ABSTRACT

DENG, YULIN. Influence of Automation Transparency on Driver Skill Decay in Use of High-level Automated Driving Systems. (Under the direction of Drs. David Kaber and Jing Feng).

Developing high-level vehicle automation has been a continuing endeavor of automobile manufacturers, and automated vehicles have realized increasing popularity with the public. Earlier literature suggests that it is highly likely for drivers to experience manual skill decay with exposure to vehicle automation, which might lead to degraded vehicle control and deteriorated hazard negotiation performance. However, few, if any, empirical studies have focused on this issue, leaving a gap in driving research. Therefore, the first phase of the present study aimed to identify possible driving skill decay caused by vehicle automation, as indicated by deteriorated hazard negotiation performance and manual driving performance after exposure to automated driving. The second phase of the study attempted to enhance driver performance under automated driving. Some prior studies in the aviation domain have tested higher levels of automation transparency to address the “pitfalls” of automation and support pilot performance; however, generalizability of findings to the automobile domain has been rarely discussed. Therefore, Phase 2 investigated whether high-level vehicle automation transparency could mitigate negative impacts of vehicle automation on driving skills. The third part of this study attempted to develop a novel prediction model of crash outcomes, presenting the likelihood of crashes based on vehicle automation features and driver pupilography. While a large volume of studies has focused on crash prediction, little attention is paid to automated driving, and few studies have considered using automated vehicle settings and real time driver pupil size change as predictors.

In Phase 1, 18 drivers took part in a driving simulation study. Three levels of automation: 1) manual driving; 2) adaptive cruise control (ACC); and 3) integration of ACC and lane keeping

assistance (Level 2 type of automation) were simulated. Participants' driving performance was compared after two durations of exposure to vehicle automation (short and extended). Results from this phase demonstrated skill decay associated with vehicle automation, as indicated by degraded hazard negotiation performance, and deteriorated lane and speed maintenance skill post automation exposure, compared to manual driving.

The second phase of experiments included 24 drivers. Similar to Phase 1, duration of exposure (short and extended) and levels of automation (ACC and Level 2 type of automation) were manipulated. Additionally, automation transparency was introduced as a new variable, which was manipulated by varying the presence of audio information on vehicle automation system states. Results further demonstrated skill decay related to higher level of vehicle automation, as indicated by degraded lane and speed control. Findings also provided evidence of high-level automation transparency improving hazard negotiation performance. Furthermore, high-level automation transparency had some merit for enhancing driving performance after automation exposure. Post-trial questionnaires also supported recommendations for utilizing high-level automation transparency.

The third part of the study adopted an all (data) subset approach to identify a best fitting crash prediction model. All combinations of driver performance predictors were investigated for predicting crash outcomes and the model with the highest r-square value was selected. However, contrary to expectations, even the selected model had a relative low goodness-of-fit. This outcome may have been due to an inadequate number of predictors and small experiment sample sizes. The results of these analyses, therefore, need to be interpreted with caution.

Results of this research are expected to advance our understanding of driver behavior changes related to use of vehicle automation. The work also provides a basis for enhancing the design of information systems in automated vehicles. Finally, the study provides a reference for future development of a comprehensive crash outcome prediction model.

Influence of Automation Transparency on Driver Skill Decay in Use of High-level Automated Driving Systems

by
Yulin Deng

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Industrial Engineering

Raleigh, North Carolina

2019

APPROVED BY:

Dr. David Kaber Committee Co-Chair	Dr. Jing Feng Committee Co-Chair
Dr. Xu Xu	Dr. Peter Bloomfield

BIOGRAPHY

Yulin Deng obtained a bachelor’s degree from The Hong Kong Polytechnic University with double degrees in Industrial Engineering and Business Administration. She came to North Carolina State University in 2015 in hope of pursuing a Ph.D. She received her master’s degree from Industrial and System Engineering Program at North Carolina State University.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to all the persons who have helped me in this endeavor.

Foremost, I'm heartily grateful to my advisor, Dr. David Kaber, whom for without his generous guidance, support and understanding I wouldn't have made it this far.

I would also like to thank my co-chair, Dr. Feng, for all her expertise and support for completion of this project. In addition, I want to thank my committee members, Dr.Xu, Dr. Bloomfield, and graduate school representative Dr. Ghosh. Their inputs are highly valued and appreciated.

I would also like to thank all the scholars I have worked with throughout the years of my study, especially Mr. Chris Cunningham from Institute of Transportation Research and education (ITRE). It has been a great pleasure working with him and the colleagues in ITRE. In addition, I'm grateful to the members of the Ergonomics Lab who have helped me out throughout my journey at North Carolina State.

Finally, I owe my profound gratitude to my family, who have always been supporting and encouraging me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1. Automation in Personal Vehicles

1.1.1. Common Automated Driving technologies

Within the past 5 years there has been a dramatic increase in the number and type of automated vehicles. Increasingly powerful automation systems have also been developed, and semi-autonomous features are now available in both luxury and less expensive vehicle models (Dikmen & Burns, 2016). The wide variety of relevant technologies inspired the International Society of Automotive Engineers' (SAE) to define five levels of driving automation (see Table 1). Each succeeding level of automation, from zero to five, indicates increasing vehicle automation functionality.

Table 1. Summary of SAE levels of automation (SAE, 2016).

SAE Level	***Name***	***Description***
Level 0:	No automation	Manual driving
Level 1:	Driver Assistance	Automation system performs sustained execution of either lateral or longitudinal vehicle motion control.
Level 2	Partial Automation	Automation system performs sustained execution of both steering and acceleration/deceleration. Drivers must remain engaged and monitor the environment at all times.
Level 3	Conditional Automation	Drivers are not required to monitor the environment at all time, but must respond appropriately to a request to intervene in automated control.
Level 4	High Automation	Vehicle is capable of performing all driving functions under some Driving modes (under some geographic areas, roadway types, weather conditions, etc.). Drivers are not required to respond appropriately to a request to intervene.
Level 5	Full Automation	Vehicle is capable of performing all driving functions under all Driving modes.

Many of today's commercial vehicles have driver assistance capabilities (Level 1), typically adaptive cruise control (ACC) or lane keeping assist (LKA; Smith, 2015). A small number of vehicle models have incorporated both functions to make them capable of partial automation (Level 2; Smith, 2015). To date few commercial vehicle models have include conditional automation (level 3) capabilities. Therefore, the present study will focus on the three available levels of automation, namely manual driving (level 0), driver assistance (level 1), and partial automation (level 2).

Among the above-mentioned vehicle automation technologies, adaptive cruise control (ACC) has been widely implemented in modern automated vehicles. A large number of automobile manufacturers, including Tesla, Volvo, BMW, Acura, Chevrolet, Dodge, Audi, and Ford, have made the ACC system available in their vehicle models (Trimble et al., 2014). ACC manages the longitudinal speed of a vehicle and automatically maintains a constant safe headway to a lead vehicle; drivers remain responsible for control of lateral speed of the vehicle. The drivers also remain responsible for supervising the system and take over when necessary (SAE, 2016). Adaptive cruise control, when used alone belongs to Level 1 vehicle automation (SAE, 2016). Figure 1 presents a typical in-vehicle ACC control interface.

Figure 1: ACC Control Interface in 2014 Jeep Cherokee (from https://www.edmunds.com/jeep/cherokee/2014/long-term-road-test/2014-jeep-cherokee-limited-active-cruise-control-works-well.html).

Lane Keeping Assistant (LKA), also referred to as active steering (AS), is another common form of vehicle automation technology. LKA provides steering inputs and maintains the vehicle in the center of a detected lane. This is enabled by onboard cameras that can detect the lane markings and the position of the vehicle within the lane (Young & Stanton, 1998). LKA is available in commercial vehicles developed by major manufacturers, including Toyota, Ford, Audi, Tesla, Honda, and Volvo (Trimble et al., 2014). Similar to ACC, under LKA, drivers are responsible for supervising the system, and they should be ready to take over when necessary. LKA belongs to Level 1 vehicle automation, when used alone (SAE, 2016). Figure 2 presents a typical in-vehicle LKA control interface.

Figure 2. LKA Control Interface in 2018 GMC Terrain (From http://sandyblogs.com/techlink/?p=8224).

Integration of LKS and ACC enables automatic lateral and longitudinal vehicle control, which has become the market trend in recent years (Ziebinski et al., 2017). Only a limited number of commercial vehicle models can provide such automation. Examples of these systems include the Toyota Highway Driving Assistant (Trimble et al., 2014), GM Super Cruise (Trimble et al., 2014), Tesla Autopilot (Tesla, n.d.), and Volvo Pilot Assist (Volvo, n.d.). The combination of ACC and LKA belongs to Level 2 vehicle automation. Therefore, all these commercial vehicle automation systems require drivers to stay attentive at all times, supervise the vehicle's actions during use, and takeover control of vehicle when necessary. Figure 3 presents a vehicle control interface with both ACC and LKA controls available.

Figure 3. Super Cruise (ACC+LKA) Control Interface in 2018 Cadillac CT6. From (https://mobilesyrup.com/2018/01/14/cadillac-super-cruise-review-learning-not-drive/).

1.1.2. Usage of Automated Driving Technologies

To investigate the use of ACC in commercial vehicles, Grove et al. (2015) studied 33,000 hours of naturalistic driving data collected throughout the United States. Results revealed that drivers used ACC systems to control the headway of their vehicles for 6.0% of the time in normal weather, and drivers set their ACC speed to an average of 63 mph. Although manufacturers of ACC systems generally do not recommend their use in adverse conditions, drivers showed similar usage of ACC in adverse conditions.

Eichelberger and McCartt (2016) conducted a survey of 183 vehicle owners on the use of ACC in 2010–2013 Toyota Sienna and Prius models. Results showed that 86% of owners had used ACC at some point, among whom 60% always used it on freeways, whereas 10% reported that they also used it on lower speed roads. Drivers most often (42%) used the longer headway setting (160 ft) compared to medium (130ft) and short (100ft) headway settings.

Eichelberger and McCartt (2016) also investigated in the use of LKA among 116 owners of 2013 Toyota Sienna and Prius models with LKA systems. According to their survey, 13% of the users reported that they always used the system, 46% reported using it sometimes, and 41% rarely or never used it. Few other studies were found on the use of LKA systems.

One other study was found to focus on the use of the integration of ACC and LKA systems. Dikmen and Burns (2016) conducted a survey with 121 Tesla owners to study the use of the Tesla Autopilot, which integrates ACC and LKA systems. Results showed that the majority of drivers actively used the automation system, with 31.2% saying they used it “always” and 57.8% saying they used it “often”.

Together, these studies suggest relatively frequent use of automated driving systems by commercial vehicle owners primarily under freeway driving conditions. The frequency of automation use highlights the need to explore the impact of vehicle automation on driver behavior and safety.

1.1.3. Crash Statistics

Recent years have seen an increasing number of crashes involving automated vehicles, as more automated vehicles are put on public roads (Schoettle & Sivak, 2015). Several recent accidents involving automated vehicles have gained national attention. In March 2018, a Volvo XC90 SUV Uber car in autonomous mode struck a pedestrian crossing the street in Arizona and the driver did not intervene in a timely manner (Wakabayashi, 2018). Later that month, a Tesla Model S was involved in a fatal crash in Utah with the Autopilot mode engaged and the driver was reported to be distracted prior to the crash (Boudette, 2018). These well-known accidents have raised extensive discussion on safety issues associated with vehicle automation.

Figure 4. Fatal Crash of a Tesla Model S in Utah. From (http://www.foxnews.com/auto/2018/05/17/data-says-tesla-driver-had-hands-off-wheel-before-utah-crash.html).

Statistics reveal that from implementation through October 2015, the Google Self-Driving Car was involved in 16 crashes (Blanco et al., 2016). Despite the relative low total number of crashes involving automated vehicles, researchers pointed out two major sources of uncertainty in the safety of automated vehicles. Schoettle and Sivak (2015) suggest that automated vehicles are driven mainly in limited and less demanding driving conditions and Blanco et al. (2016) argued that automated vehicles have relatively low public exposure.

Overall, the physical evidence presented in this section indicates need for further understanding of the safety implications of vehicle automation. The evidence also suggests that it is necessary to explore effective methods to improve driver engagement and hazard negotiation performance while using automated driving systems (ADSs).

1.2. Vehicle Automation and Driving Skills

The loss of operator skills associated with long-term use of advanced automation has been explored in many different contexts, including continuous flow production (Shiff, 1983 (bakery operations); Endsley & Kiris, 1995 (automated vehicle navigation and control)). Parasuraman, Sheridan, and Wickens (2000) outlined a theoretical approach to the analysis of human performance under automation. They concluded that if decision-making functions are consistently performed by automation, human operators might experience skill decay after a period of time. They also pointed out the gravity of cognitive skill degradation following automation failure. Bainbridge (1983) also identified the possibility of manual and cognitive skill decay as a consequence of automation.

Operator skill decay has also been observed in the field of aviation. In the pilot community, the problem of operator skill degradation has been generally recognized and has been attributed, in part, to failure to exercise those skills over time (Wickens, 1995). Through discussion with individuals managing pilot training, Wiener and Curry (1980) observed perceptible skill loss in pilots who use automatic equipment. McClumpha (1991) conducted a survey with civil pilots regarding their attitude towards automation, and the majority of pilots reported erosion of flying skills by automation.

In view of all that has been mentioned so far, one may suppose that it is possible for drivers to experience skills degradation as well, if they are exposed to vehicle automation for a period of time. Therefore, the following sections of this study will discuss the components of driving skills, as well as the effect of vehicle automation on driving skills.

1.2.1. Components of Driving Skills

Driving is a complex combination of different skills (Lajunen & Summala, 1995). Many studies have endeavored to identify the set of component skills involved in driving and researchers have come to different conclusions.

It is commonly believed that driving is a perceptual-motor skill, which includes cognitive information processing and motor performance (Lewin, 1982; Lajunen & Summala, 1995). However, disagreements exist regarding the components of driving skill. Dunkan et al. (1991) decomposed driving skills into six separate skills, including: information acquisition (especially scanning), perceptual-motor coordination, assessment of traffic situations, risk estimation, setting safety margins, and balancing disparate attractions of speed and caution. Horswill, Waylen and Tofield (2004) considered driving skill to have as many as 18 components (see Table 2) and they used subjective scores of each of the 18 components to assess driving skill.

Table 2. Components of driving skill (Horswill et al., 2004).

No.	**Components of Driving Skill**
1	Appropriate use of gears
2	Hill starts
3	Adapting driving to changing conditions
4	Appropriate use of signals
5	Knowledge of the highway code(road rules)
6	Controlled emergency stops
7	Awareness and anticipation of other road users' behavior
8	Maintaining an appropriate following distance from the car in front
9	Parking
10	Maintaining appropriate speed for the conditions
11	Proper use of mirrors
12	Awareness and anticipation of pedestrian activity
13	Monitoring junctions, bends, etc.
14	Reversing and maneuvering
15	Merging with traffic and changing lanes
16	Smooth cornering
17	Knowing when to overtake
18	Aware of your own fitness to drive

According to Zhang et al. (2010), driving skill is a driver's dynamic behavior, representing their dynamic capability to operate a vehicle. Other researchers also suggest that driving skill is a predictor of driver maximum performance capability (Lajunen & Summala, 1995). For example, evidence from previous studies suggests a strong link of driving skills to accidents and crash avoidance. According to Evans (1991), lower level of driving skills is associated with increased time to detect hazards and delayed response times. In other words, at a constant level of driving task demand, increases in skills are likely to lead to increases in safety and, thus, fewer crashes. Other researchers (McKenna & Horswill, 1999; Pelz & Krupat, 1974) reported similar findings.

Two components of driving skills have been found to be critical in hazardous driving conditions, including: hazard perception (Deery & Love, 1996; Deery, 1999; Horswill, Waylen & Tofield, 2004) and vehicle control (Horswill, Waylen & Tofield, 2004; Fuller, 2005). Hazard perception is defined as "the ability of drivers to anticipate potentially dangerous road situations" (McKenna & Horswill, 1999), and vehicle control is defined as the ability to maintain full control of the vehicle while driving (Horswill, Waylen & Tofield, 2004).

These and other driving skills are not necessarily constant over time and they can either improve or deteriorate with driving experience (Duncan, Williams & Brown, 1991). Deterioration of driving skills is commonly caused by psychological factors (e.g. tension, emotion) and physiological factors (Lewin 1982). It is worth noting that the terminology of skill decay, skill degradation and perceptible skill loss are used interchangeably in the human factors literature and are considered to refer to the same concept.

1.2.2. Effect of Vehicle Automation on Driving Skills

To further understand driver behavior under automated driving conditions, a literature review was conducted on driver performance with Level 1 types of vehicle automation (i.e., ACC) and Level 2 types of automation (i.e., ACC and LKA) with a focus on those studies making comparison with manual driving.

Regarding the effect of vehicle automation on driving skills under non-critical conditions, the literature has revealed contradictory findings. Gold et al. (2013) compared driver performance under manual driving and highly automation driving conditions. They observed less stable acceleration behavior of drivers exposed to automated driving. Two other studies reported degraded lane keeping performance of drivers using ACC, as compared to manual driving (Rudin-Brown et al., 2003; Hoedemaeker & Brookhuis, 1998). However, some researchers observed comparable lane and speed maintenance performance produced by manual and automated driving (Stanton et al., 1997; Young & Stanton, 1997). In the study by Ma and Kaber (2005), drivers exhibited superior lane maintenance performance under ACC control in comparison to manual driving.

1.2.3. Effect of Vehicle Automation on Driver Hazard Negotiation Performance

Under critical conditions, drivers still play a crucial part in hazard negotiation when driving an automated vehicle, as Level 1 and Level 2 automation requires human drivers to actively monitor the driving environment and take over when necessary (Smith, 2015). As mentioned in the discussion of driving skill components, the level of driving skills can be a reliable indicator of driver hazard negotiation performance (Evans, 1991; McKenna & Horswill, 1999; Pelz & Krupat, 1974). Several studies were found that investigated the effect of vehicle automation on driver hazard negotiation performance; however, contradictory results have been reported on this issue.

Rudin-Brown et al. (2003) compared driver reaction time to safety-critical events under manual driving and ACC. Their results showed that drivers using ACC took longer to react to lead vehicle brake lights and they were more likely to crash. They concluded that cognitive or attentional resources freed-up by ACC were used for "other purposes" instead of hazard avoidance.

Merat et al. (2012) conducted a driving simulator study to compare driving performance in manual and highly automated driving (SAE Level 2). Their results showed that in the absence of secondary tasks, driver responses to critical incidents under automated driving conditions were similar to manual control. In the presence of secondary tasks, drivers exhibited degraded response performance to critical incidents under automated driving conditions.

Gold et al. (2013) also observed longer reaction time to hazards during highly automated driving, as compared to manual driving. They attributed these observations to driver inattention and distraction under automated driving conditions.

Schleicher and Gelau (2011) also investigated the effect of vehicle automation on driver responses in critical situations. They reported delayed responses to hazards with CC or ACC, as compared to manual driving. The authors observed that drivers took some time to notice a situation that posed driving demands beyond automation’s capability as well as the need to intervene and handle the issue. They explained that drivers paid reduced attention to signs or cues that required adaptation of speed, because the speed was expected to be controlled by automated systems.

Some other literature was found that did not make direct comparison of driver hazard negotiation performance under different levels of vehicle automation but provided important insights on this issue. Young and Stanton (1997) speculated that drivers under automated driving conditions would have difficulty coping with and avoiding collisions. They argued that lower mental workload in automated driving (Ma & Kaber, 2005; Shanton et al., 1997; Rudin-Brown et al., 2003; Flemisch et al., 2008) would lead to a rapid increase in mental demand when drivers were put in a hazardous situation requiring control intervention. Therefore, driver hazard negotiation performance would be deteriorated under automated driving.

Jamson et al. (2013) compared manual driving and automated driving (Level 2) under different traffic conditions. They observed that drivers demonstrated increasing symptoms of fatigue with vehicle automation. Drivers showed more of a propensity to become involved with in-vehicle tasks when using vehicle automation. They concluded that drivers were less attentive under automated driving but the inattentiveness could be mitigated as traffic conditions became heavier.

Contradictory results have also been reported regarding driver situation awareness (SA) under automated driving conditions. Ma and Kaber (2005) found that ACC led to greater driver SA on a driving task and roadway conditions due to reduced driver workload. However, three other studies reported lower levels of SA in automated driving, as compared to manual driving (Rudin-Brown et al., 2003; Young & Stanton, 1997; Gold et al., 2013). These studies found that driver attention to the roadway was degraded and they were more likely to be distracted under automated driving.

In conclusion, previous research findings on the effect of vehicle automation on driver performance have been inconsistent and contradictory. Therefore, further research is necessary to enhance understanding of driver behavior in automated driving. It is also worth noting that all of the above studies compared either Level 1 or 2 types of vehicle automation with manual driving. Few driving simulation studies have compared driver hazard negotiation performance under two or more types of vehicle automation. Therefore, it is important for future research to make comparison of all currently available forms of vehicle automation (Level 0, Level 1 and Level 2) to provide a more complete mapping of the effect of vehicle automation on driver performance.

1.3. Automation Transparency and Driver Performance

1.3.1. Definition of Automation Transparency

Automation transparency is defined as "the degree to the extent which the inner workings or logic of the automated systems are known to human operators to assist their understanding about the system" (Seong & Bisantz, 2008). Generally, automation transparency affords understanding and predictions about a system's behavior and the information shared can include: (1) what the automation is currently doing, (2) what information is being used, (3) how information is being processed, and (4) when it was provided (Westin et al., 2016). In general, systems with higher levels of automation transparency support greater levels of information sharing with operators (Westin et al., 2016).

1.3.2. Effect of Automation Transparency on Driver Performance

The effect of automation transparency on operator behavior has been researched. Yang et al. (2017) suggested that clear and efficient information communication to human operators by an automated system could enable operators to maintain proper SA on system states. De Waard et al. (1999) also suggested that system functionality needed to be communicated clearly to operators in highly automated situations. In terms of empirical work, Mercado et al. (2016) observed unmanned vehicle operator performance under different levels of automation transparency. They found that operator performance increased as a function of transparency level. Conflicting results have been reported regarding the effect of automation transparency on operator workload. While some researchers suggest that increasing levels of information presentation should increase operator workload, some believe that providing additional information can reduce operator workload (Chen et al., 2014).

With respect to vehicle automation, previous studies have indicated the need for clear feedback from vehicle automation systems. Endsley (2017) conducted a naturalistic driving study on the autonomy features of a Tesla Model S and recorded driver experiences over a six-month period. It was found that relying on manual control and instructions was not adequate for the driver to understand how the automation functions, which resulted in false assumptions and unexpected experiences created by the system. It is worth noting that this research used a sample of one; thus, the result of the research is likely not generalizable to the broad population of drivers.

Related to this study, the National Highway Traffic Safety Administration (NHTSA) recommends that manufacturers consider human-machine interaction issues in designing vehicle interfaces in order to, "minimize the potential for mode confusion to occur." They recommend the use of, "intuitive feedback from vehicle dynamics and/or warnings to the driver" (Habib, 2017). Such actions in the design process may action for poor vehicle user instruction sets.

In addition, three other relevant studies were discovered to offer insights into the effect of vehicle automation transparency on driver performance. Beggiato et al. (2015) investigated driver information needs during partially and highly automated driving. An expert focus group concluded that information should provide transparency, comprehensibility, and predictability of system actions. Information provided by automation systems should include the current system status, the remaining time to a change in the level of automation, the fallback level as well as reasons and a preview for ongoing and subsequent maneuvers. A follow-on driving simulator experiment revealed that information requirements during automated driving were primarily related to supervision of vehicle systems, including status, transparency, and comprehensibility of system actions.

A study by Koo et al. (2015) made use of a driving simulator with auto-braking functions (Level 1 vehicle automation) to examine how verbalized messages affect driver performance. Results revealed that providing information on “what” actions are being performed (e.g., “The car is braking”) and “why” actions are occurring (e.g., “Obstacle ahead”) produced the safest driving performance. The authors concluded that verbalized information accompanying a car’s autonomous action could significantly affect driver attitudes and safety performance.

Naujoks et al. (2015) conducted a driving simulator-based study to examine the effect of speech output on driver performance. Results revealed that communicating upcoming automated maneuvers by speech decreased driver workload and interaction with non-driving-related tasks.

Evidence presented in this section suggests that feedback from current commercial vehicle automation systems is not sufficient to ensure drivers have a clear and thorough understanding of vehicle automation systems. Overall, previous literature indicates a potential positive influence of vehicle automation transparency on driver performance. However, a limited number of studies have focused on the impact of automation transparency on driver performance. Among the available literature, the majority of studies concentrate on driver performance under non-critical driving conditions rather than hazardous conditions. Therefore, it is worthwhile to further examine the effectiveness of using higher levels of vehicle automation transparency to improve driver performance, under both non-critical and hazardous conditions.

1.4. Crash Outcome Prediction

It is generally believed that the likelihood of vehicle crashes is affected by driver factors, vehicle factors, and environment-related factors (Cantor et al., 2010). Given the variety of contributing factors, a considerable number of studies have adopted different approaches to predict the probability of crash involvement.

Guo and Fang (2013) developed a negative binomial regression model for predicting the likelihood of crash and near crash occurrences, using driver age, personality indicators and previous critical incident rates as predictors. Wu et al. (2014) predicted driver crash severity with a mixed logit model, which included environmental factors, vehicle types, roadway conditions and driver demographic information. Cantor et al (2010) used driver weight, demographic data, and accident history to predict truck crash likelihood. Two other studies used driver self-reported behavior questionnaires to predict accident involvement (De Winter & Dodou, 2010; Sümer, 2003). Ball et al. (1993) predicted vehicle crashes in older drivers based on eye-health status and useful visual field of view. Adanu et al.(2017) used dynamic driver performance (speed and distraction) as predictive variables in a logistic regression model to predict crash severity. However, the predictors investigated in this study are all binary variables (0 or 1). Other studies also used driver eye movement to model driver fatigue and drowsiness (Ji et al., 2004; Jo et al., 2014), but the results were not directly linked to crash involvement.

To summarize this research, although there is a strong link of dynamic driving performance to accidents and crash involvement, thus far, few studies have attempted to predict crash likelihood based on real-time driver performance. Furthermore, only a limited number of researchers have included automated driving system functionalities in crash outcome predictions.

1.5. Motivation for research

Strong evidence of operator skill decay due to aerial vehicle automation has raised similar concern of erosion of driving skills in use of vehicle automation. Recent statistics of crashes involving automated vehicles as well as contradicting views on the impact of vehicle automation on driving skills indicate needs for further study of this issue. However, a limited number of studies have discussed the possible effect of vehicle automation on driving skill degradation, especially under hazardous situations. Therefore, the present study aims to fill this knowledge gap and facilitate understanding of possible driving skill decay associated with vehicle automation use and the potential on non-hazardous and hazardous condition negotiation. It is also worth reiterating that few research studies have directly compared multiple levels of vehicle automation, including Level 0, Level 1 and Level 2, which seems necessary given the prevalent application of these three forms of automation technologies in commercial vehicles. By comparing driving skills with the three currently available levels of automation technology, the present study will provide a more complete mapping of the effect of vehicle automation on driver performance.

Previous research also indicates the need and benefit of increased automation transparency. Automation transparency tends to afford operators with understanding and the capability to predict automation system states and, thus, improve performance. As indicated by Endsley (2017), current instructions in commercial automated vehicles are not adequate for drivers to understand vehicle automation systems. However, research to date has tended to focus on automation in general rather than vehicle automation. Considering the previously mentioned potential "pitfalls" of vehicle automation, it is worthwhile to explore the possibility of improving driver performance with higher levels of vehicle automation transparency. This topic has rarely been discussed in previous driving simulation studies.

A large number of researchers have endeavored to predict crash likelihood based on driver demographic information, personality indicators, or self-reported driving behavior; however, few have used driver performance as a crash outcome predictor. Considering the link between driver performance and accident involvement, there is a need to consider dynamic driver performance in crash outcome prediction. Moreover, thus far, few studies have focused on predicting crash outcomes during automated driving. In view of the research that has been conducted to date, automated driving system functionalities, including level of automation and automation transparency, may influence driver hazard negotiation and accident involvement. Consequently, these factors should be taken into consideration in order to develop accurate crash outcome prediction models.

To summarize this review, the objective of this study is threefold. First, the work will identify possible driving skill decay caused by vehicle automation, as indicated by deteriorated driving performance under non-hazardous conditions and hazard negotiation performance. Second, the study will investigate the effectiveness of using higher level of automation transparency as a remedy for potential driving skill decay. Third, the study will develop a crash outcome prediction model, which presents the likelihood of a crash based on vehicle automation features and driver performance data.

1.6. Hypotheses

Based on the findings of the existing literature, the below hypotheses (H) were formulated.

With respect to driving performance under non-critical conditions, it was expected that higher levels of automation would produce worse driving performance, meaning higher speed deviations (H1) and higher lane deviations (H2) would occur during manual control after exposure. Longer exposure to automation was also hypothesized to produce higher speed deviations (H3) and higher lane deviations (H4). Introducing higher level automation transparency was expected to result in lower speed deviations (H5) and lower lane deviations (H6).

In terms of hazard negotiation performance, higher levels of vehicle automation were expected to lead to longer hazard reaction times (when responding with manual control; H7) and shorter time to collision (H8), as compared to completely manual driving. Longer exposure to automation was also expected to lead to degraded hazard negotiation performance, specifically, longer hazard reaction time (when responding with manual control; H9) and shorter time to collision (H10). Introducing higher level automation transparency was expected to result in shorter hazard reaction times (H11) and greater time to collision (H12).

2. METHOD

2.1. Apparatus

2.1.1. Driving Simulator

The North Carolina State University Driving Simulator is a Forum8 UC-win/Road driving simulator, developed by FORUM8 Co., Ltd (Tokyo, Japan). The setup includes six 55-inch high definition television monitors arranged to surround a driver in the simulator cab and provide a 315-degree field of view (FOV) of the driving environment (see Figure 5). The full-motion simulator integrates a Moog 6-DOF hexapod (1000Kg capacity) with Forum8's Virtual Drive advanced driving simulation modeling tool and simulation engine. An audio speaker system provides auditory feedback for drivers. Drivers interact with the simulator through a cab control unit, which consists of a modular steering unit with a full-size wheel, and a modular accelerator and brake pedal unit (see Figure 6).

The ownship was simulated as a coupe at 5.01m in length and 2.22m in width. The simulated vehicle was similar to a Toyota Echo 1300 CC model, with automatic transmission. The maximum brake force was 9800N, and the minimum turning radius was 9.8 degree. The wheels provided a rolling resistance of 137.2N. The steering wheel had a resistance factor of 3.

Figure 5: Forum 8 Driving Simulator at North Carolina State University.

Figure 6. Forum 8 Driving Simulator Cab.

2.1.2. Pupil Lab Eye Tracker

The driving simulator setup integrated a portable eye tracking system that was used to capture participant pupliometry. To collect real-time eye movement data, a Pupil Labs eye tracking headset was used, which can be worn like a pair of glasses by a driver (see Figure 7). A world-view camera was mounted on the top of the headset with a 100-degree FOV. Eye cameras were mounted on each side of the headset and could capture left and right eye movements with a sampling frequency of 200 Hz and an average gaze estimation accuracy of 0.6 degree of visual angle (Kassner, Patera & Bulling, 2014). According to prior literature, decrease in pupil diameter is associated with fatigue (Chi & Lin, 1998; Geacintov& Peavler, 1974; LeDuc et al., 2005). In the present study, pupilometry was applied as a secondary form of fatigue analysis providing physiological evidence of the presence or absence of driver fatigue, as manipulation check on the fatigue questionnaire. As noted, the questionnaire was the primary measure of driver fatigue (delivered in near real-time) and used as a basis for determining additional driving task breaks. In fact, data quality issues precluded the pupil diameter's use even in a secondary role (see section 3.1).

Figure 7. Pupil Lab Eye Tracker.

2.2. Task Description

Each participant completed 10 experiment trials. All driving simulations presented a rural highway with two lanes on each side of the highway and Level of Service A traffic on the road. Level of service A traffic is defined as:

"Free flow. Traffic flows at or above the posted speed limit and motorists have complete mobility between lanes. The average spacing between vehicles is about 550 ft. (167 m) or 27 car lengths. Motorists have a high level of physical and psychological comfort. The effects of incidents or point breakdowns are easily absorbed. LOS A generally occurs late at night in urban areas and frequently in rural areas." (Transportation Research Board, 2010).

Logo signs and road signs were included to represent a realistic driving environment. Although the road configuration was the same for all experiment trials, each trial had slightly different environmental features in order to promote driver attentiveness throughout the experiment (Tiffault & Bergeron, 2003; Benett et al., 1974).

Under manual driving conditions, participants were instructed to drive in the right lane of the freeway and maintain the exact posted speed limit (65 mph). Other than these requirements, drivers were expected to drive in their normal manner.

Throughout the experiment, two types of hazard events were presented. The first type of hazard involved an abrupt lane incursion by a leading vehicle. The second type of hazard involved sudden breaking by a leading vehicle. One instance of each type of hazard took place during the experiment. The first hazard took place in the fifth driving trial and the second hazard occurred in the tenth trial. The order of the two types of hazards were randomized. As indicated by prior literature, driving skill is a reliable indicator of a driver's maximum performance capability (Lajunen & Summala, 1995). Therefore, to ensure that drivers achieved their full potential, experimenters would inform drivers of the possibility of hazard occurrence prior to experiment trials. However, the order and timing of hazards were not revealed to prevent drivers from predicting hazard occurrence.

Twelve participants made use of an ACC system while driving. The driving simulator automatically controlled the speed of the vehicle and maintain a constant headway distance to all lead vehicles. In this experiment, vehicle speed was maintained at 65mph and the headway distance at 160 ft. Drivers were instructed to control their vehicle in the lateral direction and drive in the right lane of the freeway at all times. The same hazard events as those presented under manual driving would take place during ACC driving. The automated driving participants were also informed of the possibility of hazard exposure before test trials. At the time of a hazard, drivers were expected to react to hazards in their own manner, by braking only. After the completion of hazard negotiation, drivers were exposed to a 2-minute manual drive, during which they were assessed on their lane keeping and speed maintenance capability. Their performance during this observation period were compared with the performance of drivers in same observation period after having used manual vehicle control.

Another twelve participants drove with both ACC and LKA automation. The vehicle automation system controlled both the longitudinal and lateral speed of the vehicle. The vehicle speed and headway distance settings were the same as those under the ACC driving condition. The same hazard events as those presented under manual driving were posed to driver using the ACC and LKA. Drivers were expected to react to hazards in their own manner, by steering or braking. Similar to ACC driving condition, drivers were exposed to a 2-minute manual drive after hazard negotiation, and their lane deviation and speed deviation performance were recorded.

With reference to the definition of driving skill, in order to complete the tasks in this study, drivers needed to use cognitive information processing and motor control skills (Lewin, 1982; Lajunen & Summala, 1995) under non-critical conditions and make use of hazard perception (Deery & Love, 1996; Deery, 1999; Horswill, Waylen & Tofield, 2004) and vehicle control skills (Horswill, Waylen & Tofield, 2004; Fuller, 2005) under hazardous conditions. If we further break-down the skills needed to complete the tasks, drivers used 8 out-of the 18 driving skill components defined by Horswill et al. (2004), including: adapting driving to changing conditions, knowledge of highway code, controlled emergency stops, awareness and anticipation of other road user behaviors, maintaining appropriate speed for conditions, monitoring junctions and bends, merging with traffic and changing lanes, and maintaining appropriate following distance to a lead vehicle.

2.3. Independent Variables

Three controlled manipulations were conducted in the present study, including the level of vehicle control automation, duration of automation exposure, and the level of automation transparency. All variables were manipulated in a single experiment.

2.3.1. Level of Automation

For the present study, three levels of automation were simulated. Level 0, or manual driving, involved no automation technologies and the participants had full control of the simulated vehicle at all times. Second, the Level 1 type of automation (ACC) maintained a constant headway to the leading vehicle and maintained constant longitudinal speed when the lead vehicle was absent. In this study, the ACC was set at 105km/hr. (approximately 65mph) and the headway distance was set at 160ft, which is a frequently used headway setting by ACC users (Eichelberger & McCartt, 2016). The drivers were able to take over the control of vehicle by depressing the brake pedal. Third, the Level 2 type of automation (ACC+LKA) provided a lane-keeping function in addition to ACC functions. Drivers were freed of both longitudinal and lateral control of the vehicle. They were able to take control of the vehicle by depressing the brake pedal or rotating the steering wheel.

For all the levels of automation, drivers were required to monitor the system and environment at all times and take over vehicle control at hazard events. The level of vehicle automation was manipulated as a between-subjects variable, meaning participants was divided into three groups, with each group experiencing one level of automation.

2.3.2. Automation transparency

Automation transparency was also manipulated as a between-subjects factor with two levels, including a baseline and high-level automation transparency. Among the participants who drove under automated vehicle control (Levels 1 and 2 automation), half of them experienced the higher level of automation transparency, and the other half experienced baseline condition. High automation transparency provided drivers with information of vehicle status through audio messages. The message content included vehicle's actions as well as reasoning for actions, as such a combination is considered most effective for improving driver safety performance (Koo et al., 2015). Examples of audio messages include: "the car is decelerating due to slow traffic" and "the car is turning right due to a curve in the road". All messages were presented at the time of action. Table 3 presents specific events for which messaging will occur. The messages were triggered by roadway events. In general, 5-6 messages were triggered per trial under Level 1 automation, and approximately 8-9 messages were presented under Level 2 automation due to the use of additional automation functionalities. The other half of participants experienced baseline condition. They did not receive audio information on system status.

Table 3: Specific Events for Vehicle Status Messages.

Event	***Message***
Speed is lower than 65mph	*The car is speeding up to preset speed*
Slow traffic	*The car the braking due to slow traffic*
Following leading vehicles	*The car is maintaining the current speed to keep a safe headway*
Obstacles (if participant drive off the road)	*The car is braking due to obstacle ahead*
Curves	*The car is turning due to curve*

2.3.3. Duration of exposure

The duration of automation exposure was manipulated as a within-subjects variable. Each participant experienced 10 experiment trials, and two hazards occurred throughout the experiments: one in the fifth driving trial, and one in the tenth driving trial. Figure 8 graphically presents the timing of experiment trials and hazard events. Consequently, each participant experienced two durations of automation exposure prior to encountering a roadway hazard. The first level of exposure included 5 driving trials, which was approximately 50 minutes of driving, in total. The second level of exposure included 10 driving trials, taking approximately 100 minutes. Trial 5 and 10 took approximately 12 minutes as a result of a 2-minute manual driving segment being added following each hazard negotiation. The test trial duration was the same under all driving conditions.

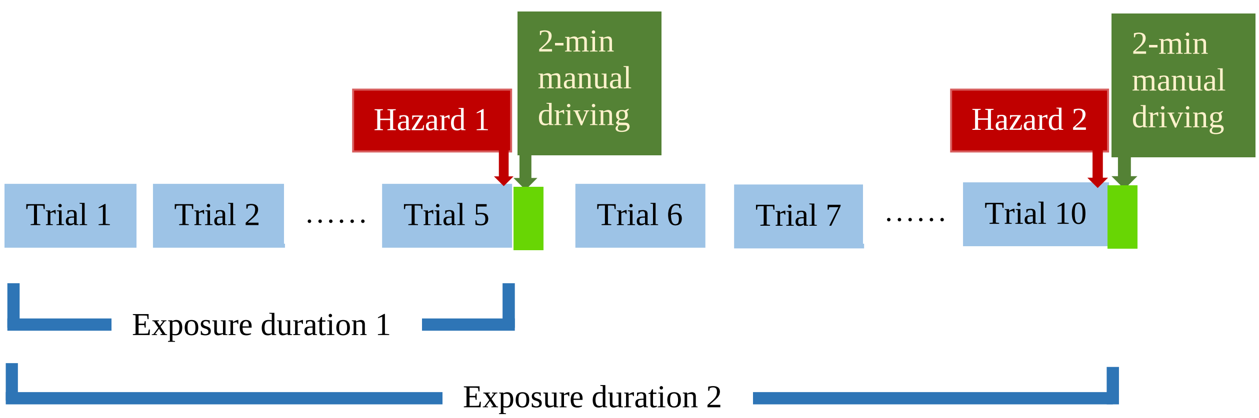

Figure 8. Duration of Exposure.

2.4. Dependent Variables

There were two types of response measures captured in this study, including driving performance under non-critical conditions and hazard negotiation performance.

2.4.1. Driving performance under non-critical condition

Lane deviation and speed deviation data was collected to assess driving performance under non-critical safety conditions. Lane deviation was defined as the average absolute deviation of the center of the vehicle from the center of an intended driving lane during an observation period. The measure has been extensively used to assess driver vehicle control performance (Mok et al., 2015; Zahabi et al., 2017, Ma & Kaber, 2005). In a recent study of vehicle automation and driver performance, Mok et al. (2015) used lane deviation to assess driver performance after taking over a fully automated vehicle. In their study, lane deviation was measured during a 2-minute manual driving period immediately following a transition of vehicle control. Naujoks et al. (2014) used a similar method in their study to evaluate driver vehicle control performance. A similar approach was adopted in the present study. The observation period for assessing lane deviation was a 2-minute segment of road following the completion of hazard negotiation, where drivers took over the control of the vehicle.

Similar to lane deviation, speed deviation is also a commonly used measure to assess driver vehicle control performance (Salvucci & Macuga, 2002; Zahabi et al., 2017; Merat et al., 2014). Merat et al. (2014) used this measure to evaluate driver performance after taking over control from an automated vehicle. In this study, speed deviation was measured during the same observation period as lane deviation. The speed deviation value was calculated as the average absolute deviation of actual vehicle speed from the posted speed limit.

2.4.2. Hazard Negotiation Performance

Hazard negotiation performance measures included hazard reaction time, time to collision, and the number of collisions during a trial. Hazard reaction time has been shown to be an effective indicator of driver hazard negotiation performance (e.g., Gold et al., 2013; Rudin-

Brown et al., 2003; Nilson el al., 2013). In the present study, hazard reaction time was recorded from when the vehicle posing a hazard initiates an offensive movement until the time at which a participant driver initiated a conscious hazard avoidance maneuver. The criteria for identifying a conscious hazard avoidance maneuver will include a driver depressing the simulator brake pedal (Louw et al., 2017). If drivers did not press the brake pedal before ownship crashed into the lead vehicle, the hazard reaction time was recorded from when the vehicle posing a hazard initiated an offensive movement until the time at which the vehicles crashed.

The second hazard negotiation performance measure was the time to collision, which was calculated as the distance gap between own ship and a lead vehicle divided by the relative speed when participant driver initiated a conscious hazard avoidance maneuver (Louw et al., 2017). This measure was used to evaluate the criticality of the hazard event at the point at which drivers begin their hazard avoidance maneuver. If drivers did not press the simulator brake pedal before ownship crashed into the lead vehicle, the time to collision would be zero.

The collision avoidance outcome (i.e., whether the driver can successfully avoid a collision) was also recorded. The number of vehicle collisions under each level of automation and level of automation transparency was determined for the two hazard presentations.

2.5. Participants

Thirty volunteers were recruited to participate in this study. All drivers had valid driver’s license and 20/20 vision (either natural or with corrective lenses) at the time of experiment participation. All participants were middle-aged drivers (30-45 years). This specific age group was selected because they are more likely to use vehicle automation technologies as compared with other age groups (Zmud & Sener, 2017). To further ensure the consistency of participant capability, all drivers had at least 5 years of driving experience, but had no prior experience with any form of vehicle automation technology.

All participants were recruited online or by flyers posted around the North Carolina State University campus. Earlier driving simulation studies on the effect of automation on driving skills serve as a basis for estimating a proper experiment sample size. Table 4 presents the sample sizes from prior studies.

Table 4: Number of Participants Recruited in Prior Studies.

Reference	***Sample Size***
Gold et al. (2013)	*32*
Rudin Brown et al. (2003)	*18*
Merat et al. (2012)	*50*
Schleicher, S., & Gelau (2011)	*22*
Ma & Kaber (2005)	*18*
Young & Stanton (1998)	*62*
Stanton et al. (1997)	*12*

The average sample size of prior studies (rounded to the nearest number) was determined to be 30, balanced by gender (15 males and 15 females). Based on the results of the prior studies, this average sample size is considered to have the potential to yield statistical sensitivity of the response measures to the controlled experiment manipulations.

Six drivers participated under each combination of vehicle automation and automation transparency (see Table 5). Participants were assigned to these conditions randomly. It should be noted that the elevated level of automation transparency only applies to Level 1 and 2 vehicle automation. Therefore, under manual driving all drivers received the baseline level of automation transparency.

Table 5. Number of Participants and Average Age for each Experiment Group.

Automation Transparency / Level of Automation	Baseline	High
0 (Manual)	6 participants, average age=37	N/A
1 (ACC)	6 participants, average age=38	6 participants, average age=35.33
2 (ACC+LKA)	6 participants, average age=39	6 participants, average age=36

2.6. Experiment Design

To attain the objectives identified in Section 1.4, the experiment included two phases. Since the high automation transparency level does not apply to manual driving, the first phase of the experiment only evaluated the baseline automation transparency condition. The primary aim of this phase was to discover possible driver skill decay caused by vehicle automation. The second phase of the experiment included manipulation of automation transparency, and the goal was to evaluate the effectiveness of using an increased level of automation transparency as a remedy for potential driving skill decay. Data collected from these two phases of the experiment were analyzed with two different statistical models.

2.6.1. First Phase

The first phase of the experiment followed a 3*2 mixed design. The duration of automation exposure represented a within-subjects variable, with two levels: a short exposure period (5 trials) and a prolonged exposure period (10 trials). Level of automation represented a between-subject variable, with three levels: Level 0 (manual driving), Level 1(ACC) and Level 2 (ACC+LKA). Table 6 presents the experiment design for the first phase of the experiment.

Table 6. Experiment Design (Phase 1).

Duration of Exposure / Level of Automation	Short	Long
Level 0 Automation		
Level 1 Automation		
Level 2 Automation		

Consequently, the following statistical model was applied to the response measure data as part from the first phase:

Equation 1: Statistical analysis model (first phase of experiment)

$$Y= \mu+V_i+P(V)_{j(i)} +S_k+VS_{ik}+\varepsilon_{ijk}$$

where μ= Grand mean; V = Level of automation (i=1,2,3); P = Participant (j=1, …,18); S = Duration of exposure (k=1,2); VS_{ik}= interaction effect of the level of automation and duration of exposure; and ε= error term. The model allows for comparison of driver performance between various levels of automation, as well as comparison between two different durations of automation exposure.

2.6.2. Second Phase

This phase of the experiment followed a 2*2*2 between-within subject design. The duration of automation exposure remained as a within-subjects factor. The level of vehicle automation in the second phase of the experiment only included those levels to which the automation transparency manipulation was applicable, specifically: Level 1 (ACC) and Level 2 (ACC+LKA). The automation transparency manipulation included two levels: baseline condition (no additional information) and high-level automation transparency (with supporting audio information). Both the level of automation and automation transparency represented between-subject variables. Table 7 presents the experiment data collection table layout for the second phase.

Table 7. Experiment Design (Phase 2).

Automation Transparency / Level of Automation	Baseline		High	
	Short duration of exposure	Long duration of exposure	Short duration of exposure	Long duration of exposure
Level 1 Automation				
Level 2 Automation				

Equation 2: Split plot statistical analysis model (second phase of experiment)

$$Y = \mu + V_i + P(VQ)_{j(il)} + Q_l + S_k + VQ_{il} + VS_{ik} + QS_{lk} + \varepsilon_{ijkl}$$

where μ= Grand mean; V = Level of vehicle automation (i=1,2); $P(VQ)_{j(il)}$ = Participant (j=1, …,24); Q = Automation transparency (l=1,2); S = Duration of automation exposure (k=1,2); VQ_{il}= interaction effect of level of automation and automation transparency; VS_{ik}= interaction effect level of automation and duration of exposure; QS_{lk} = interaction effect of automation transparency and duration of exposure; and ε= error term.

2.7. Procedures

2.7.1. Consent and Demographic Questionnaire

Upon arrival at the driving simulator lab, participants were presented with an informed consent form. Once they agreed to participate in the study, they were asked to complete a brief demographic questionnaire (see Appendix A). The questionnaire called for information about their age, gender, visual acuity, driving history, and experience with vehicle automation.

2.7.2. Simulator Sickness Questionnaire

After completion of the demographic questionnaire, participants were asked to complete a simulator sickness questionnaire (SSQ; Kennedy et al., 1993). The SSQ was used to monitor potential motion sickness symptoms during the experiment (see Appendix B). The initial survey was used to capture any baseline symptoms of nausea, disorientation or ocular-motor disturbance. The questionnaire was repeatedly administered during the course of experiment, including after every two test trials, in order to identify any substantial deviations from baseline symptoms. In case a participant presented simulator sickness symptoms (as identified based on specific criteria from Kennedy et al. (1993)), a 20-minute break was provided. If the symptoms persisted, his or her participation would be terminated, and the participant would be compensated for any time provided. In the present study, two drivers exhibited simulation sickness symptoms during the training session and quit the study. These participants were ultimately replaced in the experiment sample.

2.7.3. Training

Before beginning experiment trials, participants were introduced to the driving simulator controls, including the dashboard, steering wheel, and foot pedals. Participants assigned to automated driving conditions would also receive training on how to switch between driving modes (automated and manual control). Following this introduction, all participants completed a training session in which they became familiar with use of the simulator controls. The training session consisted of driving on a rural highway similar to that to be presented in the actual experiment. The purpose of the training scenario was to ensure that participants are capable of controlling the driving simulator. The training scenario required participants to perform vehicle maneuvers and to drive at the posted speed limit.

The training included two sessions. The first session ensured drivers were familiar with the simulator controls and assessed whether participant speed control and lane maintenance met established criteria. All participants were required to maintain 1 mph or less speed deviation and 1.37ft or less lane deviation, on average (Horrey & Wickens, 2006). If the above criteria were not satisfied, the participant repeated the training scenario. If participant training performance remained unacceptable after three trials, his or her participation in the experiment would be terminated. Driver performance data from the last training trial was recorded for all participants to serve as the performance baseline. This data was also subsequently analyzed to determine whether participants under different experiment conditions had comparable initial driving skills.

In the second training session, participants learned to use the driving systems to which they were assigned (Level 0, 1 or 2) and to negotiate hazards. This training scenario was a rural highway similar to that presented in the actual experiment. A hazard event was also included in the training scenario. The hazard event was similar to but slightly different from those presented in the actual experiment trials.

2.7.4. Fatigue Questionnaire

To minimize participant fatigue and reduce its effect on experiment results, a fatigue questionnaire was administered prior to any test trials and between every trial. The questionnaire contained a 14-question fatigue assessment scale developed by Chalder et al. (1993). The scale assessed both physical and mental fatigue symptoms (see appendix C). Since developed, this scale has been widely used in both medical and research settings (Shahid et al., 2011) and its validity is supported by prior research (Morris et al., 1998). The authors established that a fatigue score of 3 or less represents an absence of fatigue for a participant. In the present study, participants who reported a score of 3 or less would be given a 5-minute break between trials.

Participants who reported a higher fatigue score would be provided with an extra 10-minute break and the fatigue scale was administered again after the break. If the fatigue score dropped below 3, the experiment continued. If fatigue symptoms persisted for the participant, the experiment would be terminated.

2.7.5. Experiment testing

Once the training session is completed, and a participant felt comfortable driving, the Pupil lab eye tracker was calibrated. Instructions for the experiment trials was given to participants. When ready, participants started driving tasks, as described in Section 2.2. Between trials, participants would take a break and the fatigue questionnaire would be administrated. At the completion of experiment trials, participants who experienced high-level automation transparency would complete a post-trial usability questionnaire asking about their experience with the audio information they were provided with during driving. At last they filled out a payment form and were escorted out of the lab.

2.8. Statistical Data Analysis

Prior to data analysis, any outliers in response measures, due to participant failure to follow experiment instructions or equipment issues, were removed from the experiment data sets. This action was necessary to ensure that the response measures reflected actual driver performance.

Along with the fatigue scale, pupil diameter was planned to be used to identify high workload and fatigue symptoms in the present study. In total, three 2.5-minute sampling periods were identified for pupil size observations. The first one took place prior to the start of any test trials and would serve as a baseline. The second sampling period took place during Trial 5, prior to the hazard negotiation. The third sampling period took place prior to the hazard negotiation in

Trial 10. In some previous studies, researchers associated smaller pupil diameter with fatigue symptom (Chi & Lin, 1998; Geacintov& Peavler, 1974; LeDuc et al., 2005), but since there are substantial inter-individual differences in pupil size, no standard criteria have been proposed for identifying fatigue symptoms. In the present study, the planned analysis included comparing samples collected in the second and third pupil size sampling period with the baseline condition pupil size using two sample *t*-tests. However, because pupil diameter data was determined to be unreliable due to data quality and noise factors, such method was not applied to identifying potential fatigue symptoms or screening data.

Training performance was used as a baseline to determine whether drivers assigned to the five experiment conditions (6 participants under each condition) had similar driving skills prior to the start of test trials. Data collected during the training sessions included lane deviation and speed deviation. An ANOVA procedure was used to identify any differences in baseline driving performance among participant groups. If differences existed, the data sets would be further examined for any extreme values and outliers would be excluded.

With respect to the experiment trial data, outliers and abnormal response data due to equipment problems or failure of participants to follow experiment instructions correctly were initially be identified by application of Cook's D Method. These observations were removed prior to any data analysis.

With respect to the crash outcome data, the binary (crash/no crash) response was analyzed with a logistic regression procedure and a likelihood ratio test was conducted to identify any significant effects of each independent variable. All other response measures were initially be subjected to Bartlett's and Shapiro-Wilk's tests to ensure that the data met parametric test assumptions of residual normality and homoscedasticity. If the above-mentioned assumptions were not met, transformations would be applied to the responses. A MANOVA procedure was exercised but results revealed limited to no correlation among the dependent variables. In the present study, the sample size may not have been sufficient for reliability of the MANOVA. According to VanVoorhis & Morgan (2007), the number of responses (*n*) in each cell should be greater than the number of dependent variables. Since the participant term was included as an independent variable in the analysis, the *n* in each cell was two (each participant experienced two hazards and post-hazard manual driving), but the number of dependent variables was four. Therefore, the detailed results of the MANOVA analysis were not considered reliable and were not included in the present report. This data analysis was followed by an ANOVA procedure. In the event of ANOVA assumption violations for transformed responses, the measures were ranked and submitted to the parametric test to yield the equivalent of a nonparametric analysis.

2.9. Crash Prediction Model

Based on the data collected from the present study, the following logistic regression model was applied to predict the likelihood of crashes. It is worth noting that the data analysis tool used for this research (JMP) predicted the likelihood of “not crash”, by default.

$$P\ (i)=1/(1+e^{b0+b1Xi1+b2Xi2+......})$$

where P (i)= probability of crash, b_0= intercept, $b_{1...}$ b_n are coefficients for each predictor (n is the number of predictors, which were determined based on the best fitting model).

Potential predictors included: X_{i1}: Driving mode (including five levels: manual driving, Level 1 automation with baseline transparency, Level 2 automation with baseline transparency, Level 1 automation with high-level transparency, Level 2 automation with high-level transparency). X_{i2}: Duration of automation exposure (including two levels: short and extended).

Since only a small number of predictors are available for the modeling effort, an all-subset approach was used for further selection of the predictor set. In total, 4 combinations were identified (including interaction effects). The all-subset approach compared all possible models with model selection based on r-square values and the Bayesian information criterion (BIC). The R-square value is also referred to as McFadden's pseudo R-square.

The R-square value was calculated as:

$$r^2 = 1 - \log[L(M_{Candidate})]/\log[L(M_{Intercept})]$$

where $\log[L(M_{Candidate})]$ represent the log likelihood of candidate model, and $\log[L(M_{Intercept})]$ represents the log likelihood of reduced model without predictors and only the intercept present. This ratio suggests the level of improvement provided by the candidate model over the intercept model.

While adding parameters might increase the r-square value, the approach can also result in overfitting, meaning the model might fit too closely or exactly the set of available data, but fail to fit additional data or predict future observations with reliability (Babyak et al., 2004). To potentially avoid overfitting issues, the BIC value was considered as well. BIC resolves the overfitting problem by introducing a penalty term for the number of parameters in the model. Models with a lower BIC are preferred. The BIC value was calculated as:

$$BIC = -2\log[L(M_{Candidate})] + k\log(n)$$

where $\log[L(M_{Candidate})]$= the loglikelihood of the candidate model, k = the number of estimated parameters in the candidate mode, n= the number of data points (Fabozzi et al., 2014).

It is likely that models with low BIC values had low r-square values as well. Therefore, to find a model maximizing predictive utility while limiting overfitting, the following procedure was applied based on previous literature: first, the minimum BIC value among all models was identified, noted as BIC_{min}; second, candidate models with $BIC_{candidate}$ -BIC_{min}<6 were identified. This criterion was identified by earlier research (Fabozzi et al., 2014). The value is used to provide some level of confidence or assurance of the absence of model overfitting. This step was motivated by the fact that the magnitude of the difference between the $BIC_{candidate}$ and BIC_{min} values can be interpreted as evidence against a candidate model being the best model. As such, the evidence against a candidate model is weak if the difference between these terms is less than 6. Finally, among the selected candidate models, the one with the highest r-square value would be identified as the best model.

3. RESULTS

3.1. Data Screening

Driver performance data from the last training trial were recorded and analyzed to determine whether participants under different experiment conditions had comparable initial driving skills. Training data from all participants were separated into two datasets according to the two phases of experimentation, and the data was submitted to ANOVA procedure. Both datasets met ANOVA normality and equal variance assumptions.

According to the training data analysis, participants in Phase 1 exhibited comparable speed deviation performance ($F(2,17) = 0.0484$, $p=0.9529$, $1-\beta= 0.0560$) and lane deviation ($F(2,17) = 0.26$, $p=0.774$, $1-\beta= 0.083$) across all driving conditions. Similar results were observed with participants in Phase 2. Their speed deviation was comparable across automation levels ($F(1,23) = 1.1143$, $p=0.3031$, $1-\beta= 0.1720$) and automation transparency ($F(1,23) = 0.4227$, $p=0.5223$, $1-\beta= 0.0952$). Their lane deviation was comparable across automation levels ($F(1,23) = 1.1143$, $p=0.3031$, $1-\beta= 0.1720$) and automation transparency ($F(1,23) = 1.1143$, $p=0.3031$, $1-\beta= 0.1720$) as well. Therefore, no outliers were identified with this procedure.

The fatigue scale questionnaire, comprised of 14-items, indicated that none of the participants experienced perceived fatigue greater than “3” prior to any experiment trials.

As previously mentioned, the pupil size analysis was intended to provide a manipulation check on use of the fatigue questionnaires for allocating driver breaks. However, scrutinization of the pupil size data led to concerns regarding reliability. First, the manufacturer default setting for the Pupil Labs device was reported to filter all data with a confidence level greater than 0.5. However, the resulting experiment data provided evidence that such a filter is applied not applied by default and that many pupil size observations with confidence less than 0.5 are recorded. Even with application of this filter, concerns over data quality remained. Figure 9 presents a sample of baseline trial data from one participant. The figure shows pupil diameter values for the driver during a 2.5-minute period with a confidence filter of 0.5 or greater. The presence of substantial noise is obvious from the plot. In fact, a further examination of the relevant literature revealed that pupil diameter can be substantially affected by factors, including emotional state (Geangu et al., 2011; Kawai et a., 2013), lumination (Hopkinson, 1956; Pfleging et al., 2016, Laeng et al., 2012), and stress (Yamanaka et al., 2009). These factors can vary largely in driving task performance.

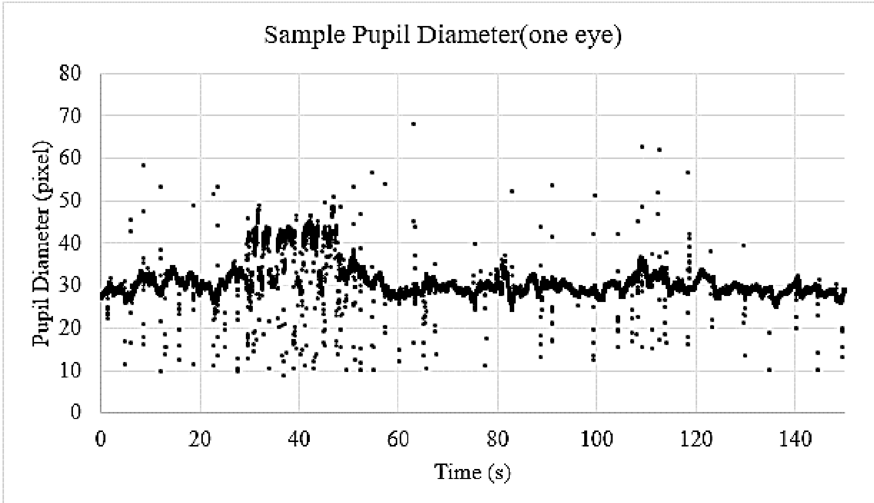

Figure 9. Sample Pupil Diameter Data.

Yet another analysis of sample pupil size data revealed further concerns with reliability. Figure 10 presents pupil diameter data from another participant, plotted against the confidence level of each data point with a minimum confidence filter of 0.5 being applied. It appeared that two “notches” were present in the pupil diameter values (see “red” arrows) with pupil diameters of ~23 and ~33 units being absent from the dataset. Given these outcomes from the data collection with the Pupil Labs device, the pupil diameter data was determined to be unreliable and this potential measure of fatigue symptoms was disregarded.

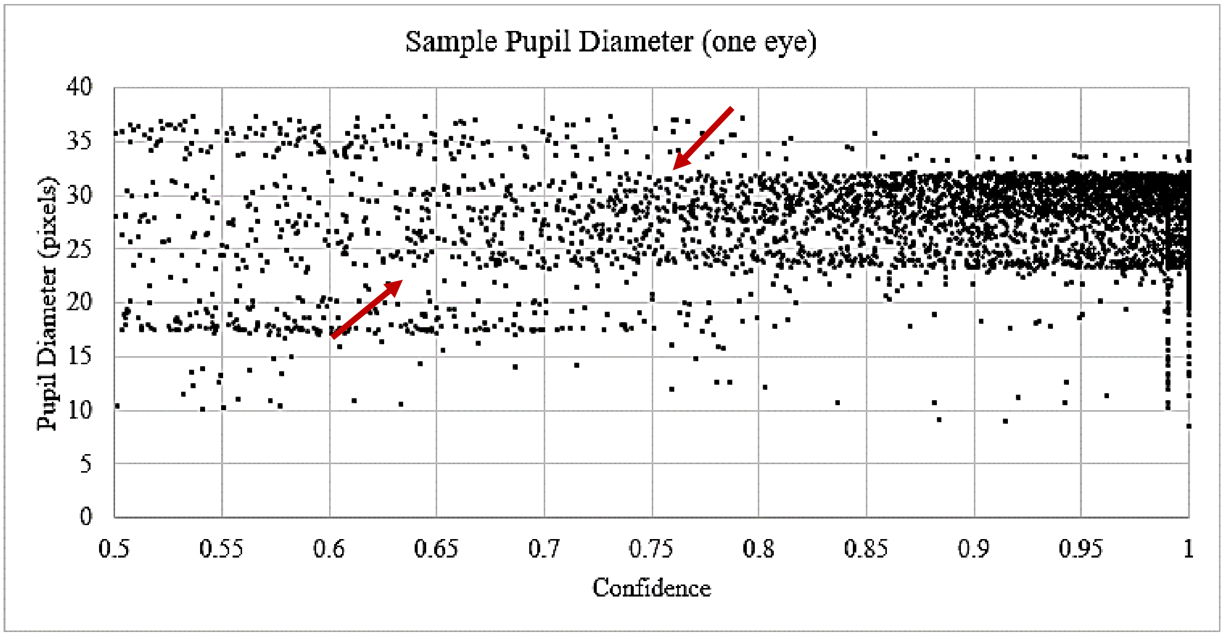

Figure 10. Sample Pupil Diameter Data Plotted against Confidence Level

3.2. Phase 1 Data Analysis Results

3.2.1. Post-Hazard Manual Driving Performance

Driver speed deviation data was subjected to a square root transformation to meet the normality assumptions. The main effect of levels of automation was found to be highly significant (F (2,35) =25.1760, p<0.0001, 1-β=1.000). However, the main effect of duration of exposure (F (1,35) =0.3523, p=0.5623, 1-β= 0.0858) and its interaction with driving mode (F (2, 35) =1.14, p=0.3478, 1-β=0.2108) were not significant. Post-hoc Tukey's tests revealed that Level 2 automation produced degraded speed maintenance as compared with manual control and Level 1 automation (Figure 9).

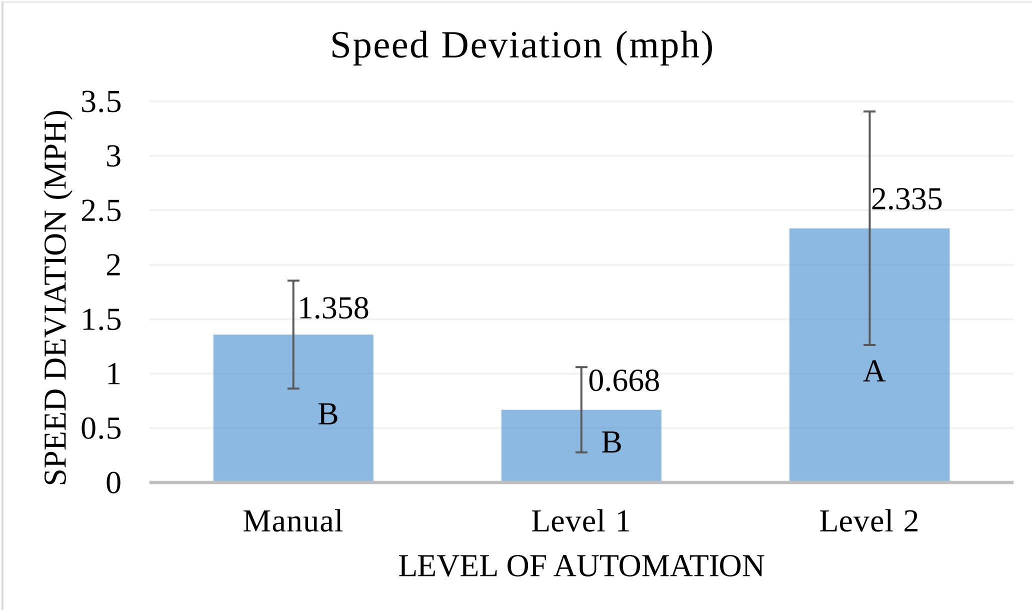

Figure 9. Effect of Level of Automation on Speed Deviation.

Driver lane deviation data violated the normality assumptions and was subjected to a rank transformation. The main effect of levels of automation was found to be significant ($F(2,35) = 14.1894$, $p=0.0003$, $1-\beta=0.9934$). However, the main effect of duration of exposure ($F(1,35) = 0.2636$, $p=0.6151$, $1-\beta=0.0769$) and its interaction with driving mode ($F(2, 35) = 1.7484$, $p=0.2077$, $1-\beta=0.1248$) were not significant. Post-hoc Tukey's tests revealed that Level 2 automation produced degraded lane maintenance performance as compared with manual control and Level 1 automation (Figure 10).

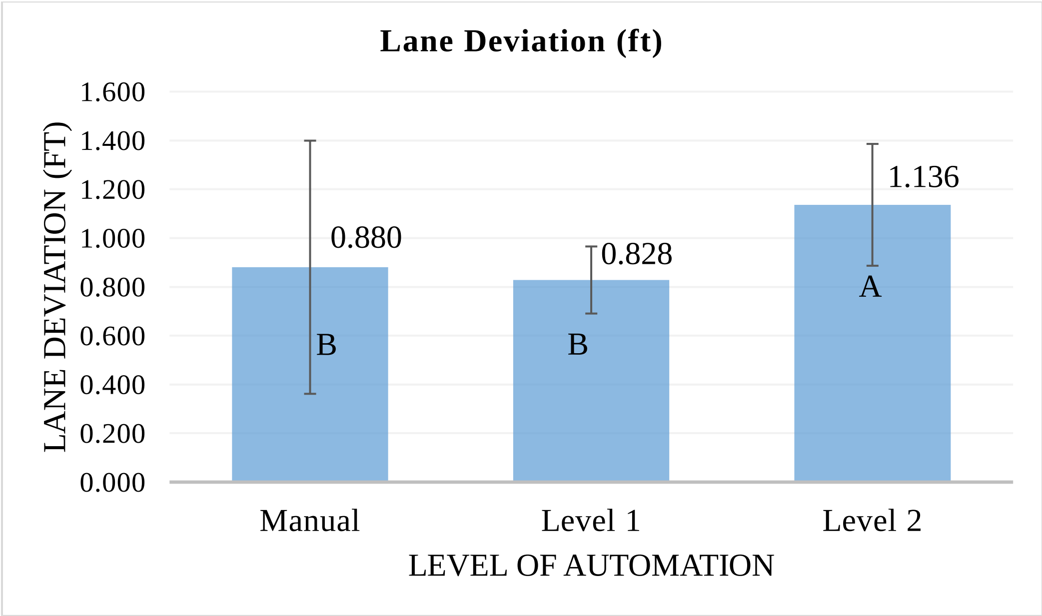

Figure 10. Effect of Level of Automation on Lane Deviation.

3.2.2. Hazard Reaction Performance

Hazard reaction time data was subjected to a rank transformation to meet the ANOVA assumptions. Rank transformed and untransformed data produced similar results; thus, results from the untransformed data are presented here. Level of vehicle automation had a significant effect on driver hazard reaction time ($F(2,34) = 6.2917$, $p=0.0104$, $1-\beta=0.8245$). Post-hoc Tukey's tests revealed that manual driving produced faster hazard reaction than Level 1 and Level 2 automation (Figure 11). No significant effect was found from duration of exposure ($F(1,34) = 0.7488$, $p=0.4005$, $1-\beta=0.0512$) and its interaction with level of automation ($F(2,34) = 0.0100$. $p=0.99$, $1-\beta=0.1353$). For visual interpretation, a graph of hazard reaction time across driving modes is presented in Figure 11. Error bars in the graph represent +/- 1 standard deviation of the response measure. The labels A/B represent significance groupings.

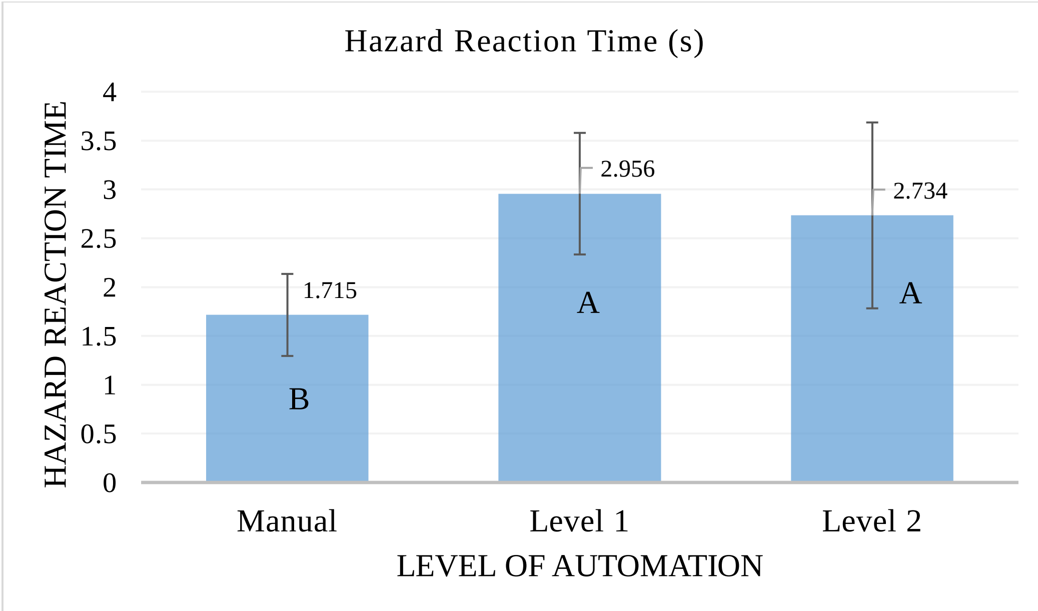

Figure 11. Effect of Level of Automation on Hazard Reaction Time.

Time-to-collision data met ANOVA assumptions. Level of vehicle automation ($F(2,34) = 6.0241$, $p=0.0130$, $1-\beta=0.8006$) was significant in effect. Drivers under manual driving produced significantly longer time-to-collision than Level 1 automation. Duration of exposure ($F(1,34) = 0.3248$, $p=0.5778$, $1-\beta=0.130$) and its interaction with level of automation ($F(2,34) = 2.3340$, $p=0.1324$, $1-\beta=0.3957$) were not significant. The effects are presented in Figure 12.

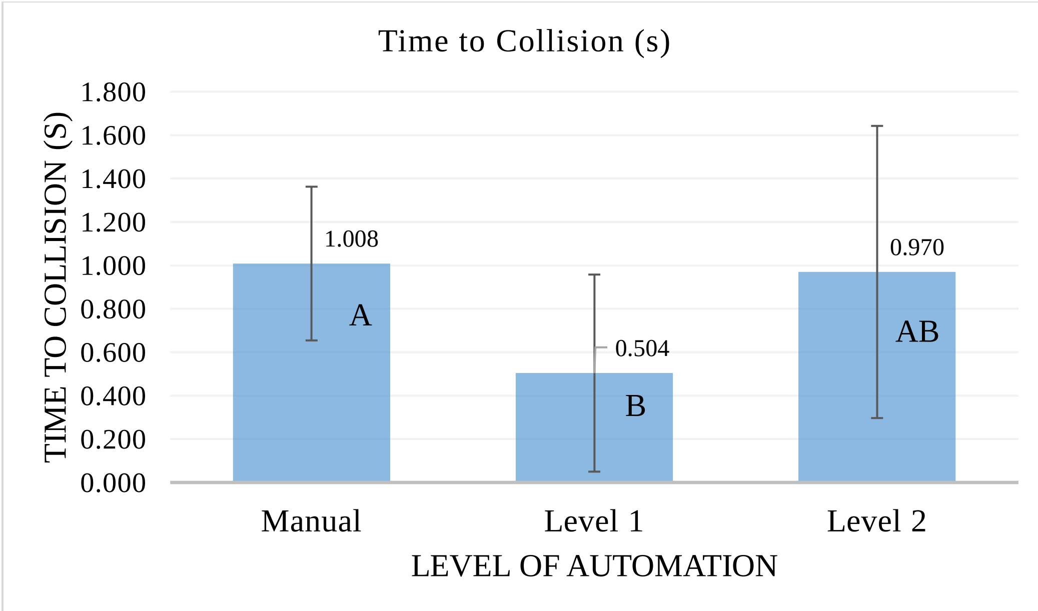

Figure 12. Effect of Level of Automation on Time to Collision.

3.3. Phase 2 Data Analysis Results

3.3.1. Post-Hazard Manual Driving Performance

Driver speed deviation data was subjected to log transformation to meet ANOVA assumptions. As presented in Figure 13, Level 2 automation led to higher speed deviation than Level 1 ($F(1,47) = 24.7709$, $p < 0.0001$, $1-\beta > 0.999$). No significant effect was identified from automation transparency ($F(1,47) = 0.1465$, $p = 0.7064$, $1-\beta = 0.100$) or duration of exposure ($F(1,47) = 0.0135$, $p = 0.9089$, $1-\beta = 0.0685$).

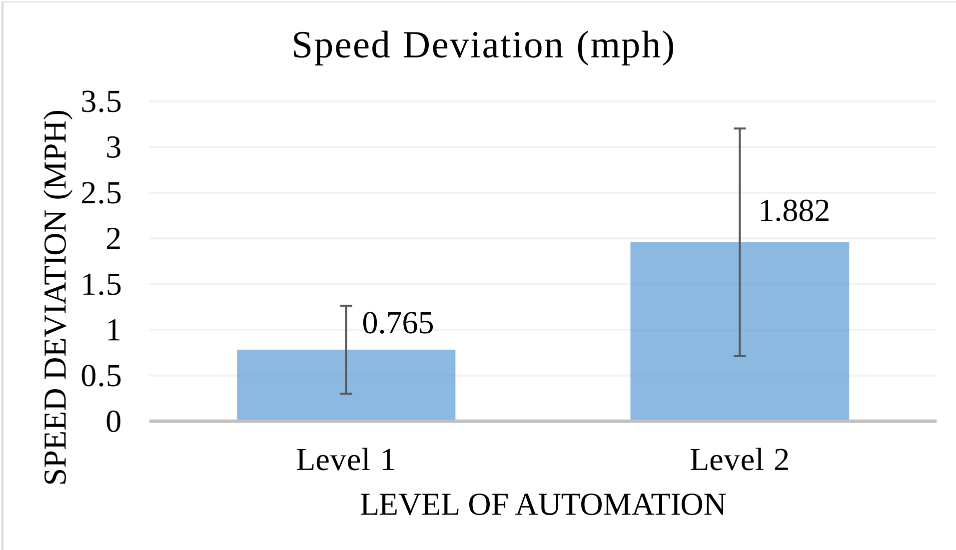

Figure 13. Effect of Level of Automation on Speed Deviation.

The interaction effect of automation transparency and level of automation was also significant (F (1,47) =6.9817, p=0.0192, 1-β=0.6576). Higher level of automation produced degraded speed control but such effect was mitigated when audio messages were present. The interaction effect is presented in Figure 14.

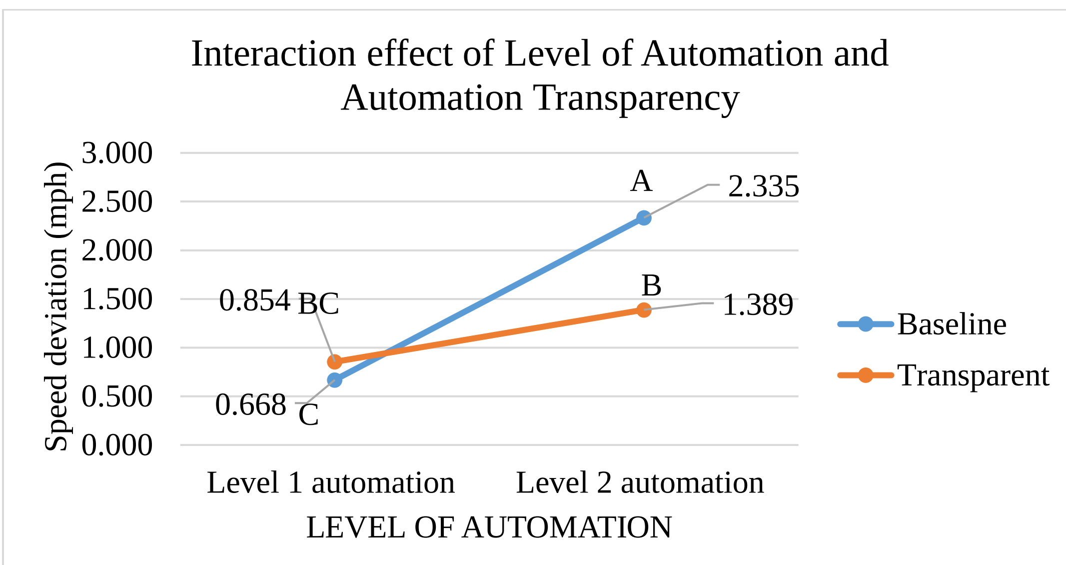

Figure 14. Interaction Effect on Speed Deviation: Level of Automation and Automation Transparency.

Driver lane deviation data did not meet ANOVA assumptions and a square-root transformation was applied. The effect of automation transparency ($F(1,47) = 0.0372$, $p = 0.8493$, $1-\beta = 0.0503$), level of automation ($F(1,47) = 0.3590$, $p = 0.5565$, $1-\beta = 0.1427$) and duration of exposure ($F(1,47) = 0.0052$, $p = 0.943$, $1-\beta = 0.0505$) were not significant. Only the interaction between automation transparency and level of automation ($F(1,47) = 23.6452$, $p = 0.0005$, $1-\beta = 0.9810$) was significant. Automation transparency produced lower lane deviations with level 2 automation but the effect of automation transparency was reverse with level 1 automation. Figure 15 presents this interaction effect.

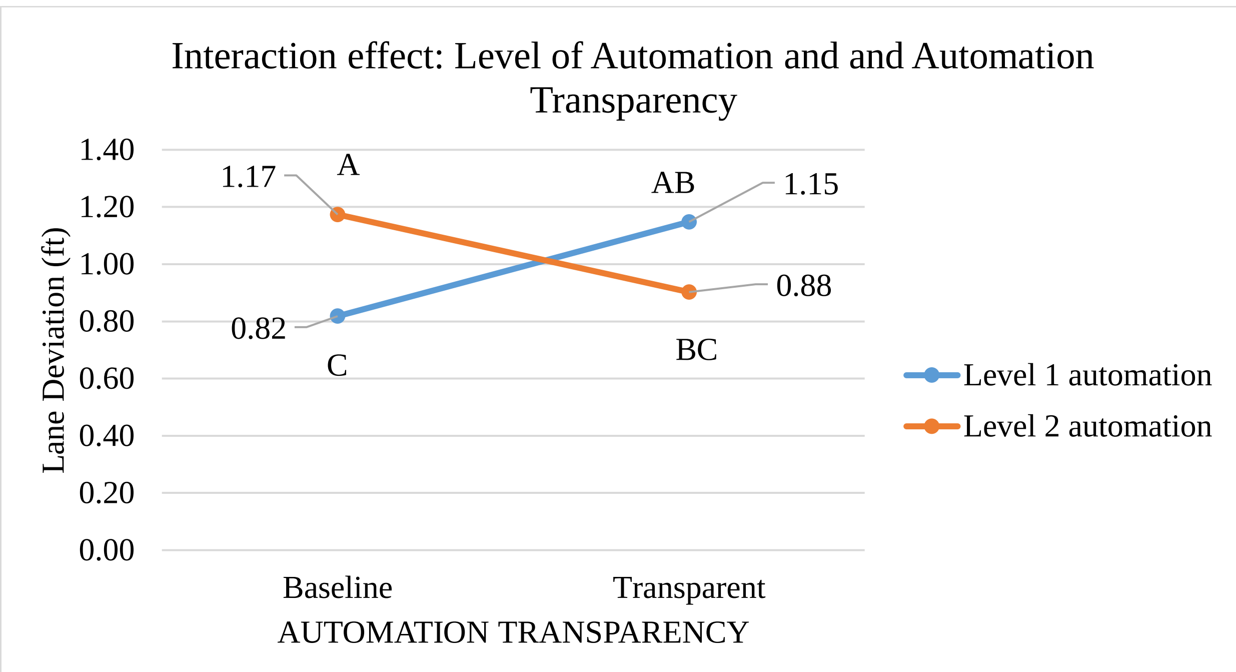

Figure 14. Interaction Effect on Lane Deviation: Duration of Exposure and Automation Transparency.

3.3.2. Hazard Reaction Performance

Hazard reaction time data was subjected to a rank transformation. Rank transformed and untransformed data produced similar results; therefore, results from the untransformed data analysis are presented. Automation transparency had a significant effect on driver hazard reaction time ($F(1,46) = 7.9262$, $p=0.0110$, $1-\beta=0.7614$). High-level automation transparency produced faster hazard reaction than the baseline condition (Fig 15). No significant effects of level of automation ($F(1,46) = 0.3227$, $p=0.5767$, $1-\beta=0.050$) or duration of exposure ($F(1,46) = 0.1382$, $p=0.7142$, $1-\beta=0.0533$) were observed.

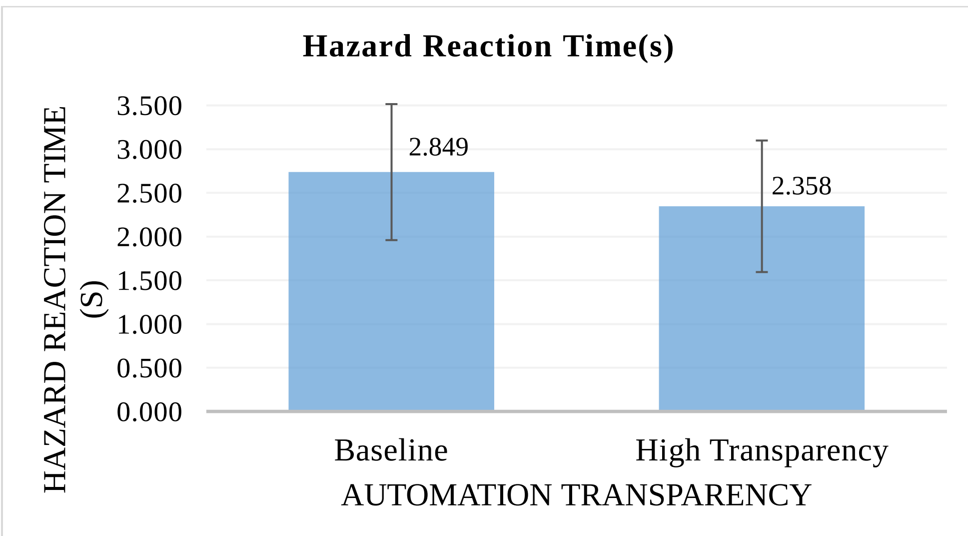

Figure 15. Effect of Automation Transparency on Hazard Reaction Time.

Time-to-collision data did not meet ANOVA assumptions and a square root transformation was applied. Automation transparency ($F(1,46) = 7.2628$, $p=0.0143$, $1-\beta=0.7247$) had a significant effect on driver hazard reaction time. Higher level of automation transparency produced longer time-to-collision than the baseline condition (Fig 16). No significant effect was found for duration of exposure (F (1,46) =0.0001, p=0.9940, 1-β=0.0500). The level of automation (F (1,46) =4.4129, p=0.0492, 1-β=0.5136) has a marginal effect (Figure 17).

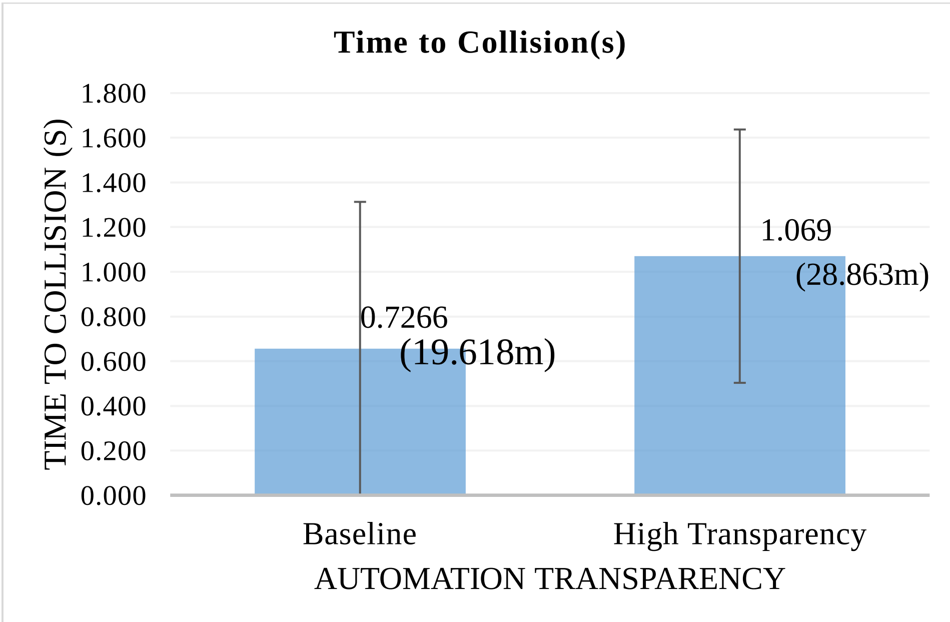

Figure 16. Effect of Automation Transparency on Time to Collision.

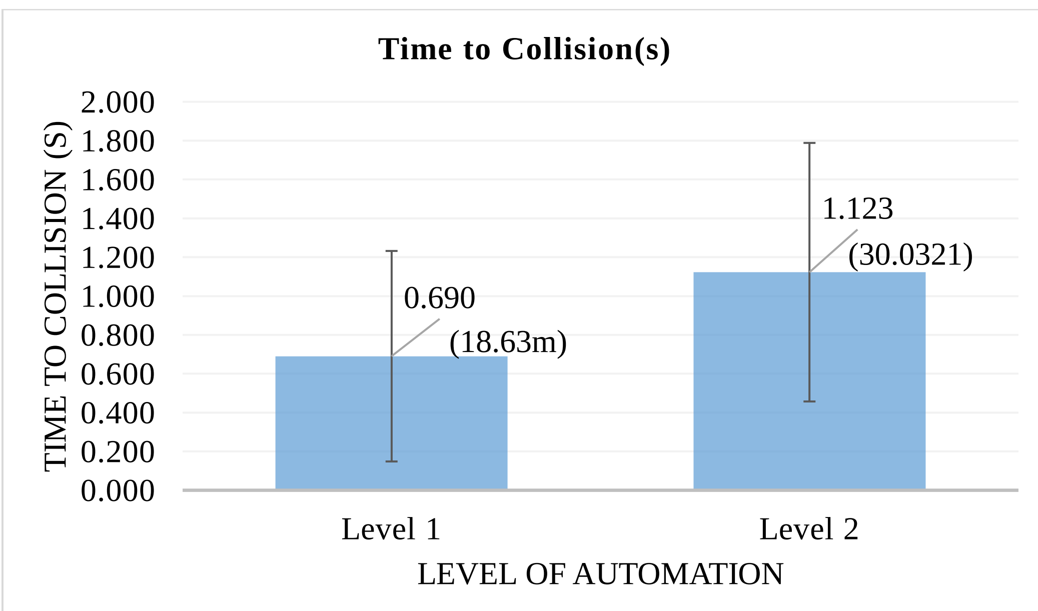

Figure 17. Effect of Level of Automation on Time to Collision.

3.3.3. Post-trial Usability Questionnaire

Participants who drove with high-level automation transparency were asked to complete a usability questionnaire after all test trials. Among these participants, six participants drove with Level 1 automation and six drove with Level 2 automation. The overall score for each question, as well as the average score for Level 1 and Level 2 automation are presented in

Figure 17.

Overall, participants found audio messages helpful. They indicated that the audio messages helped them understand automation system states and stay alert. Results suggested drivers were likely to use such audio messages, if they were provided in vehicles. It is worth noting that drivers under Level 2 automation found the audio messages to be more helpful than Level 1. They were also more likely to use such messages if they were provided in vehicles.

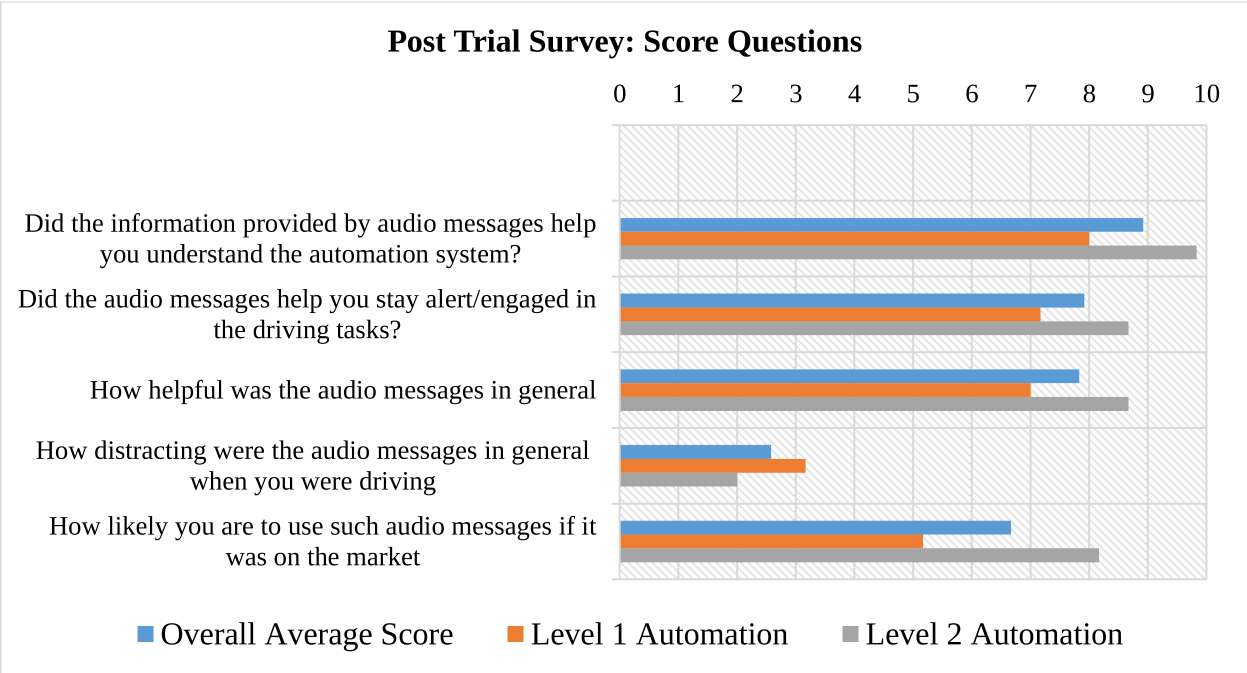

Figure 17. Post-Trial Survey: Score Questions.

In addition to the scores for questions, participants were also asked to respond to open ended questions about their recommendations or concerns. In most cases, drivers didn't have additional suggestions or concerns. All comments from participants are presented in Table 8.

Table 8. Post-Trial Usability Questionnaire: Open ended Questions.

No.	*Question*	*Level 1 Automation Response*	*Level 2 Automation Response*
6	Is there any additional information you would find helpful?	*None*	*"Numeric speed information"* *"Need to know if vehicle has stopped"*
7	Do you have any suggestions/recommendations you would like to share?	*"Audio should sound more like human"*	*"Need more human-like audio"*
8	Do you have any concerns you would like to share?	*"Audio message was helpful in the beginning but annoying beyond that"*	*None*

3.4. Crash Outcome Prediction Model

3.4.1. Model Selection

Following an "all-subset" data analysis approach, all combinations of predictors were tested. All four combinations are presented in Table 9.

Table 9. Summary table: Subsets of predictors with no significant effect.

No.	*Combinations of predictors*	*Significant effect*	*R square*	*BIC*
1	Driving Mode	*None*	*0.082*	*92.81*
2	Duration of Exposure	*None*	*0.019*	*85.53*
3	Driving mode + Duration of Exposure	*None*	*0.102*	*95.36*
4	Driving mode+ Duration of Exposure + Driving mode* Duration of Exposure	*None*	*0.143*	*108.42*

3.4.1. Best Fitting Model

On the basis of the above analysis, Combination No. 2 had the lowest BIC value (81.7377). Therefore, the BICmin was identified as 85.53. Among the models with BICcandidate – BICmin < 6, combination No.2 also had the highest r-square value among the four candidate models. However, the r-square value was low across all candidate models, including the selected model. The model is presented below.

Equation 4: Best fitting crash outcome prediction model

$$P(\text{crash}) = 1/(1+e^{\gamma}))$$

$$\gamma = \beta_0 + \beta_{1i} X_{1i}$$

where X_{1i} represents duration of exposure dummy variable, i=1,2: i=1 (short duration), j=2 (long duration, baseline effect). β_0 = intercept. β_{1i}= duration of exposure estimate parameter (i=1,2).

Parameter estimates for all baseline effects (manual driving and long duration of exposure) were zero. Table 11 reports parameter estimates for all predictors except for baseline effects (parameter estimates=0). It is worth noting that neither predictor had a significant effect on crash outcomes.

Table 11. Parameter estimates for the best fitting model.

Predictor	*Notation*	*Estimate coefficient*
Intercept	β_0	*0.0862*
Duration of exposure (short)	β_{11}	*0.2299*

4. DISCUSSION

4.1. Phase 1

Returning to the hypotheses posed at the beginning of this study, it was expected that higher levels of automation would produce higher speed deviations (H1) and higher lane deviations (H2). These hypotheses were partially supported by experiment results. Under Level 2 automation, drivers exhibited deteriorated lane and speed maintenance, but manual driving and Level 1 automation produced comparable performance. It appeared that the skill decay caused by vehicle automation was only manifested with higher level of automation where more vehicle functions were automated. This is possibly because drivers under ACC still needed to control the lateral position of the vehicle, and the lateral control task might have actively engaged drivers in observing vehicle movement and vehicle status. In this case, under Level 2 automation where both lateral and longitudinal control tasks were automated, drivers' perceptual-motor skills significantly degraded, but this effect was not as obvious as under ACC.

Longer exposure to automation was also hypothesized to produce higher speed deviations (H3) and higher lane deviations (H4). However, study results revealed comparable driving performance with long and short durations of exposure, thus these hypotheses were not supported. In the present study, the short duration of exposure to automation was around 50 minutes, and the long duration of exposure was around 100 minutes. In the aviation domain, McClumpha et al. (1991) found the greater the number of air training hours, the greater pilot support for the statement that their skills were eroded by automation. However, the minimum exposure time for drivers to experience skill decay has rarely been investigated. It is suspected that longer exposure time (> 100 min) is necessary for drivers to experience skill decay in lane and speed control tasks. In addition, in the present study, to reduce the effect of fatigue on driver performance, exposure to automation was interrupted by breaks between trials. The continuity of exposure might also be an influential factor in driving skill decay.

With respect to hazard negotiation performance, higher levels of vehicle automation were hypothesized to lead to longer hazard reaction times (H7), which was partially supported by study findings. Level 1 and Level 2 automation resulted in comparable hazard reaction times, which were significantly longer than manual driving reaction times. Higher levels of vehicle automation were also expected to produce shorter times to collision (H8), or in other words, riskier hazard avoidance performance. This hypothesis was partially supported. The time-to-collision for manual driving was longest among the three groups (least risky), but Level 1 and Level 2 automation produced comparable results. It is worth noting that the driving scenario in the present study was a two-lane freeway with smooth curves, thus the lateral vehicle control task was not strenuous for drivers. Therefore, although the lane maintenance task was automated, the difference between the two levels of automation might have been relatively simple. The experiment findings demonstrated the presence of the “out of the loop” performance problem with automated driving, as compared to manual driving, which was also observed in several previous studies (Rudin Brown et al., 2003; Gold et al., 2013; Schleicher & Gelau, 2011). It was argued that drivers under automated driving experienced distraction and inattention, and the attentional resources freed-up by automation were not used for hazard detection but “for other purposes” (Rudin Brown et al., 2003; Gold et al., 2013). On the contrary, our previous study (Deng & Kaber, 2018) observed superior hazard reaction performance with ACC, as compared to manual driving. In that study, similar to the present one, drivers were informed of the possibility of hazard occurrence, but the time to exposure was much shorter (the first hazard occurred after around 5 minutes of driving). It is likely that the longer exposure time to automation in the present study resulted in different observations than those by Deng and Kaber (2018). However, in the present research, drivers had comparable hazard reaction time and time-to-collision after short (50 min) and extended (100 min) exposure to automation, which differed from expectations (H9 and H10). It is possible that drivers experienced rapid skill decay within a short time of exposure (< 50min), but the changes became smoother over time. Since few prior studies have focused on this topic, further investigation is needed to provide a more complete mapping of driver skills and exposure time.

In general, the findings from the Phase 1 experiment demonstrated skill decay issues with automated driving, specifically, degraded hazard negotiation performance and post-hazard manual driving performance. Results suggested a need for effective methods to address such issues and mitigate the negative effects of vehicle automation.

4.2. Phase 2

Phase 2 introduced automation transparency into the experiment design. High-level automation transparency was expected to lead to lower speed deviations (H5). However, the main effect of automation transparency was not significant. Interestingly, results revealed a significant interaction effect of level of automation and automation transparency. High-level automation transparency reduced driver speed deviations but only with Level 2 automation. It was also hypothesized that a higher level of automation transparency would produce lower lane deviation (H6), which was not observed in the present study. Results suggested a significant interaction effect of levels of automation and automation transparency. Under Level 2 automation, high-level automation transparency produced lower lane deviations than the baseline condition. The effect was reversed with Level 1 automation. To further understand these observation, drivers' attitudes towards automation transparency were investigated through the post-trial questionnaire. According to the survey results, both drivers under Level 1 and Level 2 automation considered audio messages useful, but when asked "how likely you are to use such audio messages if available on the market?", drivers under Level 1 automation gave a rate of 5.17, while drivers under Level 2 automation gave a rate of 8.17. It is possible that participants, who drove with ACC, found the audio messages redundant and "annoying", because Level 1 automation involved simpler automated functions. In fact, one participant who drove with ACC mentioned that the messages were helpful at the beginning of the experiment but "annoying" beyond that. Therefore, it seems likely that the complexity of vehicle automation and drivers' attitudes could have a significant influence on the effects of automation transparency. Although automation transparency had positive effects with Level 2 automation, the effect was not shown with Level 1 due to simple automation and negative driver attitudes.

The use of high-level automation transparency to enhance hazard negotiation performance was supported by study results. Findings provided strong evidence that high-level automation transparency led to shorter hazard reaction times (H11) and greater time-to-collision (H12). These findings are in line with observations by Koo et al. (2015), who found that introducing higher levels of automation transparency produced the safest driving performance. Naujoks et al. (2015) also argued that communicating upcoming automated maneuvers by speech could lead to more attentive driving. It is possible that superior performance under high level automation transparency was due to enhanced driver understanding of vehicle functions through audio messages, or the fact that audio messages kept drivers alert, which was further demonstrated through the post-trial questionnaires. In the present research, audio messages might have actively maintained driver attention on monitoring vehicle actions and states, thus producing superior hazard negotiation performance.

In addition, in Phase 2, driver speed deviation was found to be higher with Level 2 automation than Level 1, as expected (H1). The expectation that higher level of automation could produce higher lane deviations (H2), however, was not supported. Comparable lane maintenance performance was observed under the two levels of automation. Therefore, results did provide some evidence of post-automation exposure driving skill decay with higher level automation, but this evidence was not as strong as expected. Moreover, no significant difference was detected in driver hazard reaction time between the two levels of automation. This may again be due to the ease of lateral control tasks and the simple simulated roadway conditions. Drivers under Level 2 automation exhibited higher time-to-collision but this effect was marginal and the statistical power was very low.

Similar to Phase 1, the duration of automation exposure was not influential on any response measures in the Phase 2 experiment. Therefore, Hypotheses 3, 4, 9 and 10 were not supported. These findings could again be attributed to the relatively short and discontinuous exposure of drivers to automation during the experiment.

Post-trial surveys revealed drivers' attitudes towards automation transparency. In general, drivers found higher level automation transparency to be helpful. Drivers suggested that audio messages enhanced their understanding of vehicle behavior and helped them stay alert. It is interesting to note that drivers under Level 2 automation found the messages more helpful than Level 1, and they were also more likely to use such technology if it were on the market. Since the Level 1 automation function was less complicated, drivers might have found find the automation system easier to understand, thus the audio information became redundant and "annoying". In fact, Koo et al. (2015) also found drivers under Level 1 automation to have negative emotional responses when they were presented with both "how" and "why" information about automated vehicle behavior, although the information proved to improve driving performance. According to earlier studies, such information might have increased driver workload (Chen et al., 2014), thus making drivers more “anxious” (Koo et al.,2015). Based on these findings, further investigation is necessary to determine an optimal amount of information or most appropriate time at which to present such information to users. Through the questionnaire, drivers also requested more “human like” audio, and information on vehicle speed, which may pose in-vehicle messaging system design implications for manufacturers.

4.3. Crash Outcome Prediction Model

The best crash outcome prediction model had a low *r-square* value (0.0193), which indicated a low fit to the data set. Therefore, the model could not reliably predict crash outcomes or be applied to adjust vehicle settings. Two major factors might have contributed to this result. First, the low fit might be due to unaccounted predictors. In the present model, predicting factors included duration of exposure. However, crash avoidance might be a complex outcome of reaction capability, motivation, and hazard reaction strategy. Related to this, factors like health status, occupation, or personality might be significantly influential in driver crash avoidance performance as well. In fact, previous research has suggested factors including emotion (Dingus et al., 2016), employment stability (Cantor et al., 2010), personality, etc. (Clarke& Robertson, 2005) could all have significant impact on the likelihood of a crash. Therefore, predictors in the present study might not be sufficient to precisely predict crash outcomes, and the model could likely be improved with additional driver information.

Second, it is important to note that the current study had a relatively small sample size. Given the scope of the study, only 30 participants were included. However, a rule of thumb suggests around 10 events per variable. For example, if a sample of 60 driving events is studied and 20 events do not involve a crash, such that 40 crashes occurred, the rule of thumb implies that two pre-specified predictors can reliably be fitted to the total dataset (Peduzzi, P., et al., 1996; Concato, J. et al., 1995). In this study, among 59 hazard avoidance events, 37 events involved crashes and in 22 events crashes were avoided. In this case, based on the rule of thumb, around 2 predictors could be reliably fitted with the dataset. However, the all subset approach analyzed two predictors and their interactions. Therefore, these results should be interpreted with caution.

Although driving mode and automation transparency level were reported to affect driver reaction time and time-to-collision, their effects on crash outcomes were not significant. These findings are likely due to the fact that other factors, including driver crash avoidance strategy, motivations, etc., might affect crash avoidance outcomes, which were not measured in the present study.

5. LIMITATIONS AND FUTURE DIRECTIONS

Due to the scope of this study, only drivers aged between 30 and 45 years were included in the study. Drivers from other age groups (e.g., elderly or younger drivers) might exhibit different adaption behaviors (braking, steering) when vehicle automation is active. Their attitudes towards automation transparency features may also differ. Therefore, it is necessary for future research to include drivers from different age groups.

Moreover, the pupil diameter data collected in the present study was determined to be unreliable and could not accurately predict driver physiological status. Future investigations may consider manipulating experiment settings to mitigate noise factors, or use blink rate, blink duration, or other physiological measures, such as heart rate in a complementary manner to pupil size data with the objective of accurately revealing driver physiological status. Pupil diameter data may also be considered in crash outcome prediction models by future investigations, but appropriate collection process to reduce noise level or complementary physiological measures are recommended.

In addition, the roadway condition presented in the current study was a rural freeway with clear weather, and only two types of hazard events were included. Real-life roadway conditions are significantly more complicated, and current automated driving systems are highly susceptible to such complexity (e.g., rain, unclear lane markers, pedestrians, etc.). In order to obtain further insights into driver behavior with automated driving and achieve more accurate crash prediction analysis, it is recommended for future studies that a large variety of environment conditions and hazard scenarios be investigated.

Moreover, the duration of exposure to automated driving was relatively short in the present study. The exposure was also interrupted by rest breaks between test trials to mitigate fatigue effect. In real-life cases, it is common for drivers to have longer and continuous exposure to automation. Sleep deprivation and intoxication would also be expected to lead to greater impact of automation use on performance. Future studies should also consider investigating a larger variety of exposure durations, in order to identify potential skill decay associated with each level of automation.

Further investigation into the utilization of automation transparency is also necessary. For example, additional visual messaging could be provided with similar content as audio messages. This method may ensure reliability of messaging but could lead to additional driver visual distraction. It would also be interesting to assess the impact of transparency on driver trust in automation. Future investigations could ensure similar levels of engagement across levels of vehicle automation and assess trust outcomes using subjective ratings scales (Bagheri & Jamieson, 2004).

Beyond these limitations and recommendations, with respect to the crash outcome prediction model, there might be additional influential predictors that were not considered in the present study. For example, it would be worthwhile to explore the effect of factors such as driver health status, personalities, or occupation, etc. Future studies might also benefit from a larger sample size. In the present study, the participant sample size was 30. A larger sample might be helpful to develop a more reliable crash outcome prediction model.

6. CONCLUSION

In conclusion, this study utilized driving simulation experiments to investigate driver behavior changes related to the use of vehicle automation and high-level automation transparency. The study made comparison of all currently available forms of vehicle automation (Level 0, Level 1 and Level 2) to provide a complete mapping of the effect of vehicle automation on driver performance under both critical and non-safety critical driving conditions. Findings provided some evidence of skill decay associated with automated driving, including degraded hazard negotiation performance and post-automation exposure manual driving performance. The study also investigated use of higher level of automation transparency to improve driver performance with automated driving, which proved to be a highly effective method. Participants who drove with high-level automation transparency (i.e., they were provided with audio information about vehicle automation system states) produced superior hazard negotiation performance. Automation transparency also appeared to enhance speed control and lane maintenance under Level 2 automation. Post-trial surveys revealed that drivers, especially the ones using Level 2 automation, found automation transparency to be most useful and they were likely to use such a method if it were available in the market. The current research is one of the first driving simulation studies on effects of automation transparency on driver performance. This investigation can provide a basis for enhancing the design of information systems in automated vehicles.

Finally, a novel crash outcome prediction model was developed, using vehicle settings as predictors. Limited by the scope of the study, the current research did not yield a highly fitting model for crash prediction. Therefore, this topic should be revisited in future investigations with larger sample sizes and broader sets of experimental conditions. The present study, however, may provide a useful methodological reference for future research that endeavors to develop a comprehensive and accurate real-time crash prediction model.

REFERENCES

Adanu, E. K., Smith, R., Powell, L., & Jones, S. (2017). Multilevel analysis of the role of human factors in regional disparities in crash outcomes. *Accident Analysis & Prevention, 109*, 10-17.

Babyak, M. A. (2004). What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic medicine*, *66*(3), 411-421.

Bagheri, N., & Jamieson, G. A. (2004). Considering subjective trust and monitoring behavior in assessing automation-induced "complacency.". Human performance, situation awareness, and automation: Current research and trends, 54-59.

Ball, K., Owsley, C., Sloane, M. E., Roenker, D. L., & Bruni, J. R. (1993). Visual attention problems as a predictor of vehicle crashes in older drivers. *Investigative ophthalmology & visual science*, *34*(11), 3110-3123.

Beggiato, Matthias, et al. "What would drivers like to know during automated driving? Information needs at different levels of automation." *7. Tagung Fahrerassistenzsysteme*. 2015.

Bennett, C. A., Marcellus, F. S., & Reynolds, J. F. (1974, October). Counteracting psychological fatigue effects by stimulus change. In *Proceedings of the Human Factors Society Annual Meeting* (Vol. 18, No. 2, pp. 219-224). Sage CA: Los Angeles, CA: Sage Publications.

Blanco, M., Atwood, J., Russell, S., Trimble, T., McClafferty, J., & Perez, M. (2016). *Automated vehicle crash rate comparison using naturalistic data*. Virginia Tech Transportation Institute.

Boudette, N. E. (2018, March 31). Fatal Tesla Crash Raises New Questions About Autopilot System. Retrieved May 5, 2018, from https://www.nytimes.com/2018/03/31/business/tesla-crash-autopilot-musk.html.

Bainbridge, L. (1983). Ironies of automation. In *Analysis, Design and Evaluation of Man–Machine Systems 1982* (pp. 129-135).

Cantor, D. E., Corsi, T. M., Grimm, C. M., & Özpolat, K. (2010). A driver focused truck crash prediction model. *Transportation Research Part E: Logistics and Transportation Review, 46*(5), 683-692.

Chalder, T., Berelowitz, G., Pawlikowska, T., Watts, L., Wessely, S., Wright, D., & Wallace, E. P. (1993). Development of a fatigue scale. Journal of psychosomatic research, 37(2), 147-153.

Chen, J. Y., Procci, K., Boyce, M., Wright, J., Garcia, A., & Barnes, M. (2014). *Situation awareness-based agent transparency* (No. ARL-TR-6905). ARMY RESEARCH LAB ABERDEEN PROVING GROUND MD HUMAN RESEARCH AND ENGINEERING DIRECTORATE.

Chi, C. F., & Lin, F. T. (1998). A comparison of seven visual fatigue assessment techniques in three data-acquisition VDT tasks. *Human Factors, 40*(4), 577-590.

Clarke, S., & T Robertson, I. (2005). A meta-analytic review of the Big Five personality factors and accident involvement in occupational and non-occupational settings. *Journal of Occupational and Organizational Psychology, 78*(3), 355-376.

Concato, J., Peduzzi, P., Holford, T. R., & Feinstein, A. R. (1995). Importance of events per independent variable in proportional hazards analysis I. Background, goals, and general strategy. *Journal of clinical epidemiology*, *48*(12), 1495-1501.

Deng, Y., & Kaber, D. (2018, September). Effect of Vehicle Control Format on Driver Performance and Attention Allocation under Adaptive Cruise Control. *In Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 62, No. 1, pp. 1510-1514). Sage CA: Los Angeles, CA: SAGE Publications.

De Waard, D., van der Hulst, M., Hoedemaeker, M., Brookhuis, K.A., 1999. Driver behaviour in an emergency situation in the automated highway system. *Transportation Human Factors* 1 (1), 67e82.

De Winter, J. C. F., & Dodou, D. (2010). The Driver Behaviour Questionnaire as a predictor of accidents: A meta-analysis. *Journal of safety research*, *41*(6), 463-470.

Deery, H. A. (1999). Hazard and risk perception among young novice drivers. *Journal of safety research*, *30*(4), 225-236.

Deery, H.A., & Love, A.W. (1996). The effect of a moderate dose of alcohol on the traffic hazard perception profile of young drivers. *Addiction*, 91, 815–827.

Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences*, *113*(10), 2636-2641.

Dikmen, M., & Burns, C. M. (2016, October). Autonomous driving in the real world: Experiences with tesla autopilot and summon. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 225-228). ACM.

Duncan, J., Williams, P., & Brown, I. (1991). Components of driving skill: experience does not mean expertise. *Ergonomics*, 34(7), 919-937.

Eichelberger, A. H., & McCartt, A. T. (2016). Toyota drivers' experiences with dynamic radar cruise control, pre-collision system, and lane-keeping assist. *Journal of safety research*, 56, 67-73.

Endsley, M. R. (2017). Autonomous driving systems: A preliminary naturalistic study of the Tesla model S. *Journal of Cognitive Engineering and Decision Making*, *11*(3), 225-238.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human factors*, 37(2), 381-394.

Evans, L. (1991). Traffic safety and the driver. Science Serving Society.

Fabozzi, F. J., Focardi, S. M., Rachev, S. T., & Arshanapalli, B. G. (2014). *The basics of financial econometrics: Tools, concepts, and asset management applications*. John Wiley & Sons.

Flemisch, F., Kelsch, J., Löper, C., Schieben, A., Schindler, J., & Heesen, M. (2008). Cooperative control and active interfaces for vehicle assistance and automation.

Fuller, R. (2005). Towards a general theory of driver behaviour. *Accident Analysis & Prevention*, 37(3), 461-472.

Geacintov, T., & Peavler, W. S. (1974). Pupillography in industrial fatigue assessment. *Journal of Applied Psychology*, *59*(2), 213.

Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). "Take over!" How long does it take to get the driver back into the loop?. In Proceedings of the *Human Factors and Ergonomics Society Annual Meeting* (Vol. 57, No. 1, pp. 1938-1942). SAGE Publications.

Guo, F., & Fang, Y. (2013). Individual driver risk assessment using naturalistic driving data. *Accident Analysis & Prevention*, 61, 3-9.

Grove, K., Atwood, J., Hill, P., Fitch, G., DiFonzo, A., Marchese, M., & Blanco, M. (2015). Commercial motor vehicle driver performance with adaptive cruise control in adverse weather. *Procedia Manufacturing*, *3*, 2777-2783.

Habib, K. (2017). *DOT NHTSA ODI Document* (Rep.). NHTSA. doi: https://static.nhtsa.gov/odi/inv/2016/INCLA-PE16007-7876.pdf

Harrell, F.E., 2001. Regression Modeling Strategies: With Applications to LinearModels, Logistic Regression, and Survival Analysis. Springer Science and Business. Media, Inc., New York, NY.

Hoedemaeker, M., & Brookhuis, K. A. (1998). Behavioural adaptation to driving with an adaptive cruise control (ACC). *Transportation Research Part F: Traffic Psychology and Behaviour*, *1*(2), 95-106.

Hopkinson, R. G. (1956). Glare discomfort and pupil diameter. *JOSA*, *46*(8), 649-656.

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

Horswill, M. S., Waylen, A. E., & Tofield, M. I. (2004). Drivers' ratings of different components of their own driving skill: A greater illusion of superiority for skills that relate to accident involvement. *Journal of Applied Social Psychology*, *34*(1), 177-195.

Jamson, A. H., Merat, N., Carsten, O. M., & Lai, F. C. (2013). Behavioural changes in drivers experiencing highly-automated vehicle control in varying traffic conditions. *Transportation research part C: emerging technologies*, *30*, 116-125.

Ji, Q., Zhu, Z., & Lan, P. (2004). Real-time nonintrusive monitoring and prediction of driver fatigue. *IEEE transactions on vehicular technology*, *53*(4), 1052-1068.

Jo, J., Lee, S. J., Park, K. R., Kim, I. J., & Kim, J. (2014). Detecting driver drowsiness using feature-level fusion and user-specific classification. *Expert Systems with Applications*, *41*(4), 1139-1152.

Kassner, M., Patera, W., & Bulling, A. (2014, September). Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: Adjunct publication* (pp. 1151-1160). ACM.

Kawai, S., Takano, H., & Nakamura, K. (2013, October). Pupil diameter variation in positive and negative emotions with visual stimulus. In *2013 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 4179-4183). IEEE.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology*, 3(3), 203-220.

Koo, J., Kwac, J., Ju, W., Steinert, M., Leifer, L., & Nass, C. (2015). Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust, and performance. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, *9*(4), 269-275.

Lajunen, T., & Summala, H. (1995). Driving experience, personality, and skill and safety-motive dimensions in drivers' self-assessments. *Personality and Individual Differences*, *19*(3), 307-318.

Laeng, B., & Endestad, T. (2012). Bright illusions reduce the eye's pupil. *Proceedings of the National Academy of Sciences,* 109(6), 2162-2167.

LeDuc, P. A., Greig, J. L., & Dumond, S. L. (2005). Involuntary eye responses as measures of fatigue in US Army Apache aviators. *Aviation, space, and environmental medicine, 76*(7), C86-C91.

Lewin, I. (1982). Driver training: a perceptual-motor skill approach. *Ergonomics, 25*(10), 917-924.

Louw, T., Markkula, G., Boer, E., Madigan, R., Carsten, O., & Merat, N. (2017). Coming back into the loop: Drivers' perceptual-motor performance in critical events after automated driving. *Accident Analysis & Prevention*, 108, 9-18.

Ma, R., & Kaber, D. B. (2005). Situation awareness and workload in driving while using adaptive cruise control and a cell phone. *International Journal of Industrial Ergonomics, 35(10), 939-953.*

McClumpha, A. J., James, M., Green, R. G., & Belyavin, A. J. (1991, September). Pilots' attitudes to cockpit automation. In *Proceedings of the Human Factors Society Annual Meeting* (Vol. 35, No. 2, pp. 107-111). Sage CA: Los Angeles, CA: SAGE Publications.

McKenna, F. P., & Horswill, M. S. (1999). Hazard perception and its relevance for driver licensing. *Journal of the International Association of Traffic and Safety Sciences*, 23, 26-41.

Merat, N., Jamson, A. H., Lai, F. C., & Carsten, O. (2012). Highly automated driving, secondary task performance, and driver state. *Human factors*, 54(5), 762-771.

Mercado, J. E., Rupp, M. A., Chen, J. Y., Barnes, M. J., Barber, D., & Procci, K. (2016). Intelligent agent transparency in human–agent teaming for Multi-UxV management. *Human factors*, *58*(3), 401-415.

Mok, B. K. J., Johns, M., Lee, K. J., Ive, H. P., Miller, D., & Ju, W. (2015, June). Timing of unstructured transitions of control in automated driving. In *Intelligent Vehicles Symposium (IV), 2015 IEEE* (pp. 1167-1172). IEEE.

Morriss, R., Wearden, A., & Mullis, R. (1998). Exploring the validity of the Chalder Fatigue scale in chronic fatigue syndrome. *Journal of psychosomatic research*, 45(5), 411-417.

Naujoks, F., Mai, C., & Neukum, A. (2014). The effect of urgency of take-over requests during highly automated driving under distraction conditions. *Advances in Human Aspects of Transportation*, 7(Part I), 431.

Naujoks, F., Forster, Y., Wiedemann, K., & Neukum, A. (2016). Speech improves human-automation cooperation in automated driving. *Mensch und Computer 2016–Workshopband*.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans, 30*(3), 286-297.

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of clinical epidemiology*, *49*(12), 1373-1379.

Pfleging, B., Fekety, D. K., Schmidt, A., & Kun, A. L. (2016, May). A model relating pupil diameter to mental workload and lighting conditions. *In Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 5776-5788). ACM.

Pelz, D. C., & Krupat, E. (1974). Caution profile and driving record of undergraduate males. *Accident Analysis and Prevention*, 6,45-58.

Rudin-Brown, C. M., Parker, H. A., & Malisia, A. R. (2003). Behavioral adaptation to adaptive cruise control. In P*roceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 47, No. 16, pp. 1850-1854). SAGE Publications.

SAE. (2016). U.S. Department of Transportation's New Policy on Automated Vehicles Adopts SAE International's Levels of Automation for Defining Driving Automation in On-Road Motor Vehicles. Retrieved January 24, 2017, from https://www.sae.org/news/3544/

Shahid, A., Wilkinson, K., Marcu, S., & Shapiro, C. M. (2011). Chalder Fatigue Scale. In *STOP, THAT and One Hundred Other Sleep Scales* (pp. 97-98). Springer, New York, NY.

Schoettle, B., & Sivak, M. (2015). A preliminary analysis of real-world crashes involving self-driving vehicles. *University of Michigan Transportation Research Institute.*

Schleicher, S., & Gelau, C. (2011). The influence of Cruise Control and Adaptive Cruise Control on driving behaviour–A driving simulator study. *Accident Analysis & Prevention*, *43*(3), 1134-1139.

Seong, Y., & Bisantz, A. M. (2008). The impact of cognitive feedback on judgment performance and trust with decision aids. *International Journal of Industrial Ergonomics*, *38*(7-8), 608-625.

Smith, B. W. (2015). *Automated and Autonomous Driving: Regulation Under Uncertainty* (Rep.). International Transport Forum. doi:https://www.itf-oecd.org/sites/default/files/docs/15cpb_autonomousdriving.pdf

Stanton, N. A., Young, M., & McCaulder, B. (1997). Drive-by-wire: the case of driver workload and reclaiming control with adaptive cruise control. *Safety science, 27(2),149-159.*

Stanton, N. A., & Young, M. S. (1998). Vehicle automation and driving performance. *Ergonomics*, *41*(7), 1014-1028.

Stern, J. A., Boyer, D., & Schroeder, D. (1994). Blink rate: a possible measure of fatigue. *Human factors*, *36*(2), 285-297.

Summala, H., Hakkanen, H., Mikkola, T., & Sinkkonen, J. (1999). Task effects on fatigue symptoms in overnight driving. *Ergonomics, 42*(6), 798-806.

Sümer, N. (2003). Personality and behavioral predictors of traffic accidents: testing a contextual mediated model. *Accident Analysis & Prevention, 35*(6), 949-964.

Tesla. (n.d.). Autopilot. Retrieved April 07, 2018, from https://www.tesla.com/autopilot

Thiffault, P., & Bergeron, J. (2003). Monotony of road environment and driver fatigue: a simulator study. *Accident Analysis & Prevention*, 35(3), 381-391.

Transportation Research Board (2010). Highway Capacity Manual (5th Edition).Washington, D.C.: TRB.

Trimble, T. E., Bishop, R., Morgan, J. F., & Blanco, M. (2014, July). *Human factors evaluation of level 2 and level 3 automated driving concepts: Past research, state of automation technology, and emerging system concepts.* (Report No. DOT HS 812 043). Washington, DC: National Highway Traffic Safety Administration.

Viti, F., Hoogendoorn, S.P., Alkim, T.P., and Bootsma, G., 2008. Driving behavior interaction with ACC: results from a field operational test in the Netherlands. In: Proceedings of the IEEE e Intelligent Vehicle Symposium 2008, June 4e6, Eindhoven, the Netherlands.

Volvo. (n.d.). *Pilot Assist.* Retrieved April 07, 2018, from https://support.volvocars.com/uk/cars/Pages/owners-manual.aspx?mc=v526&my= 2016&sw= 15w46&article=548956727ac6edfbc0a80151522a4edc

Wakabayashi, D. (2018, March 23). Uber's Self-Driving Cars Were Struggling Before Arizona Crash. Retrieved from https://www.nytimes.com/2018/03/23/technology/uber-self-driving-cars-arizona.html

Westin, C., Borst, C., & Hilburn, B. (2016). Automation Transparency and Personalized Decision Support: Air Traffic Controller Interaction with a Resolution Advisory System. *IFAC-PapersOnLine*, *49*(19), 201-206.

Wickens, C. D. (1995). Designing for situation awareness and trust in automation. In *Integrated systems engineering* (pp. 365-370).

Wiener, E. L., & Curry, R. E. (1980). Flight-deck automation: Promises and problems. *Ergonomics*, *23*(10), 995-1011.

Wu, Q., Chen, F., Zhang, G., Liu, X. C., Wang, H., & Bogus, S. M. (2014). Mixed logit model-based driver injury severity investigations in single-and multi-vehicle crashes on rural two-lane highways. *Accident Analysis & Prevention*, 72, 105-115.

Yamanaka, K., & Kawakami, M. (2009). Convenient evaluation of mental stress with pupil diameter. *International journal of occupational safety and ergonomics*, 15(4), 447-450.

Yang, X. J., Unhelkar, V. V., Li, K., & Shah, J. A. (2017, March). Evaluating Effects of User Experience and System Transparency on Trust in Automation. In *HRI* (pp. 408-416).

Young, M. S. & Stanton, N. A. (1997). Automotive automation: Investigating the impact on drivers' mental workload. *International Journal of Cognitive Ergonomics*, 1(4), 325-336

Young, M. S., & Stanton, N. A. (2007). What's skill got to do with it? Vehicle automation and driver mental workload. *Ergonomics*, 50(8), 1324-1339.

Zahabi, M., Machado, P., Pankok, C., Lau, M. Y., Liao, Y. F., Hummer, J., ... & Kaber, D. B. (2017). The role of driver age in performance and attention allocation effects of roadway sign count, format and familiarity. *Applied ergonomics, 63*, 17-30.

Ziebinski, A., Cupek, R., Grzechca, D., & Chruszczyk, L. (2017, November). Review of advanced driver assistance systems (ADAS). In *AIP Conference Proceedings* (Vol. 1906, No. 1, p. 120002). AIP Publishing.

Zhang, Y., Lin, W. C., & Chin, Y. K. S. (2010). A pattern-recognition approach for driving skill characterization. *IEEE transactions on intelligent transportation systems, 11*(4), 905-916.

Zmud, J. P., & Sener, I. N. (2017). Towards an understanding of the travel behavior impact of autonomous vehicles. *Transportation research procedia, 25*, 2500-2519.

APPENDICES

Appendix A: Demographic Questionnaire (pre-trial)

Please check one box only (for multiple choice questions) unless otherwise indicated.

Section A: Demographic

1. Name (e.g., first last): ________________________________

2. Please write your age in years: ___________________________

3. Please select your gender: Male ☐ Female ☐

4. Please write your current corrected vision (e.g., 20/20):

Left ___________ Right ____________

Section B: Driving Experiences

5. About how many days per week do you drive?

a. 1-2 days per week ☐

b. 3-4 days per week ☐

c. 5-6 days per week ☐

d. Everyday ☐

6. During the last 3 years, how many minor road accidents have you been involved in? (A minor accident is one in which no-one required medical treatment AND any cost of damage to vehicles and property were $1,000 or less).

Number of minor accidents ____ (if none, write 0)

7. During the last 3 years, how many major road accidents have you been involved in? (A major accident is one in which EITHER someone required medical treatment OR costs of damage to vehicles and property were greater than $1,000, or both).

Number of major accidents ____ (if none, write 0)

Appendix B: Simulation Sickness Questionnaire

Post-exposure Simulator Sickness Questionnaire
SYMPTOM CHECKLIST (Post-exposure)
Post-exposure instruction: Circle below if any of the symptoms apply to you now.

1. General discomfort	None	Slight		Moderate		Severe
2. Fatigue	None	Slight		Moderate		Severe
3. Boredom	None	Slight		Moderate		Severe
4. Drowsiness	None	Slight		Moderate		Severe
5. Headache	None	Slight		Moderate		Severe
6. Eyestrain	None	Slight		Moderate		Severe
7. Difficulty focusing	None	Slight		Moderate		Severe
8. Salivation increase	None	Slight		Moderate		Severe
Salivation decrease	None	Slight		Moderate		Severe
(as compared with normal salivation)						
9. Sweating	None	Slight		Moderate		Severe
10. Nausea	None	Slight		Moderate		Severe
11. Difficulty Concentrating	None	Slight		Moderate		Severe
12. Mental depression	No	Yes (	Slight	Moderate	Severe)	
13. "Fullness of the head"	No	Yes (	Slight	Moderate	Severe)	
14. Blurred vision	No	Yes (	Slight	Moderate	Severe)	
15. Dizziness eyes open	No	Yes (	Slight	Moderate	Severe)	
Dizziness eyes close	No	Yes (	Slight	Moderate	Severe)	
16. Vertigo	No	Yes (	Slight	Moderate	Severe)	

17. Visual flashbacks	No	Yes (	Slight	Moderate	Severe	)
18. Faintness	No	Yes (	Slight	Moderate	Severe	)
19. Aware of breathing	No	Yes (	Slight	Moderate	Severe	)
20. Stomach awareness	No	Yes (	Slight	Moderate	Severe	)
21. Loss of appetite	No	Yes (	Slight	Moderate	Severe	)
22. Increased appetite	No	Yes (	Slight	Moderate	Severe	)
23. Desire to move bowels	No	Yes (	Slight	Moderate	Severe	)
24. Confusion	No	Yes (	Slight	Moderate	Severe	)
25. Burping	No	Yes (	Slight	Moderate	Severe	)
26. Vomiting	No	Yes (	Slight	Moderate	Severe	)
27. Other	No	Yes (	Slight	Moderate	Severe	)

Appendix C: 14-item fatigue scale

Physical symptoms

1. Do you have problems with tiredness?

2. Do you need to rest more?

3. Do you feel sleepy or drowsy?

4. Do you have problems starting things?

5. Do you start things without difficulty but get weak as you go on?

6. Are you lacking in energy?

7. Do you have less strength in your muscles?

8. Do you feel weak?

Mental symptoms

9. Do you have difficulty concentrating?

10. Do you have problems thinking clearly?

11. Do you make slips of the tongue when speaking?

12. Do you find it more difficult to find the correct word?

13. How is your memory?

14. Have you lost interest in the things you used to do?

Appendix D: Post Experiment Questionnaire

Here are few questions about audio messages. On a scale of 1 to 10, with 1 being Not at all and 10 being Very, please circle the rating that best describes your opinion:

1. Did the information provided by audio messages help you understand the automation system?

   Not at all Very
   1 2 3 4 5 6 7 8 9 10

2. Did the audio messages help you stay alert/engaged in the driving tasks?

   Not at all Very
   1 2 3 4 5 6 7 8 9 10

3. How helpful was the audio messages in general:

   Not at all Very
   1 2 3 4 5 6 7 8 9 10

4. How distracting were the audio messages in general when you were driving:

   Not at all Very
   1 2 3 4 5 6 7 8 9 10

5. How likely you are to use such audio messages if it was on the market:

   Not at all Very
   1 2 3 4 5 6 7 8 9 10

6. Is there any additional information you would find helpful?

7. Do you have any suggestions/recommendations you would like to share?

8. Do you have any concerns you would like to share?