
- 30 HZ for the face and narrow forward roadway cameras,
- 15 HZ for the remaining cameras,
- 250 HZ for the physiological signals to support EKG feature extraction for heartbeat interval detection.

Phone connectivity was supported by pairing a Samsung Galaxy S4 smartphone to the vehicle's embedded system via the vehicle's Bluetooth wireless interface.

Study 2

A 2014 Mercedes CLA was adopted for the experiment, likewise, there was a primary interface associated with the vehicle using the COMMAND infotainment system. The DVI is shown in **Figure 11** while a closer look at the display screen is shown in **Figure 12**. It should be noted that the screen was situated relatively high in the center console area. There were hard buttons in the mid-center console, and another rotational controller between the seats. A push-to-talk button was also present on the steering wheel. The rotational controller was not required for any of the voice-command involved tasks. There was also a small display screen in the center of the instrument cluster, but it was not actively used in conjunction with voice-command interactions.

The vehicle was instrumented with the same data acquisition system as was used in study 1 for synchronized recording of data. Five cameras were also utilized for the same

purposes and at the same capture frequency as in study 1. Lastly, phone connectivity was achieved by pairing a Samsung Galaxy S4 smartphone to the vehicle's embedded system via Bluetooth, which was again the same as in study 1.

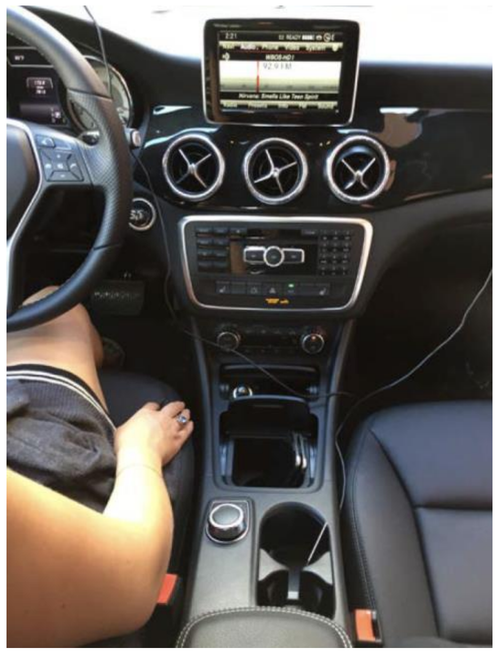

Figure 11. Layout of the DVI showing the main components of the COMMAND system

Figure 12. Close-up of the primary display screen

Study 3

A 2015 Toyota Corolla was employed for study 3, the infotainment interface was the standard production Entune Premium Audio with navigation. **Figure 13** shows the DVI of the vehicle while **Figure 14** provides a closer view of the display screen of the infotainment system. **Figure 15** gives the location of the push-to-talk button on the steering wheel, where the participant would press and engage in voice-based tasks.

Similarly, the vehicle was instrumented with the same data acquisition system as was used in study 1 & 2 for synchronized recording of data. Five cameras were also utilized for the same purposes and at the same capture frequency as in study 1 & 2. Phone connectivity was again achieved by pairing a Samsung Galaxy S4 smartphone to the vehicle's embedded system via Bluetooth, which was basically identical to study 1 & 2.

Figure 13. Layout of the DVI

Figure 14. Closer view of the display screen in the center cluster

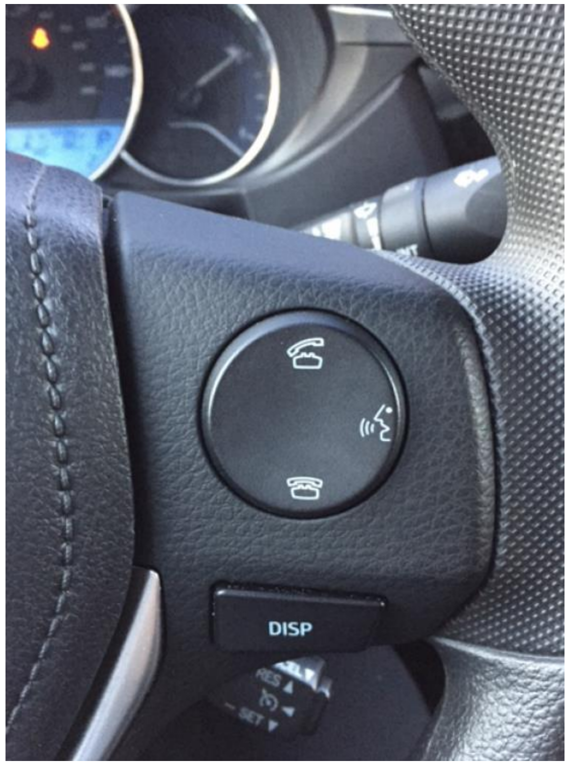

Figure 15. Location of the push-to-talk button on the steering wheel

APPENDIX B

NAVIGATION TASK DETAILS

This section aims to illustrate the detailed procedures of completing the Navigation tasks using the voice-based interface. Some facts and descriptions are based on or borrowed from the AgeLab researchers' preliminary reports (B. Mehler, Reimer, Dobres, & Coughlin, 2015; B. Mehler, Reimer, Dobres, McAnulty, et al., 2015; B. Mehler, Reimer, McAnulty, et al., 2015).

For each of the three studies, the Navigation tasks focused in this dissertation work involved entering three destination addresses via the voice-based interface:

- 177 Massachusetts Avenue, Cambridge,
- 293 Beacon Street, Boston,
- Participant's home address.

Since the system and interface differed between the three studies, the exact procedure of completing the tasks differed accordingly. In the following descriptions, the notation [XXX] represents a certain control button. Quotations with ("") are used to indicate what participants should speak, and the **bold** text indicates the auditory prompt delivered by the system. The followings are illustrations of how the first address was entered via the in-vehicle voice interface of the different study vehicles.

Study 1

- Touch [push-to-talk button]

- “Address” -> **Say the address in Massachusetts or say change state or change country** -> “177 Massachusetts Avenue, Cambridge” -> **177 Massachusetts Avenue, Cambridge, correct?** -> “Yes” -> **Complete your selection from the radio display.**
- {Optional support prompts were visible on the instrument cluster display (the additional small display) during initial steps of task, but driver was not required to look to at the display. Visual engagement was required when task shifts to again listing address options on the center console display screen. Confirmed address was shown as first option on list.}
- Touch [address on the touch screen]
- Touch [“Go” button on the touch screen]

Study 2

- Touch [push-to-talk button]
- “Enter Destination” -> **Please Say the house number, the street, and the town** -> “177 Massachusetts Avenue, Cambridge” -> **177 Massachusetts Avenue, Cambridge, accepted. Do you want to start route guidance?** -> “Yes” -> **Starting route guidance, the route is being calculated.**

Study 3

- Touch [push-to-talk button]
- **How may I help you?** -> “Address” -> **For direction to an address, please say the full address including the city and the state** -> “177 Massachusetts Avenue,

Cambridge, Massachusetts" -> **I heard 177 Massachusetts Avenue, Cambridge, Massachusetts, is that correct?** -> "Yes" -> **Starting guidance for a new route.**

BIBLIOGRAPHY

Bakhit, P. R., Osman, O. A., Guo, B., & Ishak, S. (2019). A distraction index for quantification of driver eye glance behavior: A study using SHRP2 NEST database. *Safety Science*, *119*, 106–111. https://doi.org/10.1016/j.ssci.2018.11.009

Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. Guilford Publications.

Buuren, S. V., & Groothuis-Oudshoorn, K. (2010). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 1–68.

Chiang, D. P., Brooks, A. M., & Weir, D. H. (2005, April 11). *Comparison of Visual-Manual and Voice Interaction With Contemporary Navigation System HMIs*. 2005-01–0433. SAE Transactions. https://doi.org/10.4271/2005-01-0433

Choudhary, P., & Velaga, N. R. (2017). Mobile phone use during driving: Effects on speed and effectiveness of driver compensatory behaviour. *Accident Analysis & Prevention*, *106*, 370–378. https://doi.org/10.1016/j.aap.2017.06.021

Crundall, D., Shenton, C., & Underwood, G. (2004). Eye Movements during Intentional Car following. *Perception*, *33*(8), 975–986. https://doi.org/10.1068/p5105

Cunningham, M. L., Regan, M. A., & Imberger, K. (2017). Understanding driver distraction associated with specific behavioural interactions with in-vehicle and portable technologies. *Journal of the Australasian College of Road Safety*, *28*(1), 27.

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., … Knipling, R. R. (2006). *The 100-car naturalistic driving study, Phase II-results of the 100-car field experiment (No. DOT-HS-810-593)*. United States: Department of Transportation. National Highway Traffic Safety Administration.

Donmez, B., Boyle, L. N., & Lee, J. D. (2006). The Impact of Distraction Mitigation Strategies on Driving Performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *48*(4), 785–804. https://doi.org/10.1518/001872006779166415

Fan, Y., Chen, J., Shirkey, G., John, R., Wu, S. R., Park, H., & Shao, C. (2016). Applications of structural equation modeling (SEM) in ecological studies: An updated review. *Ecological Processes*, *5*(1), 19. https://doi.org/10.1186/s13717-016-0063-3

Fernández, A., Usamentiaga, R., Carús, J., & Casado, R. (2016). Driver Distraction Using Visual-Based Sensors and Algorithms. *Sensors*, *16*(11), 1805. https://doi.org/10.3390/s16111805

Fitch, G. M., Bartholomew, P. R., Hanowski, R. J., & Perez, M. A. (2015). Drivers' visual behavior when using handheld and hands-free cell phones. *Journal of Safety Research*, *54*, 105.e29-108. https://doi.org/10.1016/j.jsr.2015.06.008

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50.

Garay-Vega, L., Pradhan, A. K., Weinberg, G., Schmidt-Nielsen, B., Harsham, B., Shen, Y., … Fisher, D. L. (2010). Evaluation of different speech and touch interfaces to in-vehicle music retrieval systems. *Accident Analysis & Prevention*, *42*(3), 913–920. https://doi.org/10.1016/j.aap.2009.12.022

Geitner, C., Sawyer, B. D., Birrell, S., Jennings, P., Skyrypchuk, L., Mehler, B., & Reimer, B. (2017). A Link Between Trust in Technology and Glance Allocation in On-Road Driving. *Proceedings of the 9th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design: Driving Assessment 2017*, 263–269. Manchester Village, Vermont, USA: University of Iowa. https://doi.org/10.17077/drivingassessment.1645

Hassan, H. M., & Abdel-Aty, M. A. (2011). Analysis of drivers' behavior under reduced visibility conditions using a Structural Equation Modeling approach. *Transportation Research Part F: Traffic Psychology and Behaviour*, *14*(6), 614–625. https://doi.org/10.1016/j.trf.2011.07.002

Hedlund, J., Simpson, H. M., & Mayhew, D. R. (2006). *International conference on distracted driving: Summary of proceedings and recommendations: October 2-5, 2005*. CAA.

Hergeth, S., Lorenz, L., Vilimek, R., & Krems, J. F. (2016). Keep Your Scanners Peeled: Gaze Behavior as a Measure of Automation Trust During Highly Automated Driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *58*(3), 509–519. https://doi.org/10.1177/0018720815625744

Hoel, J., Jaffard, M., & Van Elslande, P. (2010, April). *Attentional competition between tasks and its implications*. Presented at the European Conference on Human Centred Design for Intelligent Transport Systems, Berlin, Germany.

Hoffman, L., McDowd, J. M., Atchley, P., & Dubinsky, R. (2005). The role of visual attention in predicting driving impairment in older adults. *Psychology and Aging*, *20*(4), 610–622. https://doi.org/10.1037/0882-7974.20.4.610

Horberry, T., Anderson, J., Regan, M. A., Triggs, T. J., & Brown, J. (2006). Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. *Accident Analysis & Prevention*, *38*(1), 185–191. https://doi.org/10.1016/j.aap.2005.09.007

Hox, J. J., & Timo, T. M. (1998). An introduction to statistical equation modeling. *Family Science Review*, *11*, 354–373.

Konstantopoulos, P., Chapman, P., & Crundall, D. (2010). Driver's visual attention as a function of driving experience and visibility. Using a driving simulator to explore drivers' eye movements in day, night and rain driving. *Accident Analysis & Prevention*, *42*(3), 827–834. https://doi.org/10.1016/j.aap.2009.09.022

Lam, L. W. (2012). Impact of competitiveness on salespeople's commitment and performance. *Journal of Business Research*, *65*(9), 1328–1334. https://doi.org/10.1016/j.jbusres.2011.10.026

Lee, C., Mehler, B., Reimer, B., & Coughlin, J. F. (2015). User Perceptions Toward In-Vehicle Technologies: Relationships to Age, Health, Preconceptions, and Hands-On Experience. *International Journal of Human-Computer Interaction*, *31*(10), 667–681. https://doi.org/10.1080/10447318.2015.1070545

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors*, *46*(1), 50–80.

Lee, J D., Young, K. L., & Regan, M. A. (2008). Define driver distraction. In *Driver Distraction: Theory, Effects, and Mitigation* (pp. 31–40). Boca Raton, FL, USA: CRC Press Taylor & Francis Group.

Lee, J., Mehler, B., Reimer, B., & Coughlin, J. F. (2016). Sensation Seeking and Drivers' Glance Behavior while Engaging in a Secondary Task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *60*, 1864–1868. https://doi.org/10.1177/1541931213601425

Lee, John D., Caven, B., Haake, S., & Brown, T. L. (2001). Speech-Based Interaction with In-Vehicle Computers: The Effect of Speech-Based E-Mail on Drivers' Attention to the Roadway. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *43*(4), 631–640. https://doi.org/10.1518/001872001775870340

Liang, Y., & Lee, J. D. (2010). Combining cognitive and visual distraction: Less than the sum of its parts. *Accident Analysis & Prevention*, *42*(3), 881–890. https://doi.org/10.1016/j.aap.2009.05.001

Llerena, L. E., Aronow, K. V., Macleod, J., Bard, M., Salzman, S., Greene, W., … Schupper, A. (2015). An evidence-based review: Distracted driver. *Journal of Trauma and Acute Care Surgery*, *78*, 147–152.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power Analysis and Determination of Sample Size for Covariance Structure Modeling. *Psychological Methods*, *1*(2), 130.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *Academy of Management Review*, *20*(3), 709–734.

McLaughlin, S., Hankey, J., & Dingus, T. (2009). Driver Measurement: Methods and Applications. In D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics* (pp. 404–413). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-02728-4_43

McWilliams, T., Reimer, B., Mehler, B., Dobres, J., & McAnulty, H. (2015). A Secondary Assessment of the Impact of Voice Interface Turn Delays on Driver Attention and Arousal in Field Conditions. *Proceedings of the 8th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design: Driving Assessment 2015*, 408–414. Salt Lake City, Utah, USA: University of Iowa. https://doi.org/10.17077/drivingassessment.1602

Mehler, B., Reimer, B., Dobres, J., & Coughlin, J. F. (2015). *Assessing the Demands of Voice Based In-Vehicle Interfaces—Phase II Experiment 3—2015 Toyota Corolla (2015b)* (No. MIT AgeLab Technical Report 2015-14 (November 28, 2015)). Massachusetts Institute of Technology, Cambridge, MA.

Mehler, B., Reimer, B., Dobres, J., McAnulty, H., & Coughlin, J. F. (2015). *Assessing the Demands of Voice Based In-Vehicle Interfaces—Phase II Experiment 1—2014 Chevrolet Impala (2014b)* (No. MIT AgeLab Technical Report 2015-6A (November 30, 2015)). Cambridge, MA: Massachusetts Institute of Technology.

Mehler, B., Reimer, B., McAnulty, H., Dobres, J., Lee, J., & Coughlin, J. F. (2015). *Assessing the Demands of Voice Based In-Vehicle Interfaces—Phase II Experiment 2—2014 Mercedes CLA (2014t)* (No. MIT AgeLab Technical Report 2015-8 (November 28, 2015)). Massachusetts Institute of Technology, Cambridge, MA.

Mehler, Bruce, Kidd, D., Reimer, B., Reagan, I., Dobres, J., & McCartt, A. (2016). Multi-modal assessment of on-road demand of voice and manual phone calling and voice navigation entry across two embedded vehicle systems. *Ergonomics*, *59*(3), 344–367. https://doi.org/10.1080/00140139.2015.1081412

Mehler, Bruce, Reimer, B., Dobres, J., McAnulty, H., Mehler, A., Munger, D., & Coughlin, J. F. (2014). *Further Evaluation of the Effects of a Production Level "Voice-Command" Interface on Driver Behavior: Replication and a Consideration of the Significance of Training Method* (p. 192). MIT AgeLab Technical Report.

Misokefalou, E., Papadimitriou, F., Kopelias, P., & Eliou, N. (2016). Evaluating Driver Distraction Factors in Urban Motorways. A Naturalistic Study Conducted in Attica Tollway, Greece. *Transportation Research Procedia*, *15*, 771–782. https://doi.org/10.1016/j.trpro.2016.06.064

Mohamed, M., & Bromfield, N. F. (2017). Attitudes, driving behavior, and accident involvement among young male drivers in Saudi Arabia. *Transportation Research Part F: Traffic Psychology and Behaviour*, *47*, 59–71. https://doi.org/10.1016/j.trf.2017.04.009

Nabatilan, L. B., Aghazadeh, F., Nimbarte, A. D., Harvey, C. C., & Chowdhury, S. K. (2012). Effect of driving experience on visual behavior and driving performance under different driving conditions. *Cognition, Technology & Work*, *14*(4), 355–363. https://doi.org/10.1007/s10111-011-0184-5

National Highway Traffic Safety Administration. (2013). *(Issued Guidelines) Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices (Docket No. NHTSA-2010-0053)*. Washington, DC: U.S. Department of Transportation National Highway Traffic Safety Administration (NHTSA).

NHTSA. (2020). *Distracted driving 2018 (Research Note. Report No. DOT HS 812 926)*. National Highway Traffic Safety Administration.

Noble, A. M., Miles, M., Perez, M. A., Guo, F., & Klauer, S. G. (2021). Evaluating driver eye glance behavior and secondary task engagement while using driving automation systems. *Accident Analysis & Prevention*, *151*, 105959. https://doi.org/10.1016/j.aap.2020.105959

Oh, H. J., Ko, S. M., & Ji, Y. G. (2016). Effects of Superimposition of a Head-Up Display on Driving Performance and Glance Behavior in the Elderly. *International Journal of Human-Computer Interaction*, *32*(2), 143–154. https://doi.org/10.1080/10447318.2015.1104155

Owens, J. M., McLaughlin, S. B., & Sudweeks, J. (2011). Driver performance while text messaging using handheld and in-vehicle systems. *Accident Analysis & Prevention*, *43*(3), 939–947. https://doi.org/10.1016/j.aap.2010.11.019

Owsley, C. (2011). Aging and vision. *Vision Research*, *51*(13), 1610–1622. https://doi.org/10.1016/j.visres.2010.10.020

Papantoniou, P., Papadimitriou, E., & Yannis, G. (2017). Review of driving performance parameters critical for distracted driving research. *Transportation Research Procedia*, *25*, 1796–1805. https://doi.org/10.1016/j.trpro.2017.05.148

Peissner, M., Doebler, V., & Metze, F. (2011). Can voice interaction help reducing the level of distraction and prevent accidents. *Meta-Study Driver Distraction Voice Interaction*, *24*.

Peng, Y., Boyle, L. N., & Lee, J. D. (2014). Reading, typing, and driving: How interactions with in-vehicle systems degrade driving performance. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*, 182–191. https://doi.org/10.1016/j.trf.2014.06.001

Perez, W., & Bertola, M. A. (2011). The Effect of Visual Clutter on Driver Eye Glance Behavior. *Proceedings of the 6th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design : Driving Assessment 2011*, 180–186. Olympic Valley-Lake Tahoe, California, USA: University of Iowa. https://doi.org/10.17077/drivingassessment.1395

Pettitt, M., Burnett, G., & Stevens, A. (2005). Defining Driver Distraction. *12th World Cong. on Intelligent Transport Systems*, *1*, 12.

R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Regan, M. A., Hallett, C., & Gordon, C. P. (2011). Driver distraction and driver inattention: Definition, relationship and taxonomy. *Accident Analysis & Prevention*, *43*(5), 1771–1781. https://doi.org/10.1016/j.aap.2011.04.008

Reimer, B., Mehler, B., Dobres, J., & Coughlin, J. F. (2013). *The Effects of a Production Level "Voice-Command" Interface on Driver Behavior: Reported Workload, Physiology, Visual Attention, and Driving Performance* (No. MIT AgeLab Technical Report No. 2013-17A; p. 445). Cambridge, MA: Massachusetts Institute of Technology.

Reimer, B., Mehler, B., Dobres, J., McAnulty, H., Mehler, A., Munger, D., & Rumpold, A. (2014). Effects of an "Expert Mode" Voice Command System on Task Performance, Glance Behavior & Driver Physiology. *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '14*, 1–9. Seattle, WA, USA: ACM Press. https://doi.org/10.1145/2667317.2667320

Reimer, B., Mehler, B., Wang, Y., & Coughlin, J. F. (2012). A Field Study on the Impact of Variations in Short-Term Memory Demands on Drivers' Visual Attention and Driving Performance Across Three Age Groups. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *54*(3), 454–468. https://doi.org/10.1177/0018720812437274

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. *The Journal of Educational Research*, *99*(6), 323–338. https://doi.org/10.3200/JOER.99.6.323-338

Schumacker, R. E., & Lomax, R. G. (2010a). Confirmatory Factor Models. In *A beginner's guide to structural equation modeling*. Taylor and Francis Group, LLC.

Schumacker, R. E., & Lomax, R. G. (2010b). Developing Structural Equation Models: Part I. In *A beginner's guide to structural equation modeling*. Taylor and Francis Group, LLC.

Schumacker, R. E., & Lomax, R. G. (2010c). Developing Structural Equation Models: Part II. In *A beginner's guide to structural equation modeling*. Taylor and Francis Group, LLC.

Schumacker, R. E., & Lomax, R. G. (2010d). Introduction. In *A beginner's guide to structural equation modeling*. Taylor and Francis Group, LLC.

Schumacker, R. E., & Lomax, R. G. (2010e). Model Fit. In *A beginner's guide to structural equation modeling*. Taylor and Francis Group, LLC.

Sheykhfard, A., & Haghighi, F. (2020). Driver distraction by digital billboards? Structural equation modeling based on naturalistic driving study data: A case study of Iran. *Journal of Safety Research*, *72*, 1–8. https://doi.org/10.1016/j.jsr.2019.11.002

Simmons, S. M., Caird, J. K., & Steel, P. (2017). A meta-analysis of in-vehicle and nomadic voice-recognition system interaction and driving performance. *Accident Analysis & Prevention*, *106*, 31–43. https://doi.org/10.1016/j.aap.2017.05.013

Sivak, M. (1996). The information that drivers use: Is it indeed 90% visual? *Perception*, *25*(9), 1081–1089.

Sivak, Michael. (1998). The information that drivers use: Is it indeed 90 percent visual? *The UMTRI Research Review*, *29*(1).

Stutts, J. C., Reinfurt, D. W., Staplin, L., & Rodgman, E. A. (2001). *The Role of Driver Distraction in Traffic Crashes: (363942004-001)* [Data set]. American Psychological Association. https://doi.org/10.1037/e363942004-001

Suhr, D. (2006). *The Basics of Structural Equation Modeling*. Presented at the SAS User Group of the Western Region of the United States (WUSS), Irvine, CA.

Swedler, D. I., Bowman, S. M., & Baker, S. P. (2012). Gender and Age Differences among Teen Drivers in Fatal Crashes. *Annals of Advances in Automotive Medicine/Annual Scientific Conference*, *56*, 97. Association for the Advancement of Automotive Medicine.

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tivesten, E., & Dozza, M. (2014). Driving context and visual-manual phone tasks influence glance behavior in naturalistic driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *26*, 258–272. https://doi.org/10.1016/j.trf.2014.08.004

Treat, J. R. (1980). A study of precrash factors involved in traffic accidents. *HSRI Research Review*.

Underwood, G. (2007). Visual attention and the transition from novice to advanced driver. *Ergonomics*, *50*(8), 1235–1249. https://doi.org/10.1080/00140130701318707

Wickens, C. D., Lee, J. D., Liu, Y., & Gordon, S. E. (2004a). Introduction to Human Factors. In *An introduction to human factors engineering*. Pearson Prentice Hall.

Wickens, C. D., Lee, J. D., Liu, Y., & Gordon, S. E. (2004b). Research Methods. In *An introduction to human factors engineering*. Pearson Prentice Hall.

World Health Organization. (2011). *Mobile phone use: A growing problem of driver distraction*. Geneva: World Health Organization WHO. Retrieved from http://www.who.int/violence_injury_prevention/publications/road_traffic/distracted_driving_en.pdf

World Health Organization. (2015). *Global status report on road safety 2015*. World Health Organization.

Zhang, T., Tao, D., Qu, X., Zhang, X., Lin, R., & Zhang, W. (2019). The roles of initial trust and perceived risk in public's acceptance of automated vehicles. *Transportation Research Part C: Emerging Technologies*, *98*, 207–220. https://doi.org/10.1016/j.trc.2018.11.018

Zhao, N., Mehler, B., Reimer, B., D'Ambrosio, L. A., Mehler, A., & Coughlin, J. F. (2012). An investigation of the relationship between the driving behavior questionnaire and objective measures of highway driving behavior. *Transportation Research Part F: Traffic Psychology and Behaviour*, *15*(6), 676–685. https://doi.org/10.1016/j.trf.2012.08.001

Zhao, N., Reimer, B., Mehler, B., D'Ambrosio, L. A., & Coughlin, J. F. (2013). Self-reported and observed risky driving behaviors among frequent and infrequent cell phone users. *Accident Analysis & Prevention*, *61*, 71–77. https://doi.org/10.1016/j.aap.2012.07.019

Zhao, X., Xu, W., Ma, J., Li, H., & Chen, Y. (2019). An analysis of the relationship between driver characteristics and driving safety using structural equation models. *Transportation Research Part F: Traffic Psychology and Behaviour*, *62*, 529–545. https://doi.org/10.1016/j.trf.2019.02.004

Zheng, Y., Shokouhi, N., Thomsen, N., Sathyanarayana, A., & Hansen, J. (2016, April 5). *Towards Developing a Distraction-Reduced Hands-Off Interactive Driving Experience using Portable Smart Devices*. 2016-01–0140. https://doi.org/10.4271/2016-01-0140