DRIVER BEHAVIOR AND PERFORMANCE IN AN AGE OF INCREASINGLY INSTRUMENTED VEHICLES

EDITED BY: Oren Musicant, Haneen Farah, David Shinar and Christian Collet
PUBLISHED IN: Frontiers in Psychology

frontiers Research Topics

Frontiers eBook Copyright Statement

ISSN 1664-8714
ISBN 978-2-88971-396-7
DOI 10.3389/978-2-88971-396-7

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

DRIVER BEHAVIOR AND PERFORMANCE IN AN AGE OF INCREASINGLY INSTRUMENTED VEHICLES

Topic Editors:
Oren Musicant, Ariel University, Israel
Haneen Farah, Delft University of Technology,Netherlands
David Shinar, Ben-Gurion University of the Negev, Israel
Christian Collet, Université Claude Bernard Lyon 1, France

Citation: Musicant, O., Farah, H., Shinar, D., Collet, C., eds. (2021). Driver Behavior and Performance in an Age of Increasingly Instrumented Vehicles. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-396-7

Table of Contents

frontiers in Psychology

EDITORIAL
published: 28 July 2021
doi: 10.3389/fpsyg.2021.715239

Editorial: Driver Behavior and Performance in an Age of Increasingly Instrumented Vehicles

Oren Musicant[1]*, *Haneen Farah*[2], *David Shinar*[3] *and Christian Collet*[4]

[1] Department of Industrial Engineering & Management, Ariel University, Ariel, Israel, [2] Department of Transport & Planning, Delft University of Technology, Delft, Netherlands, [3] Department of Industrial Engineering & Management, Ben-Gurion University of the Negev, Beersheba, Israel, [4] Inter-University Laboratory of Human Movement Biology - Mental Processes, Cerebral Plasticity and Motor Performance Team (MP3), Université Claude Bernard Lyon 1, Lyon, France

Keywords: driver behavior, driver performance, advanced driver assistance system, older (elderly) drivers, autonomuos vehicles

Editorial on the Research Topic

Driver Behavior and Performance in an Age of Increasingly Instrumented Vehicles

Driver behavior and performance have been studied extensively in the last decades. Researchers have developed theories and models of driver situational awareness, driver decision-making and errors, information processing, and mental workload. Meanwhile, other scholars have focused on creating and optimizing in-vehicle technology-based interventions to increase safety. The purpose of this special issue is to bring together a collection of empirical and theoretical work focusing on understanding Driver Behavior and Performance in an Age of Increasingly Instrumented Vehicles. Several themes characterize this special issue:

ADVANCED DRIVER ASSISTANCE SYSTEMS (ADAS)

ADAS offer timely advice and feedback and can even actively take (or yield) control of the vehicle. This particular theme includes several applications of future ADAS. Ahmed et al. studied the usefulness of a connected ADAS that presents information from a control center. A group of professional drivers experienced the connected ADAS in a driving simulator. They indicated that the messages from the ADAS were most helpful when visibility was poor. In addition, the most valuable warnings were forward-collision (with other connected vehicles) and rerouting. Such findings are helpful for designers of ADAS that can receive information from control centers. In an interesting naturalistic study, Davis et al. discovered different behavioral patterns for elderly drivers: with Alzheimer disease (group 1), without Alzheimer disease (group 2), and with early signs of Alzheimer disease (group 3). The early Alzheimer disease group (#3) had fewer speeding and g-based driving events per driving distance traveled than the other groups. This result indicates sufficient orientation to self-regulate risk-taking while driving. Taking a broader perspective, this study provides an example of how ADAS can help detect (and perhaps take preventive measures) in situations related to driver health, e.g., when transferring from early signs to more advanced phases of the disease. Therefore, this theme is linked to the next theme in our Research Topic—the driver state monitoring theme (described below). ADAS technology may also have adverse outcomes. ADAS may cause deterioration of driving skills, encourage the diversion of attention from the driving task to other stimuli, and impair risk perception. Cohen-Lazry and Borowsky studied a novel multi-touch interface for an in-vehicle infotainment system and compared it with a functionally similar control interface. Participants using the multi-touch interface needed less time to complete secondary tasks, were quicker at identifying potential hazards, and reported lower subjective workload.

OPEN ACCESS

Edited and reviewed by:
Giacinto Barresi,
Italian Institute of Technology (IIT), Italy

****Correspondence:***
Oren Musicant
musicant.oren@gmail.com

Specialty section:
This article was submitted to Performance Science, a section of the journal Frontiers in Psychology

Received: *26 May 2021*
Accepted: *30 June 2021*
Published: *28 July 2021*

Citation:
Musicant O, Farah H, Shinar D and Collet C (2021) Editorial: Driver Behavior and Performance in an Age of Increasingly Instrumented Vehicles. Front. Psychol. 12:715239. doi: 10.3389/fpsyg.2021.715239

DRIVER STATE MONITORING

The second theme that characterizes several studies in this special issue is driver state monitoring. Current in-vehicle sensors can detect functional indices such as head position, eyelid, gaze, respiration rate, heart rate, and skin conductance to infer fatigue, mental workload, distraction, and risk perception. A significant portion of the studies in this special issue focused on using objective physiological indices to estimate mental workload and stress. For example, even in a simple common task, as searching for a parking space, Ponnambalam and Donmez found marginally significant changes in skin conductance and heart rate, indicating an increased mental workload. In conditional automation, drivers may engage in secondary tasks that increase the mental workload and impair driving performance. To continuously monitor the driver's mental workload, Meteier et al. recommend a combination of physiological indices (based on skin conductance, electrocardiogram, and respiration), the length of the time window for data processing, and machine learning models to identify elevated states of mental workload. Meteier et al. focused on measuring mental workload during a continuous performance of secondary tasks. But what about more discrete events without secondary tasks? Sahar et al. found that even everyday braking events trigger physiological responses of heart rate and heart rate variability. Moreover, they tested a novel intuitive, instantaneous, and unintrusive measure of stress—the steering wheel grip force (i.e., the force applied on the steering wheel in a non-steering-related task). They found that the steering wheel grip force correlated with braking intensity (a performance index), heart rate, and heart rate variability (stress indices), demonstrating its validity as a measure of stress. Another exciting research direction within the driver state monitoring theme is searching for drivers' hazard perception indicators. One option is to use biomarkers for hazard anticipations. In their study, Chirles et al. found that experienced drivers (compared to learners) had greater electrodermal responses to hazards in a hazard perception test before the hazard manifested itself.

AGE

Older drivers are known to make adjustments and self-regulation to accommodate cognitive, sensory, and motor capabilities changes. Such adjustments involve, for example, reducing long-distance drives. Here too, ADAS may be helpful. Shichrur et al. found that older drivers that used ADAS that provide collision warnings almost doubled their travel distance compared to the period before using the ADAS. It is possible that older drivers feel that the technology helps them compensate for reduced skills. Ironically, older drivers also struggle the most to adapt to ADAS technology. For example, according to Cooper et al., older drivers (compared to younger drivers) experienced increased workload when interacting with in-vehicle information systems: Older drivers were slower to respond to visual task demands and required more time to complete tasks such as entering a navigation destination, texting, calling, and dialing, and tuning the radio. These results remain consistent across three Human Machine Interfaces (HMIs) (two visual-motor and one vocal-based interaction). Although not obvious, human factors still play an essential role even in a fully autonomous vehicle. Stephenson et al. looked at the physiological responses of older passengers in an autonomous vehicle who faced expected and unexpected stops. After the unexpected stops, skin conductance sensors indicated increased passenger stress. This result suggests a need for interventions to reduce stress from unexpected events. Taking a broader perspective, the use of physiological sensors can serve to monitor passengers' stress as well as driver stress.

CONCLUSION

In the coming years, the human factor will continue to have an essential role in driving. The driver behavior and performance special issue includes studies examining drivers' behavior and state considering a range of autonomous levels. Emerging topics include novel methods in sensing and inferring driver states, novel human-machine interfaces, novel ADAS capabilities, and increasing interest in the elderly population that can benefit the most from ADAS yet have the greatest difficulty in adopting it.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

The guest editors would like to personally thank all the authors and reviewers who contributed their time and effort to this Research Topic. We also want to thank the help of Frontiers Editorial Office for all the help.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 03 June 2020
doi: 10.3389/fpsyg.2020.01154

Age-Related Differences in the Cognitive, Visual, and Temporal Demands of In-Vehicle Information Systems

Joel M. Cooper, Camille L. Wheatley, Madeleine M. McCarty, Conner J. Motzkus, Clara L. Lopes, Gus G. Erickson, Brian R. W. Baucom, William J. Horrey and David L. Strayer**

Applied Cognition Lab, Department of Psychology, The University of Utah, Salt Lake City, UT, United States

OPEN ACCESS

Edited by:
Oren Musicant,
Ariel University, Israel

Reviewed by:
Rachel Shichrur,
Ariel University, Israel
Philippe Rauffet,
Université Bretagne Sud, France

****Correspondence:***
Joel M. Cooper
joel.cpr@gmail.com
David L. Strayer
david.strayer@utah.edu

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Psychology

Received: *10 February 2020*
Accepted: *05 May 2020*
Published: *03 June 2020*

Citation:
Cooper JM, Wheatley CL, McCarty MM, Motzkus CJ, Lopes CL, Erickson GG, Baucom BRW, Horrey WJ and Strayer DL (2020) Age-Related Differences in the Cognitive, Visual, and Temporal Demands of In-Vehicle Information Systems. Front. Psychol. 11:1154. doi: 10.3389/fpsyg.2020.01154

In-vehicle information systems (IVIS) refer to a collection of features in vehicles that allow motorists to complete tasks (often unrelated to driving) while operating the vehicle. These systems may interfere, to a greater extent, with older drivers' ability to attend to the visual and cognitive demands of the driving environment. The current study sought to examine age-related differences in the visual, cognitive and temporal demands associated with IVIS interactions. Older and younger drivers completed a set of common tasks using the IVIS of a representative sample of six different vehicles while they drove along a low-density residential street. Evaluation measures included a Detection Response Task (DRT), to assess both cognitive and visual attention, and subjective measures following each condition using the NASA Task Load Index (TLX). Two age cohorts were evaluated: younger drivers between 21 and 36 years of age, and older drivers between 55 and 75 years of age. Participants completed experimental tasks involving interactions with the IVIS to achieve a specific goal (i.e., using the touch screen to tune the radio to a station; using voice commands to find a specified navigation destination, etc.). Performance of tasks varied according to different modes of interaction available in the vehicles. Older drivers took longer to complete tasks, were slower to react to stimuli, and reported higher task demand when interacting with IVIS. Older drivers stand to benefit the most from advancements in-vehicle technology, but ironically may struggle the most to use them. The results document significant age-related costs in the potential for distraction from IVIS interactions on the road.

Keywords: driving, reaction time, aging, technology, attention, workload

Abbreviations: AAAFTS, AAA Foundation for Traffic Safety; DRT, Detection Response Task; FHWA, Federal Highway Administration; HMI, Human-Machine Interface; IRB, Institutional Review Board; ISO, International Organization for Standardization; IVIS, in-vehicle information system; LCD, Liquid-Crystal Display; LED, Light-Emitting Diode; MVA, Motor Vehicle Accidents; NHTSA, National Highway Traffic Safety Administration; OEM, Original Equipment Manufacturer; OS, Operating System; PFC, Prefrontal Cortex; SAE, Society of Automotive Engineers; SuRT, Surrogate Reference Task; TEORT, Total Eyes-Off-Road Time; USB, Universal Serial Bus.

INTRODUCTION

Operating a motor vehicle is one of the riskiest activities that adults engage in on a regular basis. In fact, roadway crashes are one of the leading causes of unintentional injury and death (World Health Organization [WHO], 2011; National Safety Council [NSC], 2017) and a significant percentage of crashes involve some form of distraction or inattention (e.g., Dingus et al., 2016). To safely operate a motor vehicle, drivers must maintain their eyes on the forward roadway and keep their mind focused on the drive. This becomes increasingly difficult with the prevalence of in-vehicle electronics. These systems change the way that drivers manage their attention behind the wheel, potentially leading to increases in driver distraction—especially as systems provide more information, functions, and features to drivers.

There are several components that factor into how distracting a secondary task is for the driver (e.g., Ranney et al., 2000; Regan et al., 2011; Strayer et al., 2011). One important factor is the cognitive demand associated with *Scanning, Predicting, Identifying, Deciding, and Executing Responses* ("SPIDER" – for a review see Strayer and Fisher, 2016). Performing cognitively demanding secondary tasks has been shown to impair each of these "SPIDER-related" processes and increase the relative risk of a crash (Fisher and Strayer, 2014). Another important factor is the visual demand associated with secondary-task interactions. Guidelines derived from the "radio tuning task" suggest that individual eye glances to a device while performing a secondary task should not exceed 2 s (Perez et al., 2013). Regardless of the type of secondary task, crash rates have been shown to systematically increase as the duration of glances away from the road increases (e.g., Simons-Morton et al., 2015). For example, when paired with the primary task of driving, texting is risky because it takes the driver's eyes off the road for an average of 4.6 s (e.g., Drews et al., 2009; Olson et al., 2009). In many cases glances away from the forward roadway involve guiding a motor response (e.g., touching a location on the center stack screen). A final factor to consider is the duration of a distracting secondary task. Shutko and Tijerina (2006) suggest that task duration is critical because it represents the time in which an unexpected event might occur. All other things being equal, tasks that takes twice as long to complete will result in twice the potential risk of an adverse event.

Driver interactions with current and emerging in-vehicle information systems (IVIS) are often characterized by lengthy, complex, visual-manual, and auditory-vocal action sequences. For example, a driver may initiate a destination entry sequence with the press of a button on the steering wheel, followed by a verbal address entry, ending with the use of the touch screen. An earlier benchmarking effort from the Crash Avoidance Metrics Partnership (CAMP; Angell et al., 2006) evaluated a variety of older secondary tasks involving different types of visual, manual and cognitive interaction. Visual-manual tasks involved tuning the radio or adjusting fan speed using physical buttons located in the center console. Auditory-vocal tasks involved listening to an audiobook or sports broadcasts and answering related questions. This CAMP analysis found distinct profiles indicating that driver's workload was multimodal and characterized by different combinations of visual, manual and cognitive components.

More recently, Strayer, Cooper, and colleagues reported on a program of research designed to understand the distraction potential associated with tasks now commonly available in new vehicles (Strayer et al., 2013; Strayer et al., 2014; Cooper et al., 2014; Strayer et al., 2017, 2018). This newer research examined some of the older task types evaluated with CAMP and also newer IVIS interactions that were not available in 2006. One important outcome of this research was a multimodal evaluation method for assessing the cognitive, visual, and temporal demands of complex multimodal IVIS interactions (see Strayer et al., 2017). Indeed, large variation in the distraction potential was observed with different tasks types (e.g., audio entertainment, calling and dialing, texting, and destination entry to support GPS navigation), and modes of interaction (e.g., center stack touchscreen; voice-commands).

Importantly, the reactions of older adults are often slower than those of younger adults, a phenomenon referred to as generalized slowing (e.g., Cerella, 1985; Salthouse, 1996). The age-related effects are magnified by the complexity of the interactions (Cerella et al., 1980). In fact, compared to differences in baseline reaction time (reflecting generalized slowing), the age-related differences more than doubled when participants used voice-based commands to select music or dial a phone number (Strayer et al., 2015). Because the duration of secondary task activities is greater for older than for younger drivers, age-related differences are expected to increase as the complexity of the secondary task increases. New in-vehicle systems and other secondary task activities may be especially problematic for older drivers (Albert et al., 2018). Ongoing research seeks to understand age-related differences in multitasking (Clapp et al., 2011) and the technology barriers that older drivers encounter (Vaportzis et al., 2017). However, little is known about the way in which drivers of any age interact with these complex multimodal In-Vehicle Information Systems (IVIS). These technologies have the potential to make driving safe and enjoyable. If they are not carefully implemented; however, they will decrease attention to the roadway.

Watson et al. (2011) suggested that the U-shaped function depicting crash rates and age is closely aligned with the maturation and decline in prefrontal cortical (PFC) regions of the brain (e.g., an inverted U-shaped function across the lifespan). The PFC regions are involved in a wide variety of higher-level cognitive/executive functions that support driving-related attention (e.g., scanning, predicting, identifying, deciding, and executing responses). In fact, laboratory studies have found greater multitasking costs for older adults (e.g., Craik, 1977; Hartley, 1992; Kramer and Larish, 1996; Hartley and Little, 1999; McDowd and Shaw, 2000) and Strayer et al. (2015) observed that older drivers experienced greater levels of cognitive demand with voice-based IVIS systems.

Current Research

This study examined the cognitive and visual demand of younger and older drivers as they performed a variety of task types while driving a vehicle on a section of residential roadway. Workload measures were compared across two age groups and six different vehicles supporting different IVIS. The current research addressed two questions related to the use of these IVIS interactions.

Q1: Do the demands of IVIS interactions differ for older and younger drivers? If so, how?

Prior research has demonstrated that senescence is associated with declines in physical and cognitive performance that can impact safe driving. When older drivers interact with IVIS, they are more likely to experience cognitive, visual, or temporal interference. Furthermore, some types of IVIS interactions may present unique demands for older drivers.

Q2: Are some interfaces more difficult for older drivers to use? If so, why?

Research on IVIS voice interactions has found that older drivers experience higher cognitive demands when completing common tasks (e.g., Strayer et al., 2015). It is not clear, however, whether workload differences exist between older and younger drivers when completing tasks using controls housed in the center stack or when using center console controls. The ways in which older and younger drivers interact with IVIS may change the level of demand that they experience.

MATERIALS AND METHODS

Participants

125 participants (52 females) were recruited via flyers, social media posts and local newsprint advertising with approval from the University of Utah Institutional Review Board (IRB). Eligible participants were native English speakers, had normal or corrected-to-normal vision, and held a valid driver's license. Participants were also required to have proof of medical insurance and no accident involvement within the past 2 years. To ensure participants held a clean driving record, a Motor Vehicle Record report was obtained by the University of Utah's Division of Risk Management.

All participants belonged to one of two age cohorts: younger drivers between 21 and 36 years of age (M = 24.8 years, St Dev = 2.97), and older drivers between 55 and 75 years of age (M = 65.8 years, St Dev = 5.36). Following University of Utah policy, participants were required to take and pass a 20-min online defensive driving course and certification test. Compensation was prorated at $20 per hour.

All participants were all healthy adult drivers with no physical or mental deficits. Younger and older participants self-reported health was 5.95 (0.85) vs. 6.10 (0.63) on a 7-point scale, a difference that was not statistically significant, $p > 0.10$. Younger and older participants drove an average 9.0 (7.19) vs. 8.8 (6.56) hours per week, a difference that was not statistically significant, $p > 0.7$. Younger and older participants reported an average of 7.38 (1.16) vs.7.38 (0.85) hours of sleep the night before testing, a difference that was not statistically significant, $p > 0.98$. Finally, no participants reported a history of neurological disorders.

Twenty-four individuals from each age cohort were tested in six unique vehicles, resulting in 48 participants per vehicle (i.e., each cell in the 6 × 2 factorial design had 24 participants). The study design allowed participants to drive all six vehicles, however this was not always possible. Participants were sample-matched by age and number of driving sessions in each of the evaluated vehicles; this was done to ensure that each age cohort was comprised of similar numbers of naive and repeat participants for the vehicle. The number of exposures were matched across vehicles and age cohorts as closely as possible; however, due to factors such as order of testing and availability of participants, exact matching was not possible. Thus, a planned missing data design was used (e.g., Graham et al., 2006; Little and Rhemtulla, 2013) as only eight individuals drove all six vehicles. Among the younger age cohort (21–36), 20 participants drove 1 vehicle, 16 drove 2 vehicles, 11 drove 3 vehicles, 7 drove 4 vehicles, 2 drove 5 vehicles, and 4 drove 6 vehicles. Among the older age cohort (55–75),: 28 participants drove 1 vehicle, 14 drove 2 vehicles, 10 drive 3 vehicles, 3 drove 4 vehicles, 6 drove 5 vehicles, and 4 drove 6 vehicles. Participants were initially naïve to the specific systems and tasks but were trained until they felt competent and confident performing each type of task while driving.

Stimuli and Apparatus

Vehicles

The vehicles that were used for the study are listed below with the native infotainment system for each shown in parentheses. These cars were selected for inclusion in the study based on market diversity, availability, and IVIS functionality. Vehicles were acquired through Enterprise Rent-A-Car or purchased for testing.

- 2018 Audi A6 Premium (Man and Machine Intersect or MMI®)
- 2018 Cadillac CT6 Premium Luxury – Custom Packages (Custom User Experience or CUE®)
- 2018 Lincoln Navigator Select L (SYNC 3®)
- 2018 Mazda CX-5 Grand Touring (Mazda Connect®)
- 2018 Nissan Pathfinder SL (NissanConnect®)
- 2018 Volvo XC90 Momentum – Custom Packages (Sensus Connect®)

Equipment

Identical Google Pixel 2 phones on the T-Mobile network w Bluetooth-paired with each vehicle. An iPad Mini 4 (20.1 cm diagonal LED-backlit Multi-Touch display) was used to administer a visual-manual reference task (detailed below) and to survey participants on their self-reported measures of workload.

Each vehicle utilizes a variety of functions that facilitate interaction with the system such as touch screens, physical buttons, voice commands, touch/trackpads, and rotary wheels. Features were grouped into three Modes of Interaction: Voice Commands, Center Console, and Center Stack. IVIS functions were grouped into four Task Types: Audio Entertainment, Calling and Dialing, Text Messaging, and Navigation Entry.

Participants completed tasks involving interactions with the IVIS to achieve a goal (i.e., using the touch screen to tune the radio to a station, using voice commands to find a specified navigation destination, etc.) while driving. Tasks were categorized into four Task Types and three Modes of Interaction.

The possible Task Types performed by the participant were:

- Audio Entertainment: Participants tuned the radio to specific AM and FM frequencies and selected music from a USB connected iPad mini, using designated categories such as song titles, music genres, artist names, and album titles.
- Calling and Dialing: A list of 91 contacts with a mobile and/or work number was created for participant testing. Participants were instructed to call designated contacts and the associated number type was specified when applicable. In vehicles capable of dialing phone numbers, participants were instructed to dial the phone number 801-555-1234 as well as their own phone number.
- Text Messaging: Participants were provided with hypothetical scenarios in which they received text messages from various contacts and were instructed to interact with the messages using specified Modes of Interaction. Vehicles varied in their SMS capabilities. A portion of the system/Mode of Interaction combinations allowed users to listen to messages and reply with predetermined responses, or solely listen to the messages and not respond. Other vehicles and Modes of Interaction allowed users to respond to text messages using free dictation.
- Navigation Entry: Participants started and canceled route guidance to different locations based on a hypothetical situation they were given that differed slightly according to the options available in each system.

The Modes of Interaction with each system are described below. Interaction modalities were selected based on compatibility with the specific tasks.

- Center Stack: Visual-manual tasks were performed using the center stack interfaces found in the middle of the dash to the right of the driver. Center stack systems generally include a touchscreen to integrate visual/manual interactions so that drivers can select options and navigate menus via touch, scroll bars, seek arrows, etc. to complete tasks using options displayed on the screen. Some vehicles provide physical buttons near the touch screen for selection of options.
- Center Console: Vehicles utilizing center console controls replace or augment a touchscreen interface or manual center stack controls with an interface usually consisting of a rotary wheel and manual buttons in the center console to the right of the driver. The center console controls facilitate interactions with the center stack display located in the middle of the dash. The rotary can be spun to scroll through menus and used like a button to make selections. In some cases, the rotary wheel interfaces can be maneuvered in various directions to navigate menus, like a joystick. In the case of the Audi A6, the center console controls also incorporated a touch-sensitive pad that could be used to draw letters and numbers in search functions or select preset radio stations.
- Auditory Vocal: The voice-based interaction with each IVIS system is initiated by the press of a physical voice recognition button on the steering wheel. Microphones installed in the vehicle process the driver's verbal commands and assist them while performing tasks in the vehicle. Possible voice command options may be presented audibly or displayed on the vehicle's center stack or instrument cluster to assist users in achieving their goal.

The configuration of Task Types and Modes of Interaction depended on each system's unique capabilities. All vehicles supported Voice Commands, however each vehicle differed on visual/manual interaction (e.g., touchscreen, manual buttons, center console controls). Furthermore, different systems required specific syntax or commands to be given in a systematic order to accomplish tasks in different interaction modes. Task lists were developed to test the various combinations of features and functions available in each system. Tasks were standardized across systems as much as possible, given the variability in system interactions. The tasks supported for each vehicle system are noted in **Table 1**. A complete list of all task instructions for each vehicle is provided in **Supplementary Appendix A**.

Detection Response Task

Participants were trained to respond to both a vibrotactile stimulus and a remote visual stimulus (cf. International Organization for Standardization, 2015). A vibrotactile stimulus was positioned under the participant's left collarbone, and a remote LED light was placed along a strip of Velcro on the dashboard in such a manner that the participants only saw the reflection of the light as it changed from orange to red in the windshield directly in the forward line of sight (see Castro et al., 2016; Cooper et al., 2016). This variant of the standard DRT was used to maximize sensitivity to both cognitive and visual attention. Reaction time to the vibrotactile stimulus was used to assess cognitive workload while hit rate to the forward LED was used as a measure of competing visual demand.

A microswitch (i.e., small button) was attached to either the index or middle finger of the left hand and pressed against the steering wheel by participants when they felt a vibration or saw the light change colors. Each press of the switch was counted and recorded but only the first response was used to determine response time to the stimulus. Response time with sub-millisecond resolution to the vibrotactile onset or LED light was recorded using a dedicated microprocessor that passed results over USB connection to the host computer for storage and later analysis.

Following International Organization for Standardization (2015), the vibrotactile device emitted a small vibration stimulus, like a vibrating cell phone. The remote light stimulus consisted of a change in color from orange to red. This color changing LED stimulus differed from the ISO standard (see Castro et al., 2016). The occurrences of these stimuli cued the participant to respond as quickly as possible by pressing the microswitch against the steering wheel. The tactile and light stimuli were equiprobable and were programed to occur every three to 5 s (with a rectangular distribution of inter-stimulus intervals within that range) and lasted for 1 s or until the participant pressed the microswitch.

TABLE 1 | Modes of Interaction and tasks types for the 6 different vehicles.

Vehicle	**Audi A6**		**Cadillac CT6**		**Lincoln Navigator**		**Mazda CX-5**		**Nissan Pathfinder**		**Volvo XC90**	
Mode of interaction	**CS**	**VC**	**CC**	**VC**	**CS**	**VC**	**CC**	**VC**	**CS**	**VC**	**CS**	**VC**
Audio entertainment	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Calling and dialing	✓	✓	✓	✓	✓	✓	✓	✓	✓*	✓	✓	✓
Navigation	✓	✓	✓	✓	✓*	✓	✓**	✓**	✓***	✓	✓	✓
Text messaging	✓	✓	✓	–	✓*	✓**	✓	✓	✓	✓	–	✓

Mode of Interaction: CS = Center Stack; VC = Voice Commands, CC = Center Console. Cells with checks indicate the availability of tasks for that Vehicle/Mode of interaction. Asterisks indicate cases where a variation in task instructions was required to accommodate vehicle capabilities or limitations. Empty cells represent tasks that were not available for that Vehicle/Mode of Interaction (e.g., The Cadillac CT6 did not provide Text Messaging functionality through voice controls). The full task instruction set for each Vehicle/Mode of Interaction is provided in ***Supplementary Appendix A****.*

Procedure

Driving Route

A suburban residential street with a 25-mph speed limit was used for the on-road driving study. The route was a straight road with four stop signs and two speed bumps. Participants were required to follow all traffic laws and adhere to the 25-mph speed limit. The driving route was approximately two miles long one-way with an average drive time of 6 min. A researcher was present in the passenger seat of each vehicle for safety monitoring and data collection. An image of the driving route can be seen in **Figure 1**. The respective start/end points of the driving course were 40.781944, −111.8820912 and 40.7770036, −111.8438273.

Training

Prior to the start of the study, participants were provided the time to become accustomed to the vehicle, the route, and the DRT. The initial familiarization period is as follows:

- DRT training: Participants were instructed on how to respond to the light and vibration motor using the microswitch. Researchers monitored participants as they practiced responding to 10 stimuli to ensure participants produced response times of less than 500 milliseconds, indicating a competence and understanding of the task. After initial training, participants were given the opportunity to practice responding to the DRT while driving during the practice drive described below.
- Practice route: Prior to data collection, participants were instructed to drive the route seen in **Figure 1**. Researchers pointed out all obvious and identifiable road hazards. This practice drive allowed participants to familiarize themselves with the road as well as the handling of the vehicles.

Participants were trained to interact with and complete tasks using the assigned Mode of Interaction before each condition began. Participants were required to complete three task trials without error, while simultaneously responding to the DRT prior to starting the driving task for each of the system interactions.

Experimental Blocks

During the experimental blocks, participants were instructed to complete a set of tasks administered by the researcher using an assigned Mode of Interaction with the infotainment system. Driving the vehicle was emphasized as task prioritization and was expressed to participants in verbal instructions.

Participants were asked to pull over on the side of the road at the termination of one length of the route. The subsequent experimental block, equipped with a new assigned Task Type and Mode of Interaction, began in the opposite direction on the same route and concluded in the same manner. This was repeated until all conditions were completed, resulting in alternating travel directions for each experimental block. The order of Task Types and conditions administered in each vehicle was counterbalanced across 24 participants in each Age Cohort.

Tasks were only administered in safe and normal driving conditions. Disruptions to the natural driving environment resulted in the researcher instructing the participant to terminate the current task and only administering a new task when it was safe to do so. Tasks were not administered as participants approached intersections or construction zones. Normal behaviors of other vehicles and pedestrians were within the scope of the natural driving environment.

Participants were provided with verbal hypothetical situations or commands as cues from the passenger researcher (e.g., "Jack Olsen would like you to call him on his cell phone."). Participants were trained to wait to start each task until the researcher said "go." After the completion of each task, participants were trained to say, "done." Tasks were delivered with an approximately 5 s interval between the participants' announcement of completion and the researchers' administration of the next task. Researchers denoted each task's start and end time by pressing designated keys on the data collection computer, thus indicating the timing of on-task performance on the driving route. DRT trials were considered valid for inclusion in the statistical analysis if they occurred between these start and end times. Participants were encouraged to complete tasks as efficiently as possible, however drivers were given as much time as needed to complete each task, unless the end of the route was reached in which case tasks were terminated prematurely and later omitted from analysis.

Participants also performed three control tasks while driving one length of the designated route per task. These tasks provided a standard set of performance references which included a single-task baseline, a high cognitive demand reference task (nBack), and a high visual-demand reference task (SuRT).

- Single-task Baseline: Participants performed a single-task baseline drive using the vehicle being tested on the

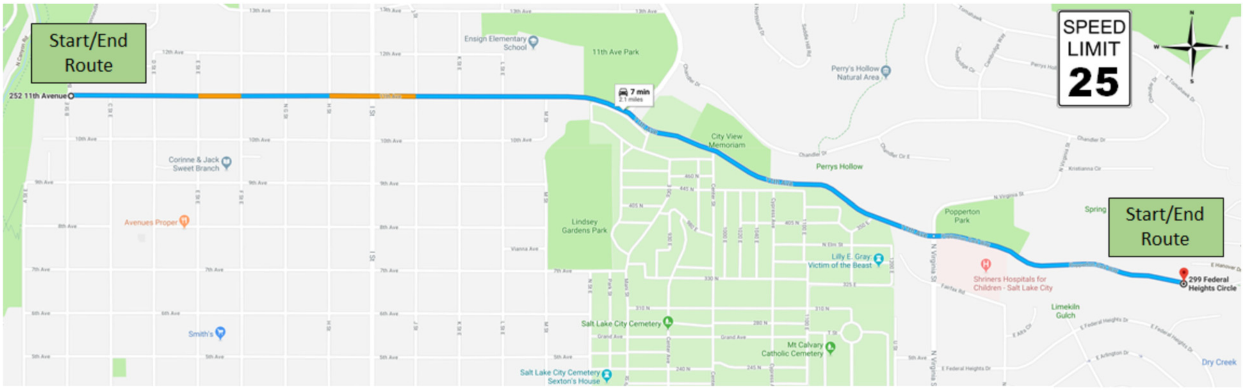


FIGURE 1 | A map image of the designated driving route, a two-mile-long residential roadway in Salt Lake City, Utah.

designated route, without interacting with the infotainment system. During the single-task baseline, participants interacted solely with the DRT stimuli, responding to both the tactile stimulus and light change as fast as possible and were asked to remain silent as to minimize distraction.

- Auditory nBack task: The auditory nBack task (Mehler et al., 2011) presented a pre-recorded, randomized set of numbers ranging from zero to nine in sequences of 10. In each sequence, numbers were spoken aloud at a rate of one digit every 2.25 s. Participants were instructed to verbally repeat the number that was presented two trials earlier as they concurrently listened for the next number in the sequence. Participants were told to respond as accurately as possible to the nBack stimuli while researchers monitored performance in real-time. During the nBack task, participants also responded to the DRT stimuli.
- Surrogate Reference Task (SuRT): A modernized version of the SuRT task (International Organization for Standardization, 2012) was presented on an iPad Mini 4 with circles printed in black on a white background. A target was presented on the display amidst 21–27 distractors. The target was an open circle 1.5 cm in diameter and the distractors were open circles 1.2 cm in diameter. Participants' were instructed to touch the location of the target. Immediately thereafter, a new display was presented with a different configuration of targets and distractors. The location of targets and distractors was randomized across the trials in the SuRT task. Participants were instructed to continuously perform the SuRT task while giving the driving task highest priority as researchers monitored performance in real-time. Researchers instructed participants to pause the SuRT task at intersections and in the event of potential hazards on the roadway. During the SuRT task, participants also responded to the DRT stimuli.

After the completion of each condition, participants completed the NASA-TLX (Hart and Staveland, 1988) to assess the subjective workload of the system presented on the iPad Mini 4. Following this assessment, participants were asked an open-ended question as an opportunity to describe or detail information not captured by the restrictive NASA-TLX questions: "Do you have any comments about the task or vehicle after this last run?"

Dependent Measures

Detection Response Task data were preprocessed following procedures outlined in International Organization for Standardization (2015). All response times faster than 100 milliseconds which were considered impossible or inadvertent responses and were not considered valid. Similarly, reaction times slower than 2500 milliseconds were eliminated from our overall calculation for Reaction Time. Non-responses or responses that were made after 2500 milliseconds from the stimulus onset were coded as a miss. System interaction was recorded by the researcher via pressing designated keys on the DRT host computer, allowing the identification of "on-task" and "off-task" segments of driving. Incomplete, interrupted, or otherwise invalid tasks, were marked with a key-flag and excluded from the analysis. The dependent measures obtained in the study are listed below:

- DRT – Reaction Time: Defined as the sum of all valid reaction times to the DRT task divided by the number of valid reaction times. Reaction time to the DRT was used to calculate cognitive demand during each experimental condition. This was used to gauge the approximate mental workload and allocation of cognitive resources required by the task for each type of IVIS interaction. Reaction times to both stimuli were included in analysis.
- DRT – Hit Rate: Defined as the number of valid responses divided by the total number of valid stimuli presented during each condition. Hit Rate to the DRT was used to calculate visual demand, or how much visual attention was required by the task during each experimental condition. In order to maximize the sensitivity to divided visual attention effects (e.g., looking away from the forward roadway), analyses were only conducted on responses to the remote LED stimulus.

Task Completion Time was obtained from the time stamp on the DRT data file defined as the time researchers said "go" and participants first initiated an action to the time when that action was completed, and the participant said, "done." Tasks with irregular occurrences and errors in administration or performance that may have affected Task Completion Time were marked as abnormal during data collection and were not included in subsequent analyses. When assessed using the visual occlusion methodology, the NHTSA guidelines provide an implicit upper limit of 24 s of total task time (National Highway Transportation Safety Administration [NHTSA], 2013). While originally intended for visual/manual tasks, these guidelines provide a reasonable upper limit for task durations of any type.

Experimental Design

The experimental design was a 2 (Age Cohort: older or younger drivers) x 4 (Task Type: Audio Entertainment, Calling and Dialing, Navigation Entry, and Text Messaging) x 3 (Mode of Interaction: Auditory/Vocal, Center Stack, Center Console) factorial with 24 participants evaluated in each cell of the factorial. However, not all system interactions offered the full factorial design (i.e., not all Task Types and Modes of Interaction were available in all vehicles).

Data Analysis

Linear mixed effects analyses were performed using R 3.5.1 (R Core Team, 2018) and lme4 version 1.1-18-1 (Bates et al., 2015). In the analyses reported below, models containing the Age Cohort plus one additional factor (i.e., Task Type and Mode of Interaction) were compared to a model without the effect in question. All main effects and two-way interactions with Age Cohort were analyzed and are included below. *P*-values were obtained by likelihood ratio tests comparing the full linear mixed effects model to the partial linear mixed effects model (see Winter, 2013). This linear mixed modeling analysis has the advantage of analyzing all available data while adjusting fixed effect, random effect, and likelihood ratio test estimates for missing data. The full analysis script is available for download here: https://github.com/utahcdst/Aging-Report-Frontiers.

Pairwise comparisons for each of the analyses are provided in a tabular format see **Tables 2–9**. Pairwise comparisons were extracted through the sequential evaluation of each model in question using a factor re-referencing approach. Each table of pairwise comparisons is structured to provide the (1) the means and standard deviations at each factor level by age, (2) the pairwise comparisons for the age contrast, which indicates whether the effect of age was significant at each level of the factor in question, (3) the pairwise comparisons of the effect in questions, which indicates whether factor levels differed from each other, and (4), the pairwise comparisons for the effect of age at each level of the factor in question. This indicates whether the age effects at each factor level differed from each other. This selective set of pairwise comparisons addresses the core effects of interest.

For each independent variable (Task Completion Time, DRT Reaction Time, DRT Hit Rate, Subjective Workload), Linear Mixed Effects Models were built to explore the main effects of Age Cohort, Task Type, and Mode of Interaction as well as all 2-way interactions with Age Cohort (e.g., Age Cohort by Task Type and Age Cohort by Mode of Interaction).

Where appropriate, results were analyzed and modeled with the inclusion of the baseline tasks (Single-task, SuRT, nBack). Baseline tasks were not included in the analysis of Task Completion Time. Results address the question of whether there were significant age differences in the associations of interactions with the vehicle technology with the independent variables.

Mean results for each of the main effects are provided in the units in which they were recorded, along with the standard error (SE) in parentheses. Due to the number of statistical comparisons performed, we used a more conservative $\alpha = 0.01$ and $\alpha = 0.001$ to denote varying levels of statistical significance. Effects that reach these levels are flagged with a single '*' and a double '**' respectively. This more conservative significance level helps to reduce the likelihood of false positives in the statistical analysis.

RESULTS

Task Completion Time

Main Effects

Results indicated that Task Completion Time differed by Age Cohort, $\chi^2(1) = 51.42$, $p < 0.001$ (Younger: $M = 23.5$, $SD = 9.83$; Older: $M = 30.2$, $SD = 15.2$) with older drivers taking significantly longer to complete tasks than younger drivers. Additionally, there were significant main effects of Task Completion Time for Task Type, $\chi^2(3) = 785.85$, $p < 0.001$ (Audio Entertainment: $M = 21.6$, $SD = 9.09$; Calling and Dialing: $M = 20.0$, $SD = 5.76$; Text Messaging: $M = 30.7$, $SD = 35.6$; Navigation Entry: $M = 30.2$, $SD = 15.2$), and Mode of Interaction, $\chi^2(2) = 119.56$, $p < 0.001$ (Auditory Vocal: $M = 29.4$, $SD = 14.9$; Center Console: $M = 27.4$, $SD = 11.7$; Center Stack: $M = 22.6$, $SD = 9.64$).

Age Cohort by Task Type

The analysis revealed a significant two-way interaction between Age Cohort and Task Type, $\chi^2(1, 3) = 12.52$, $p = 0.006$. Age contrasts indicated that the effects of age reached significance at all levels of Task Type and that all levels of Task Type differed from each other. Furthermore, Task Type contrasts by Age Cohort suggests that older drivers had an especially difficult time with the Navigation Entry task. Notably, only the median Task Completion Time for the Audio Entertainment and Calling and Dialing tasks came in under the 24-s referent for both age groups (see Strayer et al., 2013).

Age Cohort by Mode of Interaction

The interaction between Mode of Interaction and Age Cohort was not significant, $\chi^2(1, 2) = 1.09$, $p = 0.57$, suggesting that the increased Task Completion Time for older drivers was similar across all three Modes of Interaction. That is, while Age Cohort and Mode of Interaction both affected Task Completion Time, the impact of each was not dependent on the other. Age contrasts indicated that the effect of age reached

TABLE 2 | Pairwise comparisons for task completion time as a function of task type and age cohort.

	Task completion time by task type	Audio entertainment	Calling and dialing	Text messaging	Navigation entry
Means and SD	Ages 21–36 (*n* = 24) Mean (SD)	18.0 (5.01)	17.7 (3.64)	27.7 (11.8)	31.4 (8.63)
	Ages 55–75 (*n* = 24) Mean (SD)	25.4 (10.7)	22.4 (6.48)	33.8 (20.4)	40.0 (14.1)
Age cohort contrasts: *t*-value		7.00**	4.64**	5.53**	8.00**
Task type contrasts: *t*-value	Audio entertainment				
	Calling and dialing	−2.86*			
	Text messaging	14.53**	17.25**		
	Navigation entry	23.86**	26.71**	8.20**	
Task type contrasts by age cohort: *t*-value	Audio entertainment				
	Calling and dialing	−2.35			
	Text messaging	−1.11	1.12		
	Navigation entry	1.02	3.36**	2.08	

*$p < 0.01$, **$p < 0.001$.*

TABLE 3 | Pairwise coparisons for task completion time as a function of mode of interaction and age cohort.

	Mode of interaction	Auditory vocal	Center console	Center stack
Means and SD	Ages 21–36 (*n* = 24) Mean (SD)	26.2 (11.3)	23.2 (7.85)	19.7 (6.42)
	Ages 55–75 (*n* = 24) Mean (SD)	32.7 (17.2)	31.4 (13.3)	25.7 (11.3)
Age cohort contrasts		6.98**	5.93**	6.01**
Mode of interaction contrasts	Auditory vocal			
	Center console	−4.10**		
	Center stack	−11.03**	−4.36**	
Mode of interaction contrasts by age cohort	Auditory vocal			
	Center console	−0.95		
	Center stack	−0.98	−0.98	

*$p<0.01$, **$p<0.001$.*

TABLE 4 | Pairwise comparisons for reaction time as a function of task type and age cohort.

	Task type	Single	Audio entertainment	Calling and dialing	Text messaging	Navigation entry	nBack	SuRT
Means and SD	Ages 21–36 (*n* = 24) Mean (SD)	526 (110)	778 (154)	762 (150)	763 (158)	773 (157)	760 (174)	703 (147)
	Ages 55–75 (*n* = 24) Mean (SD)	639 (135)	946 (184)	934 (166)	951 (176)	954 (175)	944 (151)	881 (182)
Age cohort contrasts		4.50**	7.98**	7.34**	7.91**	7.70**	7.47**	7.16**
Task type contrasts	Single							
	Audio entertainment	38.56**						
	Calling and dialing	35.34**	−3.94**					
	Text messaging	35.25**	−2.46	1.28				
	Navigation entry	37.49**	−1.27	2.66*	1.25			
	nBack	31.17**	−2.57	0.65	−0.44	−1.53		
	SuRT	24.33**	−10.53**	−7.31**	−8.14**	−9.48**	−6.89**	
Task type contrasts by age cohort	Single							
	Audio entertainment	4.99**						
	Calling and dialing	3.99**	−1.22					
	Text messaging	4.90**	0.10	1.26				
	Navigation entry	4.56**	−0.52	0.70	−0.59			
	nBack	4.31**	−0.01	0.99	−0.09	0.42		
	SuRT	3.85**	−0.55	0.45	−0.61	−0.12	−0.46	

*$p<0.01$, **$p<0.001$.*

significance for all levels of Mode of Interaction. Mode of Interaction contrasts indicated that performance in each Mode of Interaction differed between performance in the other Modes of Interaction. Mode of Interaction contrasts by Age Cohort suggested that the magnitude of the Cohort effect was not dependent on the specific.

TABLE 5 | Pairwise comparisons for reaction time as a function of mode of interaction and age cohort.

	Mode of interaction	Single	Auditory vocal	Center console	Center stack	nBack	SuRT
Means and SD	Ages 21–36 (n = 24) Mean (SD)	526 (110)	764 (161)	775 (147)	775 (148)	760 (174)	703 (147)
	Ages 55–75 (n = 24) Mean (SD)	639 (135)	954 (188)	957 (160)	942 (164)	944 (151)	881 (182)
Age cohort contrasts:		4.49**	8.37**	7.34**	7.35**	7.47**	7.16**
Mode of interaction contrasts	Single						
	Auditory vocal	40.08**					
	Center console	34.28**	0.89				
	Center stack	37.84**	0.04	−0.78			
	nBack	31.09**	−0.94	−1.46	−0.91		
	SuRT	24.27**	−9.60**	−8.72**	−9.08**	−6.87**	
Mode of interaction contrasts	Single						
by age cohort	Auditory vocal	5.63**					
	Center console	4.05**	−0.88				
	Center stack	3.98**	−1.92	−0.61			
	nBack	4.30**	−0.22	0.49	1.13		
	SuRT	3.84**	−0.80	0.00	0.58	−0.46	

*$p<0.01$, **$p<0.001$.

TABLE 6 | Pairwise comparisons for hit rate as a function of task type and age cohort.

	Task Type	Single	Audio entertainment	Calling and dialing	Text messaging	Navigation entry	nBack	SuRT
Means and SD	Ages 21–36 (n = 24) Mean (SD)	0.90 (0.11)	0.59 (0.24)	0.66 (0.24)	0.66 (0.21)	0.60 (0.24)	0.76 (0.19)	0.65 (0.18)
	Ages 55–75 (n = 24) Mean (SD)	0.80 (0.20)	0.29 (0.24)	0.37 (0.26)	0.35 (0.23)	0.33 (0.23)	0.53 (0.25)	0.44 (0.26)
Age cohort contrasts		−3.79**	−10.34**	−9.78**	−10.01**	−9.29**	−7.26**	−6.67**
Task type contrasts	Single							
	Audio entertainment	−34.73**						
	Calling and dialing	−28.55**	7.56**					
	Text messaging	−28.47**	6.18**	−1.02				
	Navigation entry	−32.37**	2.85*	−4.69**	−3.45**			
	nBack	−15.12**	17.27**	11.10**	11.58**	14.92**		
	SuRT	−22.17**	9.19**	3.01*	3.76**	6.85**	−7.02**	
Task type contrasts by	Single							
age cohort	Audio entertainment	−8.69**						
	Calling and dialing	−7.92**	0.94					
	Text messaging	−8.22**	0.22	−0.67				
	Navigation entry	−7.24**	1.76	0.83	1.46			
	nBack	−4.55**	3.43**	2.66*	3.13**	1.99		
	SuRT	−3.77**	4.35**	3.58**	4.02**	2.90*	0.79	

*$p<0.01$, **$p<0.001$.

Age Cohort by Mode of Interaction by Task Type

The three-way interaction between each of these factors was significant, $\chi^2(1, 3, 2) = 24.8$, $p < 0.001$. This higher order interaction suggests that the effect of Age was dependent on the specific Task/Mode combination **Figure 2**.

DRT Reaction Time

Main Effects

Results indicated that DRT Reaction Time differed by Age Cohort, $\chi^2(1) = 49.8$, $p < 0.001$ (Young: $M = 739$, $SD = 168$; Older: $M = 914$, $SD = 193$), with older drivers taking significantly longer, on average, to respond to the DRT than younger drivers. Additionally, there were significant main effects for Task Type, $\chi^2(6) = 1466$, $p < 0.001$ (Audio Entertainment: $M = 870$, $SD = 194$; Calling and Dialing: $M = 848$, $SD = 180$; Text Messaging: $M = 856$, $SD = 191$; Navigation Entry: $M = 863$, $SD = 189$), and Mode of Interaction, $\chi^2(5) = 1450$, $p < 0.001$ (Auditory Vocal: $M = 858$, $SD = 199$; Center Console: $M = 868$, $SD = 179$; Center Stack: $M = 856$, $SD = 177$).

Age Cohort by Task Type

Analysis of the two-way interaction between Age Cohort and Task Type indicated that the interaction reached significance, $\chi^2(1, 6) = 31.6$, $p < 0.001$. Inspection of the data reveals a highly consistent effect of task engagement and age across each of the Task Types. Age Cohort contrasts indicated that the effect of age reached significance at each level of Task Type. Task Type contrasts suggest that Reaction Time to the Audio Entertainment

TABLE 7 | Pairwise comparisons for hit rate as a function of mode of interaction and age cohort.

	Mode of interaction	Single	Auditory vocal	Center console	Center stack	nBack	SuRT
Means and SD	Ages 21–36 (n = 24) Mean (SD)	0.90 (0.11)	0.70 (0.22)	0.54 (0.23)	0.56 (0.23)	0.76 (0.19)	0.65 (0.18)
	Ages 55–75 (n = 24) Mean (SD)	0.80 (0.20)	0.40 (0.25)	0.21 (0.18)	0.30 (0.22)	0.53 (0.25)	0.44 (0.26)
Age cohort contrasts:		−3.85**	−10.53**	−10.26**	−9.42**	−7.38**	−6.77**
Mode of interaction	Single						
contrasts	Auditory vocal	−29.47**					
	Center console	−37.62**	−16.84**				
	Center stack	−39.48**	−16.84**	3.02*			
	nBack	−16.00**	9.32**	20.72**	20.47**		
	SuRT	−23.46**	−0.01	12.92**	11.68**	−7.42**	
Mode of interaction	Single						
contrasts by age cohort	Auditory Vocal	−9.61**					
	Center console	−8.82**	−0.99				
	Center stack	−7.80**	1.84	2.25			
	nBack	−4.83**	3.53**	3.71**	2.05		
	SuRT	−4.00**	4.60**	4.61**	3.06*	0.85	

*$p<0.01$, **$p<0.001$.

task was slightly more delayed than to the Calling and Dialing task and that Reaction Time between the Calling and Dialing and Navigation Entry tasks also reached significance. Reaction time to the Single Task baseline was faster than to all other Task Types. Reaction time to the SuRT task was also faster than to the other Task Type (except Single Task), while Reaction Time to the nBack task did not differ from the Reaction Times to the four IVIS Task Types. Task Type contrasts by Age Cohort suggest that the effect of Age Cohort was smallest for the Single Task condition but that it did not differ between any of the other conditions.

Age Cohort by Mode of Interaction

The interaction between Age Cohort and Mode of Interaction also reached significance, $\chi^2(1, 2) = 33.05$, $p > 0.001$. Age Cohort contrasts indicated that the effect of Age Cohort on Reaction Time reached significance for each level of the Mode of Interaction. Mode of Interaction contrasts indicated that the Single Task and SuRT tasks differed from all other Modes of Interaction but that none of the other Modes of Interaction differed from each other. Mode of Interaction contrasts by Age Cohort indicated that the effect of Age Cohort was smallest in the Single Task baseline and similar in all other conditions.

Age Cohort by Mode of Interaction by Task Type

The three-way interaction between each of these factors was not significant, $\chi^2(1, 3, 2) = 1.69$, $p = 0.946$. This lack of interaction is clearly visible in **Figure 3** where a main effect of age is apparent with highly consistent effects of Mode of Interaction and Task Type.

DRT Hit Rate

Main Effects

Results indicated that Hit Rate differed by Age Cohort, $\chi^2(1) = 70.2$, $p < 0.001$ (Young: $M = 0.67$, $SD = 0.23$; Older: $M = 0.41$, $SD = 0.28$), with younger drivers detecting and accurately responding to the onset of the forward LED significantly more often than older drivers. Consistently across factors, older drivers failed to respond to the light stimulus more frequently resulting in lower Hit Rates. Additionally, there was a significant main effect of Task Type, $\chi^2(6) = 1232$, $p < 0.001$ (Audio Entertainment: $M = 0.44$, $SD = 0.28$; Calling and Dialing: $M = 0.51$, $SD = 0.29$; Text Messaging: $M = 0.50$, $SD = 0.27$; Navigation Entry: $M = 0.47$, $SD = 0.27$), and Mode of Interaction, $\chi^2(5) = 1577$, $p < 0.001$ (Auditory Vocal: $M = 0.55$, $SD = 0.28$; Center Console: $M = 0.38$, $SD = 0.26$; Center Stack: $M = 0.43$, $SD = 0.26$).

Age Cohort by Mode of Interaction

The interaction between Age Cohort and Mode of Interaction was also significant, $\chi^2(1, 5) = 116$, $p > 0.01$, suggesting that the effect of age cohort on Hit Rate depended on the Mode of Interaction. Age contrasts indicated that the effect of Age Cohort reached significance at all levels of the Mode of Interaction. Mode contrasts suggest that Hit Rates were highest in the Single Task condition, followed by the nBack task, then the SuRT and Auditory Vocal tasks, and lowest in the Center Stack and Center Console Modes of Interaction. Mode of Interaction contrasts by Age Cohort suggest that the effect of Age Cohort was similarly large for each of the Modes of Interaction, somewhat smaller in the nBack and SuRT tasks, and smallest in the Single Task.

Age Cohort by Mode of Interaction by Task Type

The three-way interaction between each of these factors was not significant, $\chi^2(1, 3, 2) = 1.33$, $p = 0.969$. This lack of interaction is visible in **Figure 4** where a main effect of age is apparent with highly consistent effects of Mode of Interaction and Task Type.

Subjective Workload

Main Effects

Results indicated that the composite TLX scores (average of the 21-point rating across each of the TLX subscales) differed by Age Cohort, $\chi^2(1) = 8.69$, $p = 0.003$ (Young: $M = 7.96$, $SD = 4.44$; Older: $M = 9.23$, $SD = 4.72$), with older drivers reporting IVIS task interactions to be more difficult than younger

TABLE 8 | Pairwise comparisons for NASA-TLX subjective responses as a function of task type and age cohort.

	Task type	Single	Audio entertainment	Calling and dialing	Text messaging	Navigation entry	nBack	SuRT
Means and SD	Ages 21–36 (n = 24) Mean (SD)	3.75 (2.65)	8.23 (4.33)	7.04 (4.15)	7.26 (3.88)	8.45 (4.32)	11.4 (3.73)	10.4 (4.25)
	Ages 55–75 (n = 24) Mean (SD)	5.08 (3.27)	10.3 (4.79)	8.03 (4.53)	8.68 (4.23)	10.0 (4.54)	12.1 (4.09)	10.1 (4.53)
Age cohort contrasts		2.51	4.29**	2.15	2.85*	3.16*	1.50	−0.11
Task type contrasts	Single							
	Audio entertainment	20.17**						
	Calling and dialing	12.93**	−8.86**					
	Text messaging	14.24**	−6.80**	1.85				
	Navigation entry	19.88**	−0.37	8.50**	6.45**			
	nBack	26.28**	10.19**	17.42**	15.61**	10.50**		
	SuRT	20.95**	4.01**	11.25**	9.54**	4.31**	−5.37**	
Task type contrasts by age cohort	Single							
	Audio entertainment	1.65						
	Calling and dialing	−0.75	−2.93					
	Text messaging	0.08	−1.87	0.98				
	Navigation entry	0.38	−1.56	1.37	0.36			
	nBack	−1.10	−2.91	−0.53	−1.32	−1.64		
	SuRT	−2.86*	−4.95**	−2.56	−3.32	−3.68	−1.75	

*$p<0.01$, **$p<0.001$.

drivers. Additionally, there were significant main effects for Task Type, $\chi^2(6) = 789$, $p < 0.001$ (Audio Entertainment: $M = 9.27$, $SD = 4.68$; Calling and Dialing: $M = 7.53$, $SD = 4.37$; Text Messaging: $M = 7.96$, $SD = 4.12$; Navigation Entry: $M = 9.22$, $SD = 4.49$), and Mode of Interaction, $\chi^2(5) = 958$, $p < 0.001$ (Auditory Vocal: $M = 7.33$, $SD = 4.17$; Center Console: $M = 9.70$, $SD = 4.27$; Center Stack: $M = 9.75$, $SD = 4.60$).

Age Cohort by Task Type

Analysis of the two-way interaction between Age Cohort and Task Type on Subjective Workload indicated that the interaction was not significant, $\chi^2(1, 6) = 28.4$, $p > 0.001$. Age Cohort Contrasts suggest that the effect of Age Cohort was not significant in any of the baseline Task Types but that it reached significance in all IVIS Task Types except the Calling and Dialing task. Task Type Contrasts suggest that drivers found the nBack task to be the most difficult, followed by the SuRT task, then the Navigation Entry and Audio Entertainment tasks, then the Calling and Dialing and Text Messaging tasks, followed by the Single Task. Task Type Contrasts by Age Cohort suggest that the effect of Age Cohort was somewhat larger in the Audio Entertainment task than the SuRT task but that the Age Cohort effect was otherwise indistinguishable.

Inspection of the data reveals a relatively consistent effect of age across each of the in-vehicle tasks.

Age Cohort by Mode of Interaction

The interaction between Age Cohort and Mode of Interaction was also significant, $\chi^2(1, 5) = 20.7$, $p > 0.001$. Age contrasts found no effect of age in any of the tasks. Mode of Interaction contrasts found that the subjective evaluation of workload differed between the Modes of Interaction. Specifically, drivers felt that Auditory Vocal tasks were easier to complete than tasks completed using the Center Stack or Touchscreen displays. All of which were reported to be easier than the nBack task. Mode of Interaction contrasts by Age Cohort suggest that the magnitude of the Age Cohort effect was smaller in the SuRT task than several of the other task interactions.

Age Cohort by Mode of Interaction by Task Type

The three-way interaction between each of these factors was not significant, $\chi^2(1, 3, 2) = 7.52$, $p = 0.275$. This lack of interaction is visible in **Figure 5** where a main effect of age is apparent with highly consistent effects of Mode of Interaction and Task Type.

DISCUSSION

This research investigated the unique challenges that older drivers face when completing several common tasks using the In-Vehicle Information System (IVIS) of six 2018 vehicles. Prior research has shown that compared to younger drivers, older drivers exhibit greater difficulty dividing attention between tasks. This has been shown both in general laboratory tasks and in driving. How these generally reported differences are manifest in interactions with real-world vehicle technologies has not been well studied. This research provides additional insight into the unique challenges faced by older drivers as they interact with modern in-vehicle technologies, by addressing two sets of previously unanswered questions related to IVIS use.

Q1: Do IVIS interaction demands differ for older and younger drivers? If so, how?

Results suggest that, compared to younger drivers, older drivers experienced increased workload when interacting with IVIS. Older adults were slower to respond to the DRT stimuli (higher cognitive demand), were more likely to fail to respond to the forward LED (higher visual demand) and required more time to complete all tasks (increased exposure). Measures of cognitive, visual, and temporal demand for older and younger drivers indicated nearly identical patterns between all conditions (e.g.,

TABLE 9 | Pairwise comparisons for NASA-TLX subjective responses as a function of mode of interaction and age cohort.

	Mode of interaction	Single	Auditory vocal	Center console	Center stack	nBack	SuRT
Means and SD	Ages 21–36 (n = 24) Mean (SD)	3.75 (2.65)	6.57 (3.79)	9.02 (4.24)	8.98 (4.33)	11.4 (3.73)	10.4 (4.25)
	Ages 55–75 (n = 24) Mean (SD)	5.08 (3.27)	8.12 (4.39)	10.4 (4.2)	10.6 (4.74)	12.1 (4.09)	10.1 (4.53)
Age cohort contrasts		2.55	3.42	2.93	3.21	1.53	−0.10
Mode of interaction contrasts	Single						
	Auditory vocal	13.60**					
	Center console	20.29**	11.70**				
	Center stack	23.46**	15.56**	0.69			
	nBack	27.00**	20.54**	8.29**	8.66**		
	SuRT	21.52**	13.61**	2.46	2.12	−5.52**	
Mode of interaction contrasts	Single						
by age cohort	Auditory vocal	0.43					
	Center console	0.29	−0.09				
	Center stack	0.34	−0.10	0.01			
	nBack	−1.13	−1.86	−1.49	−1.68		
	SuRT	−2.93*	−4.14**	−3.40**	−3.82**	−1.80	

*$p<0.01$, **$p<0.001$.

conditions that resulted in high workload for younger drivers also resulted in high workload for older drivers). However, measured workload for older drivers was consistently higher than for younger drivers.

Q2: Are some interfaces more difficult for older drivers to use? If so, why?

Older drivers had an especially difficult time maintaining visual attention to the forward roadway during secondary task interactions (as quantified by Hit Rate to the forward LED DRT stimulus – See Strayer et al., 2017), especially when completing IVIS tasks. IVIS task interactions are demanding in general but especially so for older drivers. Older drivers may benefit from interface designs that promote their continued visual attention on or near the forward roadway (e.g., careful placement of physical controls and dials, screen placement in-line with forward vision, use of voice controls, etc.).

Summary of Results

Task Completion Time

Task Completion Time is an important facet of driver workload as it represents exposure. All other demands being equal, tasks that require longer to complete will result in greater distraction potential (see Strayer et al., 2015). When compared to younger drivers, older drivers required more time to complete tasks in all experimental conditions. Some noteworthy differentiation occurred between Task Type and Mode of Interaction.

- On average, both younger and older drivers completed the Audio Entertainment and Calling and Dialing Task Types in less than the 24-s time reference. Navigation Entry and Text Messaging Task Types each required significantly more than 24-s for both younger and older drivers.
- Tasks were completed more quickly using the center stack display by both age groups. Older drivers, on average, required more than 24-s to complete the Auditory Vocal and Center Console tasks.

Reaction Time to the DRT

Reaction Time to the DRT is a reliable and valid indicator of cognitive demand (cf. International Organization for Standardization, 2015). Interestingly, Reaction Time to each of the tasks was insensitive to differences in Task Type, and Mode of Interaction. Reaction Time did, however, greatly differ between the Single Task baseline and any of the other tasks and was longer for older drivers. The stability of Reaction Time during IVIS interactions regardless of Task Type and Mode of Interaction, suggests that at least with some tasks, cognitive demand remains constant throughout task engagement. Our analysis found that:

- Older drivers were slower than younger drivers, across all Task Types.
- Older drivers were slower than younger drivers for all Modes of Interaction.

Hit Rate to the DRT

Hit Rate to the DRT indicated that older drivers had a more difficult time dividing visual attention between driving and secondary tasks. Similar to Reaction Time, the effect of Age Cohort was consistent for each of the four Task Types, and across each of the three Modes of interaction. Our analysis found that:

- Older drivers hit rate was greatly reduced across all Task Types compared with younger drivers.
- Older drivers were less likely to respond to the forward LED across each of the three Modes of Interaction.

Subjective Measures

- All drivers found the Calling and Dialing and Text Messaging Task Types to be less demanding than the Navigation Entry and Audio Entertainment Task Types. Overall, older drivers reported IVIS task interactions to be more demanding than younger drivers.
- Both older and younger drivers reported that voice commands were easier to use than the Center Console or Center Stack controls. Older drivers reported that all

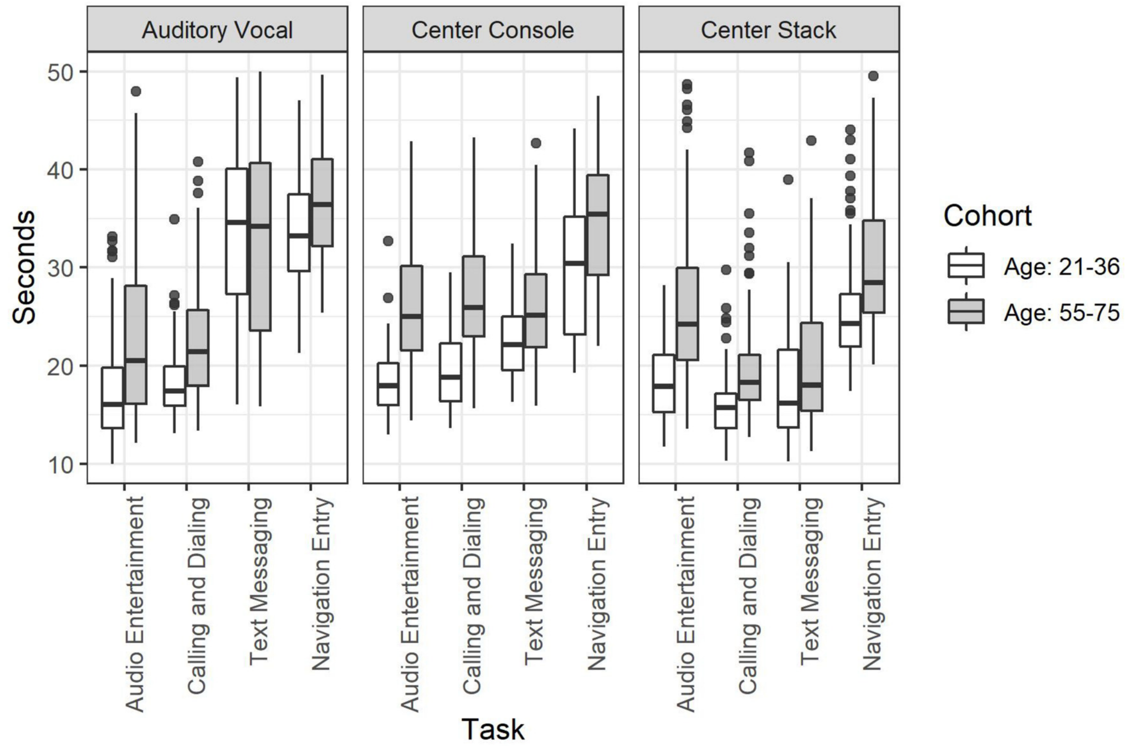


FIGURE 2 | Task completion time for the full factorial of age cohort by mode of interaction by task.

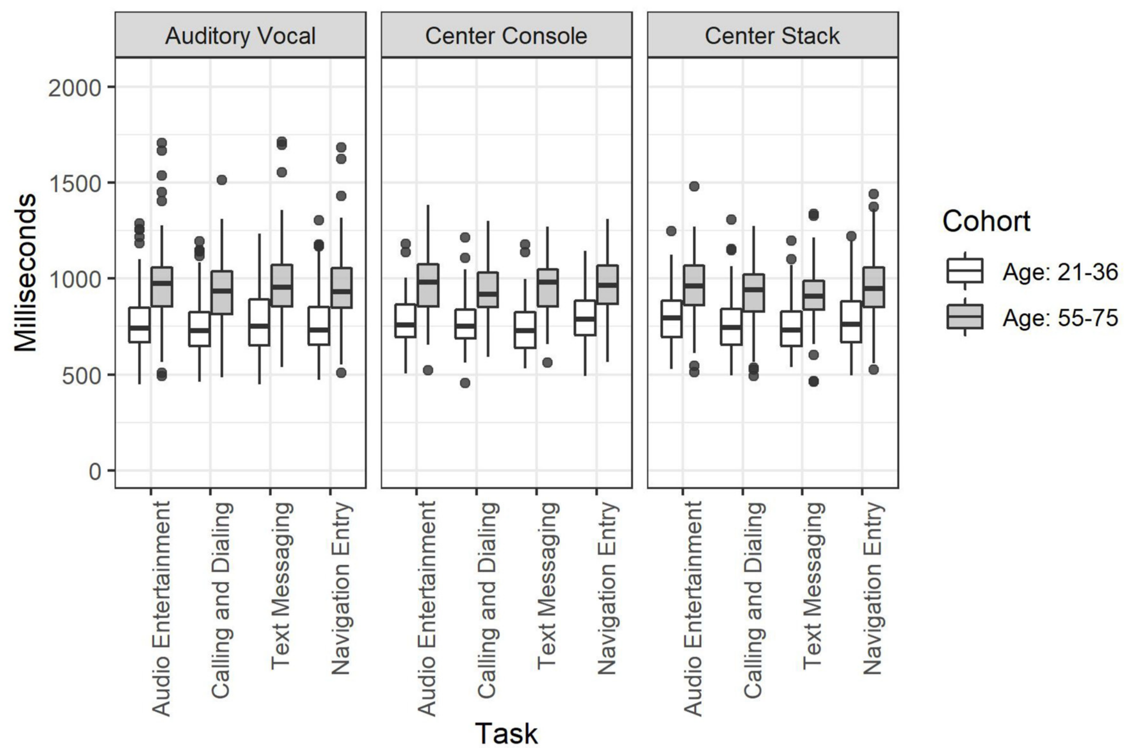


FIGURE 3 | Detection response task reaction time for the full factorial of age cohort by mode of interaction by task.

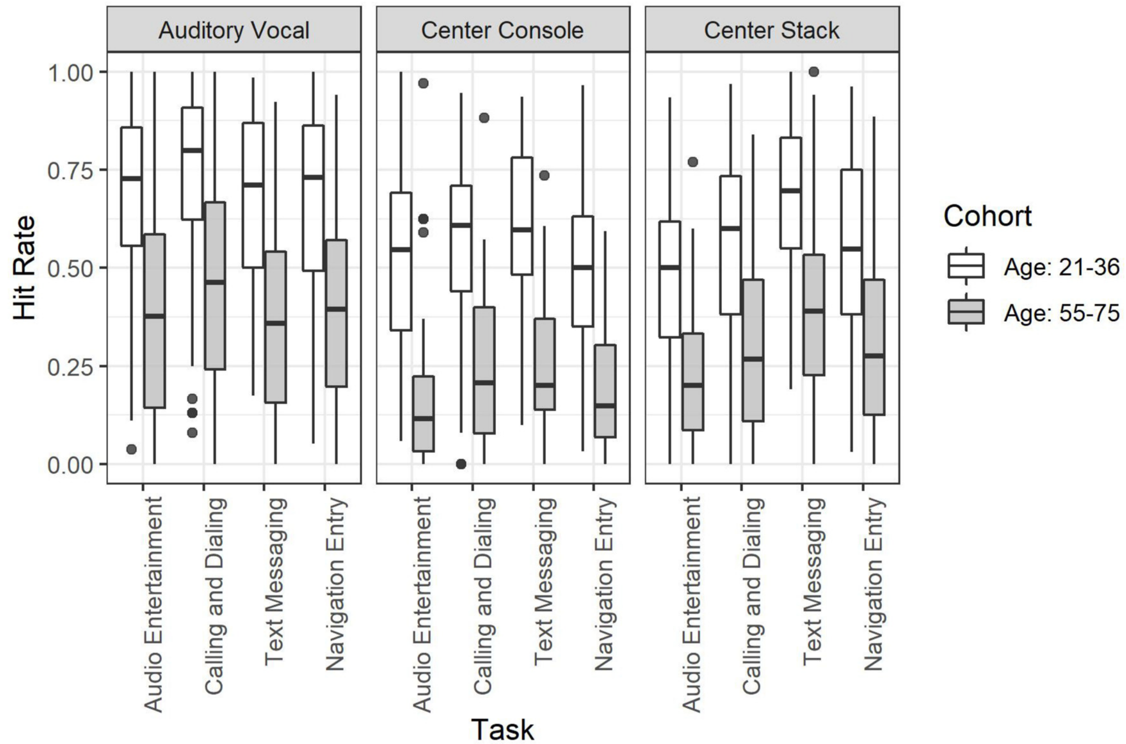


FIGURE 4 | Detection response task hit rate for the full factorial of age cohort by mode of interaction by task.

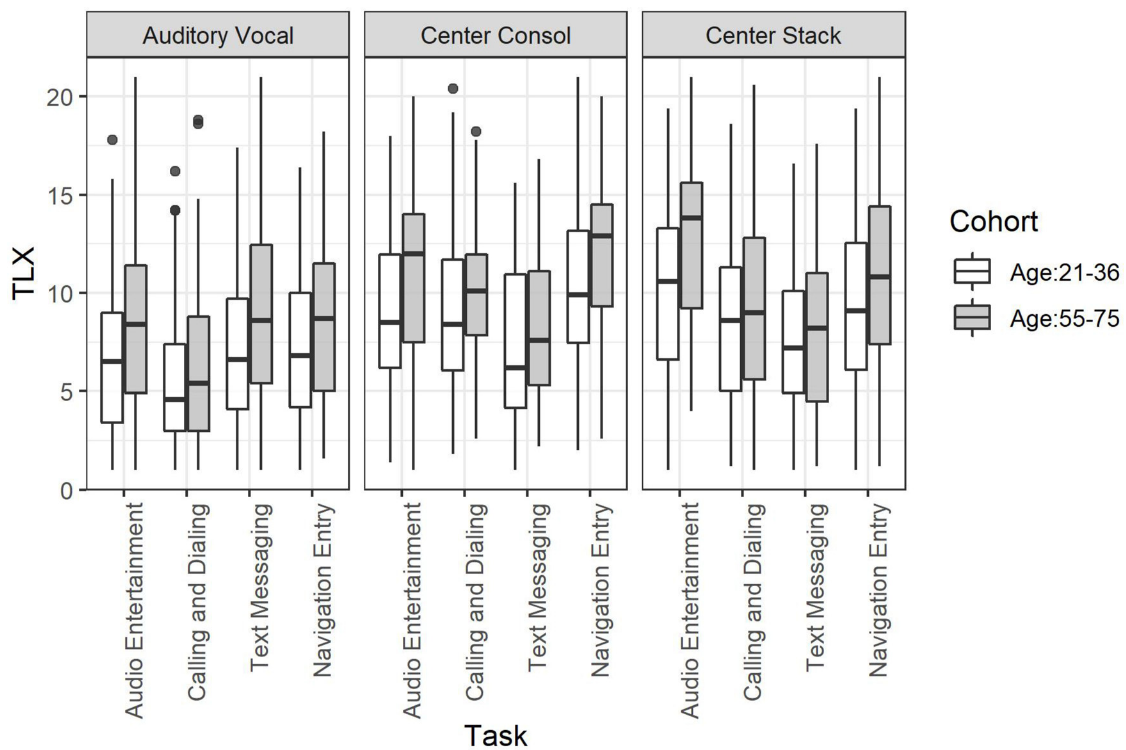


FIGURE 5 | Task load index subjective workload for the full factorial of age cohort by mode of interaction by task.

Modes of Interaction were more demanding than reported by younger drivers.

- Comments from both groups were primarily negative in tone. The similarity in comments made by drivers emphasizes a need for improvement of these in-vehicle systems for all drivers across the age range.

Further Discussion

The time required to complete tasks is a simple and effective way to evaluate general task demands. This has been noted by other researchers when applied to visual-manual interactions with IVIS systems (e.g., Green, 1999). We found that all tasks imposed cognitive and visual demands. Not surprisingly however, hit rates to the forward LED were even lower for tasks that diverted driver's eyes from the road. Thus, an assessment of task completion time with a measure of visual attention may sufficient to understand driver workload when completing the discrete tasks with IVIS.

Findings from this research suggest that workload estimates derived from younger drivers may underestimate the workload experienced by older drivers. In the Visual Manual Distraction Guidelines published by National Highway Transportation Safety Administration [NHTSA] (2013), a participant sampling strategy that includes drivers from 4 age groups is recommended. Given the consistent performance differences between younger and older drivers, we recommend that future testing give higher priority to evaluating older users. Systems that older adults find easy to use will also be usable by younger adults; however, the converse may not always be the case.

A logical and potentially incorrect generalization of these findings would be to assume that poorly performing Task Types or Modes of Interaction would result in increased on-road distraction. While this may be true, it may also be the case that drivers naturally refrain from activities that are complex, error prone, or slow to complete. Frustration arising from these tasks may cause drivers to seek out simpler ways of IVIS interaction. For example, voice recognition systems in vehicles show promise, however, driver usage of these systems continues to be low (Viita, 2014). The reason being that they often require the use of precise keywords spoken in a very specific, and rigid order. The result may be an interaction that is more complex, frustrating, and distracting than the same action completed using the touchscreen on the center-stack.

Results from this evaluation should, therefore, be interpreted as a measure of the user experience, or distraction potential (Ranney et al., 2009; Lee and Strayer, 2004), and not necessarily a reflection of the level of on-road distraction that would be expected from these Task Types and Modes of Interaction. Paradoxically, it may be the case that the most difficult and demanding systems evaluated in this research are also the least likely to result in driver distraction because they are not used. Furthermore, there is the possibility that the in-vehicle information systems that are the most cumbersome to use may ultimately result in users abandoning the IVIS in lieu of their personal cell phone to achieve the same tasks. This captures what has been described as the Usability Paradox (Lee and Strayer, 2004) wherein distraction may increase with usability. Likewise, poorly designed systems may discourage use and therefore decrease distraction potential overall. Complex user requirements may pose unnecessary system-based barriers, which could result in circumstances where older drivers are faced with no good options.

Design Recommendations

Compared to younger drivers, older drivers in this research exhibited slower Reaction Time, decreased Hit-Rate, longer Task Completion Time, and reported higher task demand when interacting with IVIS. These findings suggest that, at a minimum, older drivers should be included in a Universal Design validation as their interactions with vehicle technologies may significantly differ from that of younger drivers (Czaja et al., 2009). Relevant principles of Universal Design for vehicle manufacturers include Equity, Flexibility, Simplicity, Perceptibility, Error Recovery, and Accessibility (Farage et al., 2012). These principles may provide a framework for improvement of IVIS design. Clearly, an emphasis on simplicity would benefit drivers of all ages (Farage et al., 2012).

Both older and younger drivers exhibited difficulty dividing attention between tasks presented on the touch screen and the forward roadway. Data from this study suggests that a center console interactions are especially cumbersome for older drivers. Care should be taken to help drivers maintain their visual attention on the forward roadway without introducing unnatural interfaces that may cause interference with safe driving. While voice commands may help to reduce many of the potential problems of other interface types, they will only be used by drivers if the systems accurately process requests in a timely fashion. Even so, auditory vocal interactions imposed a relatively high level of cognitive demand on drivers. No interface is demand free and all interactions with vehicle technologies should be carefully considered and restricted when reasonable.

CONCLUSION

This research investigated the challenges faced by younger and older drivers as they completed several common tasks using the In-Vehicle Information System (IVIS) of a representative sample of six 2018 vehicles. Compared to younger drivers, older drivers exhibited significant increases in cognitive and visual workload when completing IVIS tasks. Older drivers had difficulty dividing their visual attention between IVIS tasks and the forward roadway. In some cases, older drivers responded to fewer than 25% of LED illuminations presented on the forward windscreen.

Older drivers also required significantly more time than younger drivers to complete all task interactions. An analysis of subjective workload found that drivers were generally aware of task demands but may have underestimated their actual workload, as quantified in the other measures. Comments provided by drivers after each task interaction suggested that both older and younger drivers shared similar concerns about the experience of modern IVIS. Results from this research suggest that current versions of IVIS are demanding and difficult to use, especially for older drivers. For drivers to fully realize the potential benefits of current and future vehicle technologies, a renewed focus on accessible design is required.

TERMS AND NOMENCLATURE

Detection Response Task (DRT)	The DRT is an International Standards Organization protocol (International Organization for Standardization, 2015) for measuring attentional effects of cognitive demand in driving. In this research, a vibrotactile device emitted a small vibration stimulus, similar to a vibrating cell phone or an LED light stimulus changing color from orange to red. These changes cued the participant to respond as quickly as possible by pressing the microswitch attached to a finger against the steering wheel. DRT reaction time increases and hit rate decreases as the workload of the driver increases.
In-vehicle information system (IVIS)	The collection of features and functions in vehicles that allow motorists to complete tasks unrelated to driving while operating the vehicle. In this report, the terms IVIS and system are used interchangeably. The IVIS features we tested involved up to four Task Types (see below) and up to three Modes of Interaction (see below).
Modes of Interaction	The way a user interacts with an IVIS to perform a task. Modes of Interaction were categorized into three types: Voice Commands, Center Stack, and Center Console. In this report, Mode and Mode of interaction is used interchangeably.
NASA TLX	A questionnaire-based metric assessing the subjective workload of the driver. The TLX assesses mental demand, physical demand, temporal demand, performance, effort, and frustration.
nBack task	The nBack task presented a prerecorded series of numbers ranging from 0 to 9 at a rate of one digit every 2.25 seconds. Participants were instructed to say out loud the number that was presented two trials earlier in the sequence. The nBack task places a high level of cognitive demand on the driver without imposing any visual/manual demands and was used as a high workload reference task.
Primary driving task	Activities that the driver must undertake while driving including navigating, path following, maneuvering, and avoiding obstacles.
Reference task	A task used for the purpose of comparing different tests or test results across vehicles or systems.
Single-task baseline	When the driver is performing the primary driving task (i.e., driving) without the addition of workload imposed by IVIS interactions.
Secondary-task	A non-driving related additional task.
SuRT task	The variant of the Surrogate Reference Task (SuRT, ISO TS 14198) used in this report required participants to use their finger to touch the location of target items (larger circles) presented in a field of distractors (smaller circles) on an iPad Mini tablet computer that was mounted in a similar position in all the vehicles. The SuRT task places a high level of visual/manual demand on the drivers because they must look at and touch the display to perform the task. The SuRT task served as a reference for the visual/manual demands associated with performing IVIS interactions.
Task completion time	The time to complete a task. Task completion time was defined as the time from the moment participants first initiated an action to the time when that action had terminated, and the participant said, "done." When assessed using the visual occlusion methodology, the NHTSA guidelines provide an implicit upper limit of 24 seconds of total task time. While originally intended for visual/manual tasks, these guidelines provide a reasonable upper limit for task durations of any Mode or Task Type.
Task Type	Tasks were categorized into one of four Task Types: Audio Entertainment, Calling and Dialing, Text Messaging, and Navigation, depending on vehicle capabilities. These Task Types were completed via different Modes equipped in each vehicle for each interaction.
Visual demand	The visual workload associated with the performance of a task. This would include the structural interference associated with taking the eyes off the forward roadway as well as the central interference in visual processing that arises from cognitive demand. In this report, we refer to the visual demand associated with performing IVIS tasks with different Modes of Interaction when the vehicle is in motion.
Visual reference task	A variant of the SuRT task (see above) served as the visual reference task in the current research.
Voice Commands	The Voice Commands method in which users communicate with the IVIS via voice recognition and structured commands. Voice Commands are aimed toward hands free interactions but may incorporate some visual manual interactions such as using steering wheel controls for activation. Voice Commands are one of the three Modes of Interaction evaluated in this research.
Workload	The aggregate of cognitive, visual, and manual demands on the driver. A motorist's workload reflects a combination of demands from the primary task of driving and any secondary tasks performed by the driver. The terms demand and workload are used interchangeably in this report and we develop separate metrics for cognitive workload and visual workload.

DATA AVAILABILITY STATEMENT

The datasets for this study can be found on the Utah Center for Driver Safety and Technology github repository https://github.com/utahcdst/Aging-Report-Frontiers.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The University of Utah Institutional Review Board (IRB). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was funded by a grant from the AAA Foundation for Traffic Safety, Washington, DC,

United States. An on-line technical report (pre-print) based on the data presented in this paper can be found at https://aaafoundation.org/wp-content/uploads/2019/07/2019-AAAFTS-Utah-IVIS-Age-Difference-Full-Report.pdf.

REFERENCES

Albert, G., Lotan, T., Weiss, P., and Shiftan, Y. (2018). The challenge of safe driving among elderly drivers. *Healthc. Technol. Lett.* 5, 45–48. doi: 10.1049/htl.2017.0002

Angell, L., Auflick, J., Austria, P. A., Kochhar, D., Tijerina, L., Biever, W., et al. (2006). *Driver Workload Metrics Task 2 Final Report*. Washington, DC: DOT HS.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Statist. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Castro, S., Cooper, J., and Strayer, D. (2016). "Validating two assessment strategies for visual and cognitive load in a simulated driving task," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Los Angeles, CA: Sage CA), 1899–1903. doi: 10.1177/1541931213601432

Cerella, J. (1985). Information processing rates in the elderly. *Psychol. Bull.* 98, 67–83. doi: 10.1037/0033-2909.98.1.67

Cerella, J., Poon, L. W., and Williams, D. M. (1980). "Age and the complexity hypothesis," in *Aging in the 19SOs: Psychological Issues*, ed. L. Poon (Washington. DC: American Psychological Association), 332–340. doi: 10.1037/10050-024

Clapp, W. C., Rubens, M. T., Sabharwal, J., and Gazzaley, A. (2011). Deficit in switching between functional brain networks underlies the impact of multitasking on working memory in older adults. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7212–7217. doi: 10.1073/pnas.1015297108

Cooper, J. M., Castro, S. C., and Strayer, D. L. (2016). "Extending the detection response task to simultaneously measure cognitive and visual task demands," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Los Angeles, CA.

Cooper, J. M., Ingebretsen, H., and Strayer, D. L. (2014). *Measuring Cognitive Distraction in the Automobile IIa: Mental Demands of Voice-Based Vehicle Interactions with OEM Systems*. Washington, DC: AAA Foundation for Traffic Safety.

Craik, F. I. M. (1977). "Age differences in human memory," in *Handbook of the psychology of aging*, eds J. Birren, and K. W. Schaie (New York, NY: Van Nostrand Reinhold), 384–420.

Czaja, S. J., Rogers, W. A., Fisk, A. D., Charness, N., and Sharit, J. (2009). *Designing for Older Adults: Principles and Creative Human Factors Approaches*. Boca Raton, FL: CRC press.

Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., et al. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proc. Natl. Acad. Sci. U.S.A.* 113, 2636–2641. doi: 10.1073/pnas.1513271113

Drews, F. A., Yazdani, H., Godfrey, C., Cooper, J. M., and Strayer, D. L. (2009). Text messaging during simulated driving. *Hum. Fact.* 51, 762–770. doi: 10.1177/0018720809353319

Farage, M. A., Miller, K. W., Ajayi, F., and Hutchins, D. (2012). Design principles to accommodate older adults. *Glob. J. Health Sci.* 4:2.

Fisher, D. L., and Strayer, D. L. (2014). Modeling situation awareness and crash risk. *Ann. Adv. Automot. Med.* 5, 33–39.

Graham, J. W., Taylor, B. J., Olchowski, A. E., and Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychol. methods* 11:323. doi: 10.1037/1082-989x.11.4.323

Green, P. (1999). "The 15-second rule for driver information systems," in *Proceedings of the ITS America Ninth Annual Meeting* (Washington, DC: Intelligent Transportation Society of America).

Hart, S. G., and Staveland, L. E. (1988). Development of NASA-TLX (task load index): results of empirical and theoretical research. *Adv. Psychol.* 52, 139–183. doi: 10.1016/s0166-4115(08)62386-9

Hartley, A. (1992). "Attention," in *Handbook of Aging And Cognition*, eds F. Craik, and T. Salthouse (Hillsdale, NJ: Erlbaum), 3–49.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01154/full#supplementary-material

Hartley, A. A., and Little, D. M. (1999). Age-related differences and similarities in dual-task interference. *J. Exp. Psychol. Gen.* 128:416. doi: 10.1037/0096-3445.128.4.416

International Organization for Standardization (2012). *Road Vehicles – Ergonomic Aspects Of Transport Information And Control Systems – Calibration Tasks For Methods Which Assess Driver Demand Due To The Use Of In-Vehicle Systems (Rep. ISO* 14198). Geneva: International Organization for Standardization.

International Organization for Standardization (2015). *Road Vehicles – Transport Information and Control Systems – Detection Response Task (DRT) for Assessing Attentional Effects of Cognitive Load in Driving (Rep. ISO 17488)*. Geneva: International Organization for Standardization.

Kramer, A. F., and Larish, J. L. (1996). Aging and dual-task performance. *Aging Skill. Perform.* 7, 83–112.

Lee, J. D., and Strayer, D. L. (2004). Preface to the special section on driver distraction. *Hum. Fact.* 46, 583–586. doi: 10.1518/hfes.46.4.583.56811

Little, T. D., and Rhemtulla, M. (2013). Planned missing data designs for developmental researchers. *Child Dev. Perspect.* 7, 199–204. doi: 10.1111/cdep.12043

McDowd, J. M., and Shaw, R. J. (2000). "Attention and aging: a functional perspective," in *The Handbook of Aging And Cognition*, eds F. I. M. Craik, and T. A. Salthouse (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 221–292.

Mehler, B., Reimer, B., and Dusek, J. A. (2011). *MIT AgeLab Delayed Digit Recall Task (n-back) [White paper]*.

National Highway Transportation Safety Administration [NHTSA] (2013). *Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices: Notice of Federal Guidelines (Federal Register 78:81 (26 April 2013)*. Washington, DC: NHTSA.

National Safety Council [NSC] (2017). *Injury Facts: 2017 Edition*. Chicago, IL: NSC.

Olson, R. L., Hanowski, R. J., Hickman, J. S., and Bocanegra, J. (2009). *Driver Distraction In Commercial Vehicle Operations. DOT F 1700.7 (8-72)*. Washington, DC: U.S. Department of Transportation.

Perez, M., Owens, J., Viita, D., Angell, A., Ranney, T. A., Baldwin, G. H. S., et al. (2013). *Radio Tuning Effects On Visual And Driving Performance – Simulator And Test Track Studies*. Report No. DOT HS 811 781, Washington, DC: National Highway Traffic Safety Administration.

R Core Team (2018). *R: A Language And Environment For Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ranney, T. A., Baldwin, G. H., Vasko, S. M., and Mazzae, E. N. (2009). *Measuring Distraction Potential Of Operating In-Vehicle Devices (No. HS-811 231)*. Washington, DC: NHTSA.

Ranney, T. A., Garrott, R. W., and Goodman, M. J. (2000). *NHTSA Driver Distraction Research: Past, Present, and Future. Driver Distraction Internet Forum*. Available online at: https://www.researchgate.net/profile/Elizabeth_Mazzae/publication/255669554_NHTSA_Driver_Distraction_Research_Past_Present_and_Future/links/0a85e53ab86eb9ea74000000/NHTSA-Driver-Distraction-Research-Past-Present-and-Future.pdf (accessed May 18, 2020).

Regan, M. A., Hallett, C., and Gordon, C. P. (2011). Driver distraction and driver inattention: definition, relationship and taxonomy. *Accident Anal. Prevent.* 43, 1771–1781. doi: 10.1016/j.aap.2011.04.008

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychol. Sci.* 103, 403–428. doi: 10.1037/0033-295x.103.3.403 doi: 10.1037/0033-295x.103.3.403

Shutko, J., and Tijerina, L. (2006). Eye glance behavior and lane exceedences during driver distraction. *Paper Presented at the given Driver Metrics Workshop*, Ottawa.

Simons-Morton, B., Guo, F., Klauer, S. G., Ehsani, J. P., and Pradham, A. K. (2015). Keey your eyes on the road: Young drivers crash risk increases according to duration of distraction. *J. Adolesc. Health* 54, S61–S67.

Strayer, D., Cooper, J., McCarty, M., Getty, D., Wheatley, C., Motzkus, C., et al. (2018). *Visual and Cognitive Demands Of Using Apple's Carplay, Google's Android Auto And Five Different Oem Infotainment Systems*. Washington, DC: AAA Foundation for Traffic Safety.

Strayer, D. L., Cooper, J. M., Goethe, R., McCarty, M., and Biondi, F. (2017). *Visual and Cognitive Demands of Using In-Vehicle Infotainment Systems*. Washington, DC: AAA Foundation for Traffic Safety.

Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J., Medeiros-Ward, N., and Biondi, F. (2013). *Measuring Cognitive Distraction In The Automobile*. Washington, D.C: AAA Foundation for Traffic Safety.

Strayer, D. L., Cooper, J. M., Turrill, J. M., Coleman, J. R., and Hopman, R. J. (2015). *Measuring Cognitive Distraction In The Automobile Iii: A Comparison Of 10 2015 In-Vehicle Information Systems*. Washington, DC: AAA Foundation for Traffic Safety.

Strayer, D. L., and Fisher, D. L. (2016). SPIDER: a model of driver distraction and situation awareness. *Hum. Fact.* 58, 5–12.

Strayer, D. L., Turrill, J., Coleman, J., Ortiz, E., and Cooper, J. M. (2014). *Measuring Cognitive Distraction in the Automobile: II. Assessing In-vehicle Voice-based Interactive Technologies*. Washington, DC: AAA Foundation for Traffic Safety.

Strayer, D. L., Watson, J. M., and Drews, F. A. (2011). "Cognitive distraction while multitasking in the automobile," in *The Psychology of Learning and Motivation: Advances in Research and Theory*, ed. B. Ross (San Diego, CA: Academic Press), 29–58. doi: 10.1016/b978-0-12-385527-5.00002-4 doi: 10.1016/b978-0-12-385527-5.00002-4

Vaportzis, E., Clausen, M. G., and Gow, A. J. (2017). Older adults perceptions of technology and barriers to interacting with tablet computers: a focus group study. *Front. Psychol.* 8:1687. doi: 10.3389/fpsyg.2017.01687

Viita, D. (2014). *Usage and Satisfaction of Voice Recognition and Touchscreen HMI*. Available online at: https://www.strategyanalytics.com/access-services/automotive/in-vehicle-ux/reports/report-detail/usage-and-satisfaction-of-voice-recognition-and-touchscreen-hmi?Related (accessed December 12, 2018).

Watson, J. M., Miller, A. E., Lambert, A., and Strayer, D. L. (2011). "The magical letters P,F,C, and sometimes U: the rise and fall of executive attention with the development of prefrontal cortex," in *Handbook of Lifespan Psychology*, eds K. Fingerman, C. Berg, T. Antonucci, and J. Smith (Berlin: Springer), 407–436.

World Health Organization [WHO] (2011). *Mobile Phone Use: A Growing Problem Of Driver Distraction*. Geneva: World Health Organization.

Winter, B. (2013). *A Very Basic Tutorial For Performing Linear Mixed Effects Analyses*. Available online at: https://jontalle.web.engr.illinois.edu/MISC/lme4/bw_LME_tutorial.pdf (accessed May 18, 2020).

Conflict of Interest: The Detection Response Task (DRT) devices used in this research were assembled by Red Scientific Inc., a company which is owned and operated by JC. Device specifications used for DRT assembly adhered to ISO 17488, as did all data handling and cleaning procedures.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 18 August 2020
doi: 10.3389/fpsyg.2020.01889

Assessment of Drivers' Perceptions of Connected Vehicle–Human Machine Interface for Driving Under Adverse Weather Conditions: Preliminary Findings From Wyoming

Mohamed M. Ahmed†, Guangchuan Yang† and Sherif Gaweesh†*

Department of Civil and Architectural Engineering, University of Wyoming, Laramie, WY, United States

OPEN ACCESS

Edited by:
Haneen Farah,
Delft University of Technology,
Netherlands

Reviewed by:
Pavlo Bazilinskyy,
Delft University of Technology,
Netherlands
Tuomo Kujala,
University of Jyväskylä, Finland

***Correspondence:**
Mohamed M. Ahmed
mahmed@uwyo.edu;
mahmed@knights.ucf.edu

†ORCID:
Mohamed M. Ahmed
orcid.org/0000-0002-1921-0724
Guangchuan Yang
orcid.org/0000-0001-8150-489X
Sherif Gaweesh
orcid.org/0000-0001-7977-6378

Specialty section:
This article was submitted to
Performance Science,
a section of the journal
Frontiers in Psychology

Received: *16 February 2020*
Accepted: *08 July 2020*
Published: *18 August 2020*

Citation:
Ahmed MM, Yang G and Gaweesh S (2020) Assessment of Drivers' Perceptions of Connected Vehicle–Human Machine Interface for Driving Under Adverse Weather Conditions: Preliminary Findings From Wyoming. Front. Psychol. 11:1889. doi: 10.3389/fpsyg.2020.01889

Connected vehicle (CV) technology aims to improve drivers' situational awareness through audible and visual warnings displayed on a human–machine interface (HMI), thus reducing crashes caused by human error. This paper developed a driving simulator test bed to assess the readability and usefulness of the Wyoming CV applications. A total number of 26 professional drivers were recruited to participate in a driving-simulator study. Prior to driving the simulator, the participants were trained on both the concept of CV technology and the developed CV applications as well as the operation of the driving simulator. Three driving simulation scenarios were designed. For each scenario, participants drove two times: one with the HMI turned on and another one with the HMI turned off. After driving the simulator, a comprehensive revealed-preference survey was employed to collect the participants' perceptions of CV technology and Wyoming CV applications. Results show that the Wyoming CV applications were most favored under poor-visibility driving conditions. Among the Wyoming CV applications, forward collision warning and rerouting applications were experienced as the most useful. Approximately 89% of the participants stated that the Wyoming CV applications provided them with improved road condition information and increased their experienced safety while driving; 65% of the participants stated the CV applications and the HMI did not introduce distraction from the primary task of driving. Finally, this paper concludes that the design of CV HMI needs to balance a trade-off between the readability of the warnings and drivers' capability to safely recognize and timely respond to the received warnings.

Keywords: Wyoming connected vehicle pilot, human–machine interface, driver behavior, human factors, driving simulator

INTRODUCTION

In the United States, Interstate 80 (I-80) is a major corridor for east–west freight movement and passenger travel in the country. The 402-mile I-80 freeway corridor in Wyoming is considered to be a unique freeway corridor because it is all located above 6,000 feet (1,829 m) in elevation and with very few alternate routes. As a mountainous rural freeway, the total traffic volume is not high; nevertheless, the commercial truck volume makes up 30–55% of the total traffic flow

(WYDOT, 2017). As a consequence of Wyoming's adverse winter weather conditions, such as snowstorms, strong crosswinds, icy road surface, and low visibility from blizzard and the presence of work zones, there have been remarkable traffic crash records along I-80 in Wyoming, which resulted in fatalities, road closures, and tremendous economic loss (WYDOT, 2017). In reality, it was found that more than 90% of motor vehicle crashes were attributed at least in part to human error (Nhtsa, 2015). With the booming of vehicle technology, connected vehicle (CV) technology has been widely introduced into the market at a fast pace. CV technology is designed to improve drivers' awareness of hazards and situations they cannot even see through vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), and infrastructure-to-vehicle (I2V) dedicated short-range communication (DSRC) technologies so that proactive reactions could be made to avoid potential crashes (Shladover, 2018). A handful of studies have been conducted to assess the benefits of CV applications on reducing traffic collisions (Jeong et al., 2014; Dey et al., 2016; Olia et al., 2016; Zulkefli et al., 2017). In general, these studies demonstrate that CV technology has great potential in reducing the probability of traffic collisions on various transportation facilities and under different weather and traffic conditions.

With consideration of the challenging driving conditions on I-80 in Wyoming, the United States Department of Transportation (USDOT) selected Wyoming to develop, test, and deploy a suite of CV applications that utilize V2V, V2I, and I2V real-time communication technologies to provide warnings and advisories regarding various road conditions to heavy truck and light vehicle drivers (Gopalakrishna et al., 2016). The CV applications developed in the Wyoming CV pilot are expected to enable CV drivers to have awareness of upcoming hazardous traffic and roadway situations; therefore, drivers could make proactive reactions to avoid potential crashes. One of the key components of the Wyoming CV system is the on-board human–machine interface (HMI), which delivers received real-time geospecific basic safety messages (BSMs) and traveler information messages (TIMs) to drivers. Nevertheless, to date, there still lacks a clear understanding of how drivers recognize and response to the notifications displayed on the CV HMI. In fact, a well-designed HMI has the potential to provide CV drivers with proactive decision-making supports so that CV drivers could more timely respond to an imminent hazardous traffic condition and, thus, reduce the probability of involvement in traffic collisions. However, inappropriate integration of various CV warnings and advisories may mislead, distract, or even disturb drivers from their normal driving task (Li et al., 2017; Talamonti et al., 2017). These adverse effects are particularly significant during high-workload situations or driving under inclement weather and road surface conditions.

In this regard, this research aims to assess the effectiveness of the CV applications developed by the Wyoming CV Pilot Development Program. The assessment methodologies employed in this study have two steps. First, this research developed a CV driving-simulator test bed to simulate different traffic and weather conditions on I-80 in Wyoming. Then, professional snowplow truck and highway maintenance vehicle drivers from the Wyoming Department of Transportation (WYDOT) were invited to participate in the developed driving-simulator study. After experiencing the Wyoming CV applications in a simulated environment, each participant was requested to finalize a reveal-preference questionnaire survey, in which the participant provided perceptions of effectiveness of the CV applications as well as the visual distractions caused by the Wyoming CV HMI.

The remainder of this paper is organized as follows: Section "Literature Review" presents a review of the literature regarding HMI design and evaluation. Section "Description of Wyoming CV Applications and HMI" describes the functions of the Wyoming CV applications and HMI display layout. Section "Assessment of Wyoming CV HMI" documents the development of driving-simulator testing scenarios and participants' evaluation of the CV applications after driving the developed simulation scenarios; finally, preliminary findings and discussion of the lessons learned from this pilot study are listed in Section "Concluding Remarks."

LITERATURE REVIEW

HMI Display Design

In current practice, various modalities have been employed for the development of HMI display. In general, these modalities can be classified into four categories (Péter et al., 2017): mechanical, acoustic, visual, and haptic interfaces.

Mechanical interfaces require a mechanical interaction from the driver, which could be pressed by hand, finger, or foot; pulled, slid, or rotated by hand; or touched by hand or finger. The interfaces may include pedal, steering wheel, button, switch, stalk, slider, and controller knobs. Some advanced practices have been developed in these ordinary interfaces to enhance the driving performance on roads, such as electronic throttle control, electrical braking systems, electrical steering systems, etc. (Wang et al., 2016). Acoustic interfaces are common output interfaces because an acoustic (or auditory) interface does not require drivers to take off their eyes off the road; hence, it could be considered a safer modality than the visual one. These interfaces include beeps, voice feedback (i.e., spoken messages), and voice control. Beeps are suitable for drawing drivers' attention. However, it provides unidentified information unless the driver recognizes the source of the beeper. Visual interfaces when used solely are usually used to communicate information in non-critical events. This is because visual messages could fail to deliver important information if the information displayed goes unnoticed by drivers. Over years, numerous visual interfaces were included in vehicles to suit different applications of autonomous and connected vehicles, including indicator lights, LCD displays, organic light emitting diode (OLED) displays, and head-up displays (HUD). However, the most detrimental effect of using visual interfaces is the possible increase in visual workload (Engström et al., 2015). The research also suggests that visual warnings could be used as supplemental information to an auditory or haptic warning. Haptic interfaces provide the driver with information through the driver's tactile sense, such as a lane-keeping warning system that develops reaction torque when departing from the lane (Montiglio et al., 2006), and the haptic steering interface (Steele and Gillespie, 2001; Boyle, 2012), which can give navigation by developing sequenced pulses on the wheel clockwise or counterclockwise according to the required direction.

For the design format of messages that are displayed on the HMI, the Federal Highway Administration (FHWA) emphasizes that they should adhere to standard message formats. It is highly recommended to use familiar signs and messages that are similar to what is provided in the MUTCD (FHWA, 2015). This is because drivers may get confused with regard to the meaning of non-standard signs. In addition, spatial compatibility is required for the design of the message in the context of communicating information to drivers because the selection of a response is directly related to the position of the related stimulus. Information provided on HMIs should match what is provided on real-world traffic control devices (FHWA, 2015). Péter et al. (2017) point out that HMI devices are initially developed to provide services that enhance the efficiency of driving tasks. General aspects and standards for effective HMIs include the following requirements: readability, clarity, interpretability, accessibility, and ease of handling. Sentouh et al. (2014) indicate that the implementation of HMI should address a number of challenges, including what information is important for drivers, how information is displayed, when, under what circumstances, and in what order the information should be presented to drivers. Olaverri-Monreal and Jizba (2016) summarize the issues involved in the field of human–machine interaction; it was concluded that the in-vehicle HMI should provide an intuitively meaningful indication of the presence of a warning and its timely status. In addition, it is crucial to investigate driver distraction levels as well as the modality and dimension of the visual warnings and their suitable in-vehicle locations. Biondi et al. (2017) investigated the effectiveness of auditory, vibrotactile, and multimodal (i.e., combination of two or more modalities) HMI warnings; it was found that multimodal warnings appeared to be effective in low-workload conditions. However, the effect vanishes as the overall level of workload increases.

Assessment of HMI

The most commonly used HMI design-assessment methodologies found in the literature are based on (1) stated-preference questionnaire surveys, (2) field-experiment testing using instrumented vehicles, and (3) driving-simulator testing.

For the questionnaire-survey method, Höltl and Trommer (2013) compared European drivers' perceptions of advanced driver assistance systems (ADAS) through an online questionnaire survey, which aimed at collecting each driver's rating of different ADAS applications in terms of perceived usefulness, ease of use, efficiency, and changed driving behavior. Bazilinskyy and de Winter (2015) conducted an international survey to gather drivers' opinions and preferences on auditory interfaces. The results show that the auditory interfaces are preferred for the application of parking assistance and a forward collision warning (FCW) system. Another worldwide connected vehicle survey conducted by Accenture Consulting (2016) shows that traffic information, weather information, and a speed camera are the most popular HMI applications. For the field-experiment method, Fitch et al. (2014) investigated whether collision avoidance systems should present individual crash alerts in a multiple-conflict scenario or present only one alert in response to the first conflict. This was because, in reality, secondary alerts may startle, confuse, or interfere with drivers' execution of an emergency maneuver. Testing results show that drivers who received both the FCW and lane-change merge alerts were significantly faster at steering away from the lateral crash threat than the drivers who received only the FCW alert. Song et al. (2016) evaluated drivers' response to HMI under two different types of warning systems, emergency warning and general warning, by combining various modalities. Study results show that, for emergency alerts, the most effective warning information was transmitted by integrating "voice, graphic, and text" or "repeated computer tone and text." In the case of a general warning alert, the "repeated computer tone, voice, graphic, and text" combination was indicated to be the most effective.

Bao et al. (2012) evaluated truck drivers' following behavior to an in-vehicle crash warning system in a naturalistic driving environment. Results indicate that the presence of warnings increased mean time-headway by 0.28 s, and drivers' response time to the forward collisions was 15% faster than the baseline condition (i.e., no in-vehicle crash warning system). Biondi et al. (2018) developed a rating tool for assessing HMIs of various ADASs. Based on a field-experiment testing, the authors point out issues that are related to visual, auditory, and haptic warnings; for example, auditory warnings used by the rear parking sensor were not indicative of the distance of the vehicle to obstacles, visual warnings adopted by a blindspot monitor were located in unconventional locations, and accelerations operated by the lane keep assist system were in some cases uncomfortable and jolty.

In comparison with the questionnaire survey and field-experiment methods, a driving simulator has the advantages of testing different HMI design alternatives in a safe environment, and environmental variables can be better controlled. Cummings et al. (2007) investigated the impacts of single versus multiple warnings on driver performance. It was found that participants' reaction times and accuracy rates were significantly affected by the type of collision event and alarm reliability. Moreover, the use of individual warnings did not significantly affect driving performance in terms of reaction time or response accuracy. Osman et al. (2015) tested the location of the visual HMI display in a connected-vehicle simulator experiment. Results reveal that the majority of respondents preferred the visual display to be provided as a HUD in the midsection of the windshield. Jakus et al. (2015) investigated the effectiveness of integrating multimodal interfaces and using single-modal interfaces. Three different interfaces were defined: (1) visual, (2) auditory, and (3) a multimodal auditory and visual interface. Results show that the interaction with visual and audio head-up displays was significantly faster and safer. In term of efficiency, no significant difference was found between the visual only and audiovisual modalities. However, the majority of the users preferred to use multimodal interfaces. Zhao et al. (2016) developed an integrated driving simulator and microsimulation modeling framework to evaluate the environmental benefits of CV applications. The authors point out that driving simulator–based experiments have the advantage of accounting for the response of human drivers to the recommended speed profiles, thus safely and more accurately evaluating the benefits of CV applications. Ma et al. (2016) employed a driving simulator to compare the effectiveness of physical roadside dynamic message signs (DMS) and virtual DMS (VDMS) generated by CV technology. Effectiveness was measured in terms of message comprehension, distraction, and overall difficulty level in receiving messages. It was concluded that, in general, VDMS performed better than DMS, particularly with the increase of the message length and under higher driving workload conditions. Schwarz and Fastenmeier (2017) investigated the effects of modality (e.g., auditory vs. visual) and specificity (e.g., low vs. high volumes) on warning effectiveness. Results show that the effects of specificity is dependent on the modality of the warning. Francois et al. (2017) compared three speedometer display patterns in a simulated truck-driving setting: digital, analog, and redundant speedometers. It was found that the digital speedometer is more efficient and less visually distracting for absolute and relative reading tasks, whereas the analog speedometer is more effective for detecting a dynamic speed change. The redundant speedometer has the best performance when compared to the two single types for each of the three reading tasks. Naujoks et al. (2017) explored the potential of using visual-auditory HMI to inform drivers in a non-distracting way. Based on the driving-simulator testing, it was found that participants clearly favored the HMI with additional speech-based output, which demonstrates the potential of speech to enhance the usefulness and acceptance of HMI. Houtenbos et al. (2017) examined the effects of audiovisual warning of the speed and direction of intersecting vehicles at intersections using a driving simulator. Based on a postexperiment questionnaire survey, the authors conclude that the beeps (audio modality) were regarded as more useful than the lights (visual modality). Vaezipour et al. (2018) designed a driving simulator experiment to investigate drivers' acceptance of various types of in-vehicle HMIs (i.e., visual advice only, visual feedback only, and visual advice plus feedback) and the impact of in-vehicle HMI on driving behavior. Results show that visual advice only HMI was most accepted by participants, and both advice and feedback HMIs were found to benefit eco-safe driving behavior.

With consideration of the costs of each assessment methodology and the availability of facilities at the time the research was conducted, this research employed integrated driving-simulator testing and revealed-preference questionnaire survey methods to identify drivers' perceptions of Wyoming CV HMI.

DESCRIPTION OF WYOMING CV APPLICATIONS AND HMI

CV Applications

The Wyoming CV applications were classified into five categories based on their function and communication technologies (Gopalakrishna et al., 2016): category 1: forward collision warning (FCW), category 2: distress notification (DN), category 3: situational awareness (SA), category 4: work zone warnings (WZW), and category 5: spot weather impact warning (SWIW). A detailed illustration of the existing communication and traffic control devices along the Wyoming I-80 corridor and the DSRC locations that are deployed on the corridor can be found in Ahmed et al. (2019a).

Forward Collision Warning (FCW)

Forward collision warning is a V2V communication-based safety application that issues a warning to the CV driver in case of an impending front-to-rear collision with another CV ahead in traffic in the same lane and direction of travel. FCW aims to help CV drivers avoid or mitigate front-to-rear vehicle collisions in the forward path of travel. This CV application is critically important for safety along I-80 in conditions when snowplows are moving slower than following traffic and/or low visibility conditions caused by adverse weather. The developed FCW has two warning levels: the cautionary level and the alert level. The HMI displays a yellow cautionary warning icon along with a loud beep sound when the time-to-collision is greater than 5 s but less than 9 s; drivers need to be prepared to brake when receiving the cautionary FCW. When the collision time is less than 5 s, the alert FCW is triggered; the HMI displays a red warning icon along with continuous loud beeps. Drivers need to immediately begin braking to avoid rear-ending the leading vehicle.

Distress Notification (DN)

Distress notification is a V2I communication-based safety application that enables CVs to communicate a distress status back to the Wyoming CV system when the vehicle's sensors detect an event that might require assistance from others (e.g., air bag deployed and vehicle disabled) or the CV driver manually initiates a distress notification. The DN, which includes the vehicle category, location, content, and time of the message, is sent to the nearest roadside unit (RSU). The RSU forward it to the Wyoming CV system for notifying system operators. If an RSU is out of the range of DSRC, the DN is expected to be received by nearby CVs that are traveling in the same and/or opposite direction via V2V communication. These CVs will forward the DN to an RSU that is connected to the Wyoming CV system.

Situational Awareness (SA)

The SA application adopts I2V and V2I communication technologies to assemble important travel information from Wyoming CV system operators and communicate them directly to CV drivers through both DSRC and satellite communications. SA enables delivery of up-to-date downstream traffic and road conditions that may affect driving safety to CV drivers. The SA application includes weather alerts, speed limitations, vehicle restrictions, road surface conditions, incidents ahead advisories, truck parking availability, and road closures, etc.

Work Zone Warnings (WZW)

The WZW application employs I2V communication technology to provide CV drivers information about the unsafe conditions that exist in an active work zone, such as obstructions in the vehicle's travel lane, lane shifts and closures, speed reductions, and construction vehicles/workers entering or exiting the work zone.

Spot Weather Impact Warning

SWIW is a special case of SA that warns CV drivers of local hazardous weather conditions, such as rain, snow, fog, or strong winds. The primary difference between SWIW and other SA applications is that it provides more localized information.

The majority of the visual warnings were developed following Manual on Uniform Traffic Control Devices (MUTCD) guidance (FHWA, 2009); detailed descriptions of each CV application, including its communication technology or technologies, visuals that are displayed on the CV HMI, and messages delivered by this CV application, are summarized in **Table 1**.

Layout of Wyoming CV HMI

Figure 1 illustrates how the Wyoming CV warnings are displayed on the HMI screen. In general, these CV warnings are categorized into four priority levels based on the urgency of the imminent situation: Level 1–FCW, Level 2–variable and regular speed limit, Level 3–critical warnings, and Level 4–advisory warnings, respectively. In this pilot study, critical warnings were determined to be situations that would significantly affect driver's operation of vehicle (e.g., icy road surface, work zone, severe weather, etc.) or appear beyond expectation (e.g., road closure, accident or distressed vehicle ahead, fog or strong wind ahead, etc.). Advisory warnings aimed to provide advisory information to draw drivers' awareness while driving, mainly including adverse weather conditions that may affect driving, such as rain and snow, location of rest area or parking area, etc. Within the critical and advisory warnings, in case there are multiple warnings appearing simultaneously on the HMI, warnings that are more urgent are displayed closer to the driver, i.e., on the left side of the HMI.

ASSESSMENT OF WYOMING CV HMI

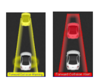
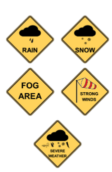
Apparatus

The CV driving simulator study was conducted at the University of Wyoming Driving Simulator Lab (WyoSafeSim). The motion-based, high-fidelity driving simulator can switch between a passenger car (2004 Ford Fusion) cockpit cab and a freight truck (2000 Sterling AT9500 18-wheeler semi-trailer) cockpit cab. It is mounted on a three-degrees-of-freedom D-Box motion platform, which comprises four electro-mechanical linear actuators to provide two rotational and one translational degrees of freedom (roll, pitch, and heave). The simulator provides motion cues to immerse the driver into a real driving experience with changes in kinematics, such as velocity, acceleration, and deceleration. In addition, a low-frequency vibration transducer is mounted on the vehicle floor to simulate vibrations generated by engine and road. The simulator has open architecture software with complete source code of simulation creator tool, which offers flexibility of building roadways and developing driving scenarios that could replicate the actual driving environments. The CV HMI was mounted on the dashboard of the simulator to provide participants with the various CV warnings as illustrated in **Figure 2**.

Participants

This research recruited a total of 26 professional drivers to participate in the CV driving-simulator study to assess the effectiveness of the Wyoming CV HMI. The participants

TABLE 1 | Summary of the Wyoming CV applications.

CV category	Technology	Visual(s)	Messages delivered
Forward collision warning (FCW)	V2V		An impending front-end collision with a CV ahead in the same traffic lane and direction of travel.
Distress notification (DN)	V2I and V2V		A distress notification is sent to other CVs as well as local traffic management center to seek emergency help.
Situational awareness (SA)	I2V and V2I		Road surface conditions: an icy or slick spot road will be encountered while driving. Variable speed limits: advisory/regulatory operating speed limits for existing road and/or weather conditions. Road closures and restrictions: road closed to all vehicle types or certain types of vehicles such as light trailers or light high profile vehicles. Parking availability: information provided for available close by parking or rest areas.
Work zones warning (WZW)	I2V		An active work zone ahead as well as distance to the work zone, lane closure, and speed limit.
Spot weather impact warning (SWIW)	I2V and V2I		An adverse weather condition, such as rain, snow, fog, strong wind, or severe weather, ahead.

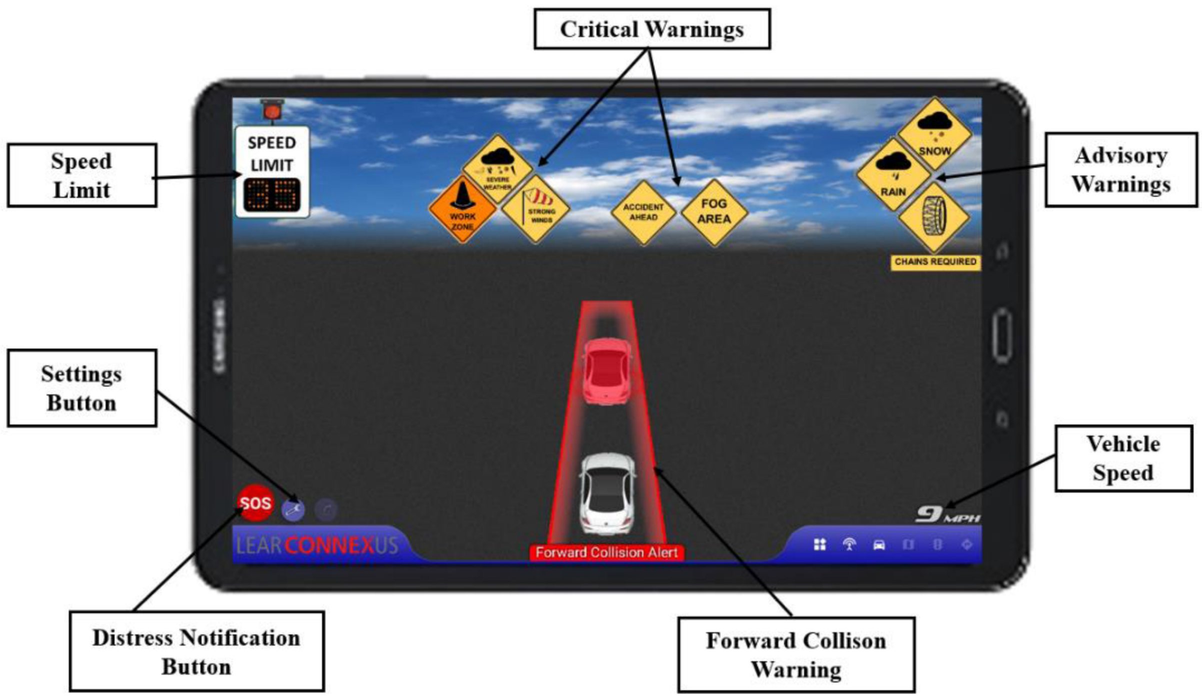


FIGURE 1 | Layout of the Wyoming CV HMI (Ahmed et al., 2019b).

are professional snowplow truck drivers and employees from WYDOT; they are expected driving connected trucks and connected maintenance vehicles after the full deployment of the CV system in Wyoming. The selection of participants considered a wide range of factors that might affect the acceptance or perception of CV technology, such as age, education level, driving experience, etc. Because potential Wyoming CV users are commercial truck drivers, WYDOT snowplow truck and freeway maintenance vehicle drivers, and Highway Patrol vehicle drivers, at this stage, all the participants were male. Based on a predrive survey questionnaire, it was summarized that the participants' ages ranged from 21 to 61 years (Mean = 42; *S.D.* = 10.3). Among the 26 participants, 15 graduated from high school, nine have a college degree, and two have a postgraduate degree. All participants had a valid commercial or class C driver's license with an average driving experience of 14.5 years (range: 0.5–36 years, *S.D.* = 11). Twenty-five of the participants reported they never had any ophthalmic surgery (one participant had laser vision correction in 2006). During the driving simulator study, all the participants were in good health condition without vision, audition, and emotional issues that might affect their normal driving (e.g., angry, depressed, dizzy, etc.). All the participants reported that they have encountered reduction in visibility due to snow, blizzards, fog, smoke, or heavy rain while driving on I-80 in Wyoming.

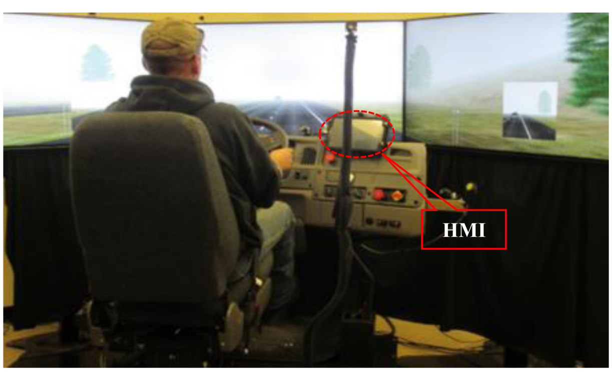


FIGURE 2 | Truck simulator with CV HMI.

Driving-Simulator Study Scenarios

Three comprehensive simulation scenarios were developed to simulate different real-world traffic and weather conditions on I-80-like freeways: work zone with FCW in fog, slippery road surface due to snowy weather, and road closure due to accident in severe weather, respectively. After a warm-up session, each participant drove each simulation scenario two times; one with the CV HMI turned on and the other one with the CV HMI turned off. To eliminate the potential impact of any learning effect on the simulation result, this research randomly assigned the sequence of these six simulation scenarios to each participant. Prior to the driving simulator study, participants were provided with training on both the basic concept of the Wyoming CV system and hands-on operation of the driving simulator under the CV environment. **Figure 3** illustrates the driving simulator study at the WyoSafeSim lab.

The driving simulator test bed was designed as a two-way, four-lane freeway segment with a 75-mph (120 km/h) speed limit to represent the basic operational conditions of I-80 in Wyoming. To control for the potential impact of the ambient traffic on participants' driving behavior, the average and standard deviation of speed of the ambient traffic was coded to match speed distributions similar to the Wyoming I-80 in alike adverse weather conditions.

The work zone simulation scenario aimed to test the Wyoming CV system's WZW and FCW applications. These CV applications are expected to help in avoiding potential collisions at a freeway work zone due to reduced visibility caused by fog. The general simulation procedure is detailed as follows:

Participants first accelerated to the normal freeway driving speed (i.e., 75 mph). A fog area was design ahead of a work zone; a "fog area" CV warning with an advisory speed limit of 65 mph (105 km/h) were displayed on the CV HMI before participants entered the fog area. In the work zone, the right lane of the freeway was closed following typical construction zone layouts in Wyoming; a series of WZWs along with an advisory speed limit of 45 mph (75 km/h) were shown on the CV HMI to alert participants to change lanes and reduce speed before entering the work zone. To simulate an FCW, a slow-moving truck was designed to appear in the work zone; a worker suddenly crossed the lane in front of the slow-moving truck, and thus, the truck made an emergency brake to yield to the worker. A proximity sensor was employed to trigger the truck, indicating that the truck could make the braking action at the designated distance in front of the simulator vehicle. Then, with V2V communication technology, an advisory and an alert FCW were displayed successively on the CV HMI to notify participants of the potential forward collision. It is worth mentioning that the foggy condition was created to allow a safe stopping sight distance for 45 mph for the simulator vehicle type, i.e., heavy truck.

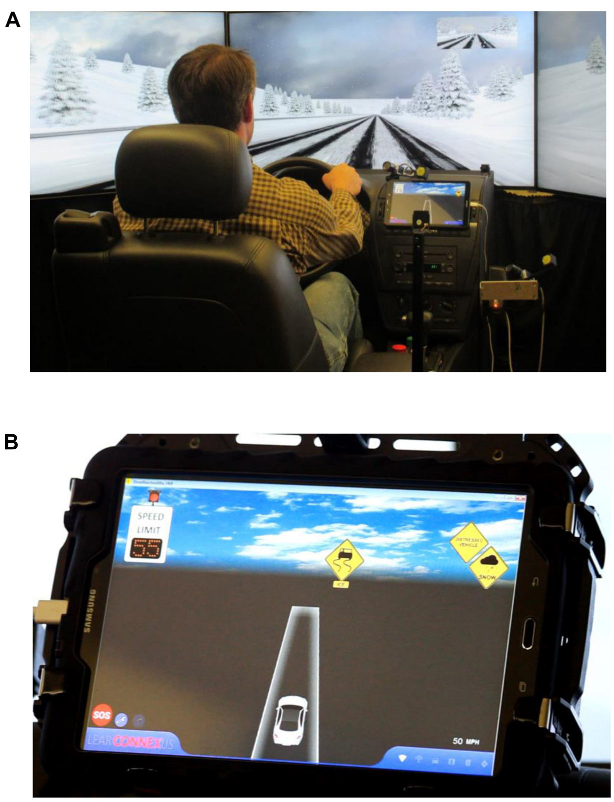


FIGURE 3 | Illustration of driving simulator study at WyoSafeSim lab: **(A)** A simulation scenario; **(B)** CV HMI.

The slippery road surface simulation scenario was designed to test the Wyoming CV system's SWIW and DN applications. Functions of these CV applications were to warn the participants to reduce speed before entering an icy road segment, thus avoiding skidding off the travel lane or being involved in a secondary crash.

This simulation scenario started with a snowy weather condition; a "snow" CV warning with an advisory speed limit of 65 mph appeared on the CV HMI. Later on, a "severe weather" CV warning with an advisory speed of 45 mph were displayed on the CV HMI. Before entering the icy road segment, an "icy surface" CV warning with a 35 mph (55 km/h) advisory speed limit were displayed on the CV HMI to warn participants to reduce speed when driving on the icy road. Prior to entering the icy curve, a "distressed vehicle" warning was received to alert participants there was a skidding-off accident ahead, indicating that participants should drive with extreme caution. For participants who lost control of the vehicle due to speeding, they were asked to use the DN application to generate and send a distress message to the TMC and other CVs on the road (after sending the DN, the simulation scenario was automatically terminated).

The road closure simulation scenario intended to test the Wyoming CV system's SWIW and SA applications. These CV applications provided participants with real-time road closure notification due to an incident as well as information about the nearest rest area to help participants avoid being jammed on the closed freeway or involve in a secondary crash. A "snow" CV warning with a speed limit of 65 mph was displayed on the CV HMI. Later on, a "severe weather" CV warning with an advisory speed of 45 mph were displayed. A pile-up crash was designed on the freeway mainline to simulate a road closure condition; the crash was located downstream of a rest area. "Accident ahead" and "road closed" warnings were displayed on the CV HMI; then, a "rest area" notification appeared on the CV HMI to inform participants about the nearest rest area. If a participant exited the freeway to the rest area, a voice message was played to inform the participant to park the vehicle and stop this driving simulator scenario. Otherwise, the participant was queued in front of the crash location on the freeway.

Sequence of the CV warnings and general layout of each driving-simulator study scenario are illustrated in **Figure 4**.

Questionnaire Survey

After experiencing the Wyoming CV application in the driving-simulator study, a comprehensive postdrive questionnaire survey was employed to collect participants' qualitative opinions regarding their preferences on different CV warning modalities

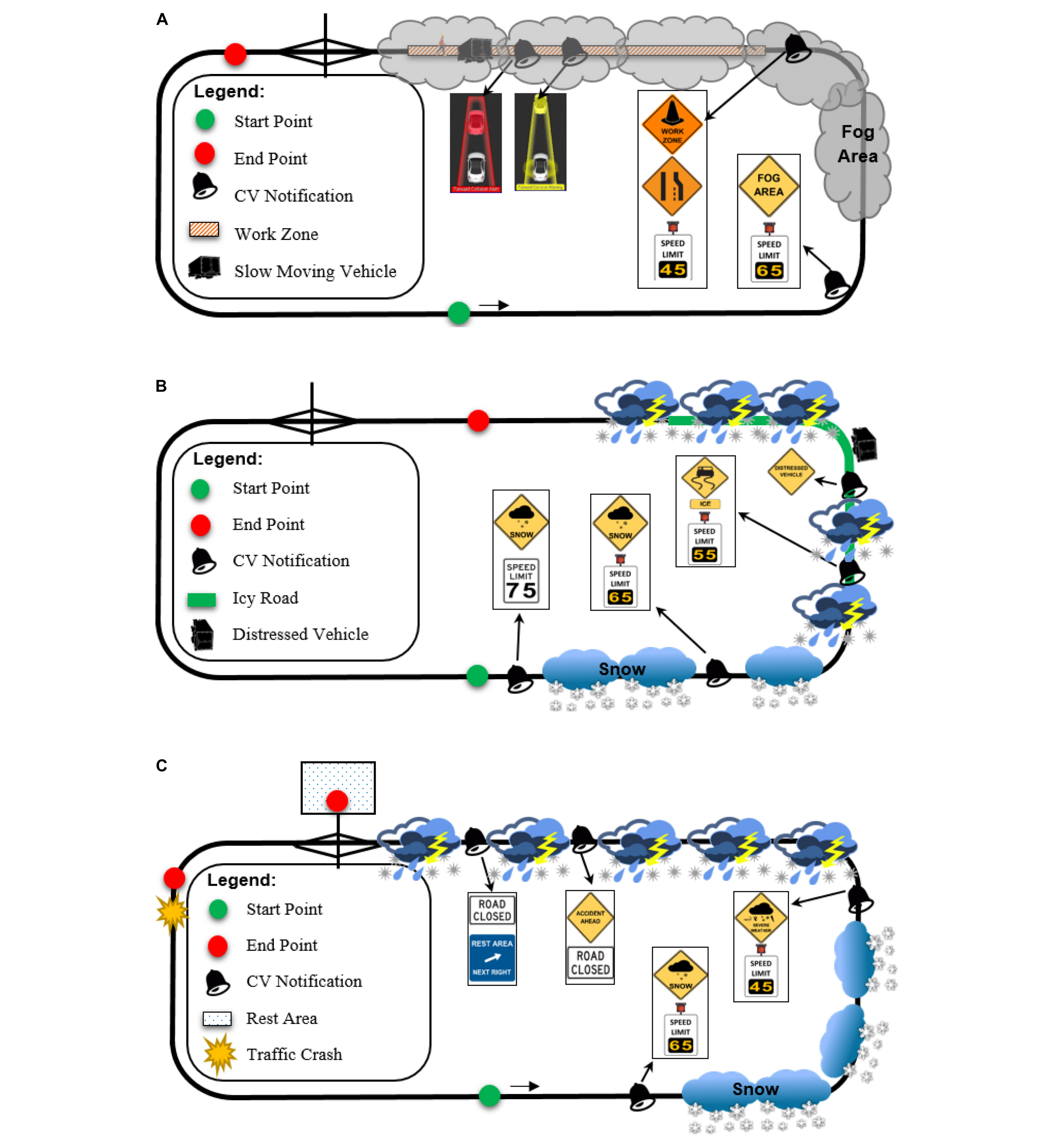


FIGURE 4 | Description of driving simulator study scenarios: **(A)** work zone with FCW in fog; **(B)** slippery road surface due to snowy weather; **(C)** road closure due to accident in severe weather.

and the effectiveness of CV technology under various real-world driving conditions. The questionnaire survey was initially designed by the University of Wyoming research team and then reviewed, revised, and approved by the USDOT Volpe National Transportation Systems Center.

Results show that the majority of participants (96.2%) preferred to have the CV warnings displayed at the combination of visual and auditory modalities. For the auditory-warning modality, it was found that using a simple "beep" sound for advisory warnings and a series of louder "beep" sounds for critical warnings would best draw a driver's attention while a repeated voice message tended to disturb normal driving. For the visual-warning modality, results show that, by grouping CV warnings to different priority levels and presenting warnings that have a higher priority closer to the driver (i.e., left side on the CV HMI), drivers tended to more easily perceive the imminent safety hazard when multiple warnings were displayed on the CV HMI. Overall, participants indicated that CV technology was most useful under poor-visibility driving conditions, such as rainy, foggy, blowing snow, and sun glare weather conditions, as illustrated in **Figure 5**. It was found that, under normal daytime driving conditions, participants felt that CV technology did not introduce significant benefits in comparison with when driving under adverse weather conditions that resulted in a slippery road surface, when the view of the road ahead was partially blocked by other vehicles on the curvy terrain, or driving at night.

In addition to qualitative descriptions, the questionnaire survey also collected participants' assessment of the readability and usefulness of the Wyoming CV applications. Readability refers to how easily participants felt they recognized a CV warning or a bundle of CV warnings; usefulness means whether a CV warning helped drivers to recognize an imminent safety hazard or assisted them in better planning their trip. The assessment contains two components: assessment of CV technology and the specific CV applications, respectively. Responses were measured on a 7-point Likert scale (example: strongly disagree to strongly agree). Accordingly, the evaluation results were converted to a 1–7 score, in which score 1 corresponds to a very negative assessment result and score 7 to a very positive assessment result. The numerical values were used for quantification of assessments and comparisons across different CV applications.

Table 2 presents 26 participants' assessment results of the Wyoming CV system. In addition to the scores generated from the Likert scale questionnaire survey, this paper categorizes

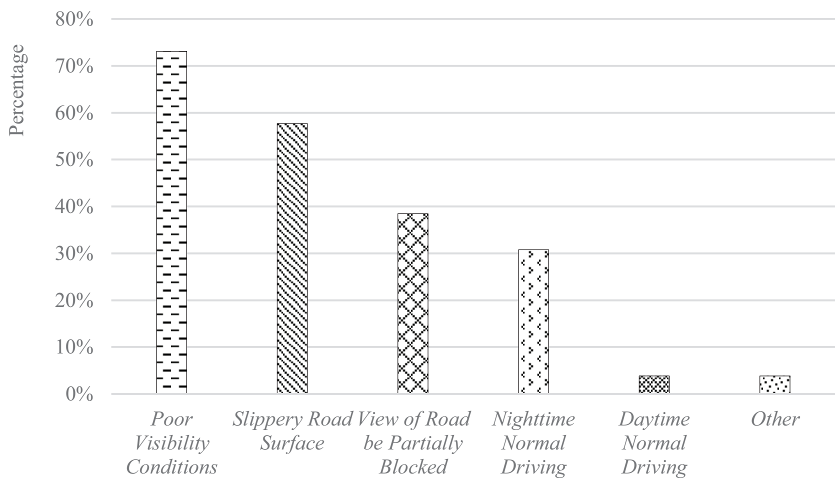


FIGURE 5 | Rank of the effectiveness of CV technology under various driving conditions. Poor visibility conditions = rain, fog, snow, sun glare weather conditions; view of road be partially blocked = view of the road ahead is partially blocked by other vehicles or the curves and other terrains; other = road work.

TABLE 2 | Participants' assessment results of CV technology.

Scale items	Mean	*SE*	Positive	Neutral	Negative
(a) Readability of CV warnings					
A1: After experiencing the CV applications, how easy was it to understand the CV technology and warnings?	6.1	0.80	96.2%	3.8%	0%
A2: Do you think that warnings among the different CV applications are confusing?	5.5	0.95	80.8%**	19.2%	0%
A3: Do you think that the CV warnings and the display unit are introducing any distraction from the main driving task?	5.2	1.37	73.1%***	15.4%	11.5%
A4: Were the visual warnings clear, obvious, and convey the required message?	5.7	0.93	84.6%	15.4%	0%
(b) Usefulness of CV Technology					
B1: Do you think the CV system provided you with improved road condition information?	5.8*	1.14	85.7%	9.5%	4.8%
B2: Do you think that having the CV applications would help to increase traffic safety and reduce crashes?	5.9	1.14	88.5%	7.7%	3.8%
B3: How likely will you be dependent on the CV applications to warn you for upcoming hazardous conditions, when fully implemented on I-80?	4.2	1.61	42.3%	30.8%	26.9%
B4: Would you like to have the CV applications in your vehicle?	4.8	1.67	65.4%	23.1%	11.5%

**Based on 21 available samples; **positive feedback means CV applications were NOT EXPERIENCED AS confusing; ***positive feedback means CV applications and the display units were not experienced as introducing any distraction.*

TABLE 3 | Participants' assessment results of the specific CV applications.

CV applications	Readability					Usefulness				
	Mean	*SE*	Positive	Neutral	Negative	Mean	*SE*	Positive	Neutral	Negative
FCW	5.9	0.99	84.6%	15.4%	0%	6.1	1.03	88.5%	11.5%	0%
DN	6.0	1.11	88.5%	7.7%	3.8%	5.7	1.12	84.6%	11.5%	3.8%
SA (road surface)	6.1	0.80	96.2%	3.8%	0%	5.7	1.08	84.6%	11.5%	3.8%
SA (rerouting)	6.1	1.13	92.3%	3.8%	3.8%	6.0	1.10	84.6%	15.4%	0%
WZW	6.2	0.97	88.5%	11.5%	0%	5.8	1.24	80.8%	15.4%	3.8%
SWIW	5.9	0.91	92.3%	7.7%	0%	5.5	1.27	73.1%	19.2%	7.7%

participants' perceptions of CV technology into three categories: positive (scores 5–7), neutral (score 4), and negative (scores 1–3). In general, the majority of participants provided positive feedback regarding the Wyoming CV applications and indicated that CV technology provided improved road condition information and would help to increase traffic safety.

Table 3 presents participants' assessment results of the readability and usefulness of the specific CV applications. Overall, the readability and usefulness of the Wyoming CV applications have been well accepted by the participants; specifically, FCW and rerouting notifications were found to be most useful.

CONCLUDING REMARKS

This study assessed the subjective experiences related to the readability and usefulness of the Wyoming CV application in a simulated environment. It was found that the majority of the participants preferred to have the CV warnings provided in a combination of visual and auditory modalities. For visual warnings, this study grouped the CV warnings into four priority levels and presented warnings that have a higher priority closer to the driver. This was considered by the participants to be an effective way for them to perceive the imminent safety hazard when multiple warnings were displayed on the HMI simultaneously. For auditory warnings, the participants reported that a simple "beep" sound for advisory warnings and a series of louder "beep" sounds for critical warnings would best draw their attention while a repeated voice message tended to disturb normal driving. The participants indicated that CV technology was most useful under poor-visibility driving conditions; FCW and rerouting were the most useful CV applications. It is worth pointing out that FCW and rerouting CV applications have the most significant potential to realize the WYDOT CV pilot's strategic goals to improve safety and mobility. Generally speaking, FCW and rerouting applications are tactical-level CV applications, which can directly help drivers to avoid a crash or being congested on the freeway. In comparison, DN, SWIW, WZW, and other SA applications are strategic-level CV applications, which aim to assist drivers more easily to recognize safety hazards or unexpected events, particularly when drivers' recognition ability is limited by visibility.

Nevertheless, assessment results reveal that there are still a couple of issues that need to be considered to further improve the design of the Wyoming CV HMI. The primary issue is the potential distraction of CV HMI. As presented in **Table 2**, approximately 27% of participants indicated that distraction could be introduced by the Wyoming CV HMI (i.e., 12% found the CV warning distracting, and 15% found them neutral). Another issue is that the usefulness of CV technology tends to be less significant during normal daytime driving conditions or when drivers can recognize hazardous conditions without receiving CV warnings. From **Table 2**, it was found that only 42% of participants stated that they are going to depend on the CV applications to identify upcoming hazards, and less than two thirds of the participants showed desirability of having CV technology in their vehicles. These findings are consistent with a previous study that found drivers may not exactly trust in advanced driver assistance systems (Kidd et al., 2017) and also further proved previous research findings that truck drivers would like to receive acceptable feedback that is designed and implemented properly (Roetting et al., 2003) and displeasure with the continuous auditory warnings (Bazilinskyy et al., 2019). Therefore, these findings indicate that, under normal daytime driving conditions, the repeated auditory or visual CV warnings might distract drivers from their driving task. With this consideration, this study suggests that the design of CV HMI needs to add a user customization capability to suit the needs of individual users, such as a CV system that can be automatically or manually deactivated under normal daytime driving conditions. Nevertheless, it is necessary to clarify that this study is highly practice-focused, which aimed at supporting the WYDOT CV pilot. At this stage, this study only recruited 18 male drivers from WYDOT and the trucking industry; findings of this pilot study presented some preliminary insights into the optimal design of CV HMI display in a way that drivers can perceive CV warnings promptly without being distracted. Considering the increasingly popularity of CV technology, future studies need to recruit a larger number of participants that cover a wider range of demographic features to further investigate general drivers' perceptions of the Wyoming CV HMI through statistical analysis and modeling, which will further benefit the design of CV HMI for general purposes (Engström et al., 2015; Biondi et al., 2017; Vaezipour et al., 2018, 2019).

In fact, safe driving is the principle task for human drivers. As specified by the National Highway Traffic Safety Administration (Nhtsa, 2009), the primary requirement of the in-vehicle HMI is to deliver desired warnings or notifications to a driver while minimizing driver distraction. Therefore, the optimal design of CV HMI needs to balance a trade-off between the readability of the messages (e.g., maximum number of messages displayed on the CV HMI, length of each message and the modality of the message, etc.) and drivers' capability to safely recognize and timely respond to the received message(s). This is particularly critical during high-workload situations or under adverse weather conditions when drivers need more response and reaction time to an unexpected event because overloaded CV HMI information may distract the driver and lead to safety issues. However, there is still a lack of a comprehensive assessment of the effectiveness of different CV HMI display designs and development of CV HMI design guidelines considering human factors. Specifically, the following aspects need to be further investigated: (1) Which kind of HMI display modality (i.e., visual, auditory, voice message, or a combination of visual and auditory) best delivers the meaning of a warning? (2) What is the maximum number of warnings that can be displayed on the HMI without confusing drivers? (3) When should an early warning be displayed and how long should the warning remain on the HMI? (4) How to prioritize different warnings when they are displayed simultaneously on the HMI. In summary, incorporating human factors into the design and development of CV HMI has become increasingly critical, which aims to minimize the potential distractions introduced by these in-vehicle technologies.

DATA AVAILABILITY STATEMENT

The data that support the findings will be available upon reasonable request from the corresponding author, MA.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Wyoming IRB. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MA, GY, and SG: study conception and design. SG and GY: experiments and data collection. GY: analysis and interpretation of results. GY and MA: drafted manuscript preparation. All authors reviewed the results and approved the final version of the manuscript.

FUNDING

This research was sponsored by the United States Department of Transportation (USDOT) (Grant No. DTFH6116RA00007) and the Wyoming Department of Transportation (WYDOT) (Grant No. RS04218).

ACKNOWLEDGMENTS

The authors thank the Volpe National Transportation Systems Center for reviewing and revising the questionnaire survey, and the participants from the WYDOT and Trihydro Corporation for participating in the driving simulator study and their feedback on the Wyoming CV HMI.

REFERENCES

Accenture Consulting (2016). *Accenture Connected Vehicle Survey: What Drivers Want*. Available online at: https://www.accenture.com/us-en/insight-automotive-connected-vehicle (accessed July, 2020).

Ahmed, M. M., Yang, G., Gaweesh, S., Young, R., and Kitchener, F. (2019a). Performance evaluation framework of Wyoming connected vehicle pilot deployment program: summary of Phase 2 pre-deployment efforts and lessons learned. *J. Intell. Connect. Vehicles* 2, 41–54. doi: 10.1108/jicv-03-2019-0006

Ahmed, M. M., Yang, G., and Gaweesh, S. (2019b). Development and assessment of a connected vehicle training program for truck drivers. *Transp. Res. Rec.* 2673, 113–126. doi: 10.1177/0361198119827904

Bao, S., LeBlanc, D. J., Sayer, J. R., and Flannagan, C. (2012). Heavy-truck drivers' following behavior with intervention of an integrated, in-vehicle crash warning system: a field evaluation. *Hum. Fact.* 54, 687–697. doi: 10.1177/0018720812439412

Bazilinskyy, P., and de Winter, J. (2015). Auditory interfaces in automated driving: an international survey. *Peer J Comput. Sci.* 1:e13. doi: 10.7717/peerj-cs.13

Bazilinskyy, P., Larsson, P., Johansson, E., and de Winter, J. C. F. (2019). Continuous auditory feedback on the status of adaptive cruise control, lane deviation, and time headway: an acceptable support for truck drivers? *Acoust. Sci. Technol.* 40, 382–390. doi: 10.1250/ast.40.382

Biondi, F., Strayer, D. L., Rossi, R., Gastaldi, M., and Mulatti, C. (2017). Advanced driver assistance systems: using multimodal redundant warnings to enhance road safety. *Appl Ergon.* 58, 238–244. doi: 10.1016/j.apergo.2016.06.016

Biondi, F. N., Getty, D., McCarty, M. M., Goethe, R. M., Cooper, J. M., and Strayer, D. L. (2018). The challenge of advanced driver assistance systems assessments: a scale for the assessment of the human-machine interface of advanced driver assistance technology. *Trans. Res. Rec.* 2672, 113–122. doi: 10.1177/0361198118773569

Boyle, R. (2012). No more GPS voice: haptic steering wheel buzzes to give you directions. *Popular Sci.*

Cummings, M. L., Kilgore, R. M., Wang, E., Tijerina, L., and Kochhar, D. S. (2007). Effects of single versus multiple warnings on driver performance. *Hum. Fact.* 49, 1097–1106. doi: 10.1518/001872007x249956

Dey, K. C., Rayamaihi, A., Chowdhury, M., Bhavsar, P., and Martin, J. (2016). Vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication in a heterogeneous wireless network – Performance evaluation. *Trans. Res. Part C* 68, 168–184. doi: 10.1016/j.trc.2016.03.008

Engström, J., Johansson, E., and Östlund, J. (2015). Effects of visual and cognitive load in real and simulated motorway driving. *Trans. Res. Part F* 8, 97–120. doi: 10.1016/j.trf.2005.04.012

FHWA, (2009). *Manual on Uniform Traffic Control Devices for Streets and Highways*. Washington, D.C: U.S. Department of Transportation – FHWA.

FHWA, (2015). *Multiple Sources of Safety Information from V2V and V2I: Redundancy, Decision Making, and Trust-Safety Messages Design Report*. Washington, D.C: U.S. Department of Transportation – FHWA.

Fitch, G. M., Bowman, D. S., and Llaberas, R. E. (2014). Distracted Driver Performance to Multiple Alerts in a Multiple-Conflict Scenario. *Hum. Fact.* 56, 1497–1505. doi: 10.1177/0018720814531785

Francois, M., Crave, P., Osiurak, F., Fort, A., and Navarro, J. (2017). Digital, analogue, or redundant speedometers for truck driving: impact on visual distraction, efficiency and usability. *Appl. Ergon.* 65, 12–22. doi: 10.1016/j.apergo.2017.05.013

Gopalakrishna, D., Garcia, V., Ragan, A., English, T., Zumpf, S., Young, R. K., et al. (2016). *Connected Vehicle Pilot Deployment Program Phase 1, Comprehensive Pilot Deployment Plan, ICF/Wyoming*. Report No. FHWA-JPO-16-297. Washington, D.C: U.S Department of Transportation.

Höltl, A., and Trommer, S. (2013). Driver Assistance Systems for Transport System Efficiency: influencing Factors on User Acceptance. *J. Intell. Trans. Syst.* 17, 245–254. doi: 10.1080/15472450.2012.716646

Houtenbos, M., de Winter, J. C. F., Hale, A. R., Wieringa, P. A., and Hagenzieker, M. P. (2017). Concurrent audio-visual feedback for supporting drivers at intersections: a study using two linked driving simulator. *Appl. Ergon.* 60, 30–42. doi: 10.1016/j.apergo.2016.10.010

Jakus, G., Dicke, C., and Sodnik, J. (2015). A user study of auditory, head-up and multi-modal displays in vehicles. *Appl. Ergon.* 46, 184–192.

Jeong, E., Oh, C., Lee, G., and Cho, H. (2014). Safety Impacts of Inter-vehicle Warning Information Systems for Moving Hazards in Connected Vehicle Environments. *Trans. Res. Rec.* 2424, 11–19. doi: 10.3141/2424-02

Kidd, D. G., Cicchino, J. B., Reagan, I. J., and Kerfoot, L. B. (2017). Driver trust in five driver assistance technologies following real-world use in four production vehicles. *Traffic Inj. Prevent.* 18:su1.

Li, Y., Xing, L., Wang, W., Wang, H., Dong, C., and Liu, S. (2017). Evaluating impacts of different longitudinal driver assistance systems on reducing multi-vehicle rear-end crashes during small-scale inclement weather. *Accident Anal. Prevent.* 107, 63–76. doi: 10.1016/j.aap.2017.07.014

Ma, J., Smith, B. L., and Fontaine, M. D. (2016). Comparison of In-Vehicle Auditory Public Traffic Information With Roadside Dynamic Message Signs. *J. Intell. Trans. Syst.* 20, 244–254. doi: 10.1080/15472450.2015.1062729

Montiglio, M., Martini, S., and Murdocco, V. (2006). "Development of a lane keeping support system for heavy trucks," *Proceedings of the 3th ITS World Congress*, London, U.K.

Naujoks, F., Forster, Y., Wiedemann, K., and Neukum, A. (2017). Improving Usefulness of Automated Driving by Lowering Primary Task Interference through HMI Design. *J. Adv. Trans.* 2017:6105087.

Nhtsa. (2009). *Technology Applications for Traffic Safety Program: A Primer*. Report No. DOT-HS-811-040. Washington, D.C: U. S.Department of Transportation.

Nhtsa. (2015). *Traffic Safety Facts: Crash Status – A Brief Statistical Summary*. Report No. DOT-HS-812-115. Washington, D.C: U.S. Department of Transportation.

Olaverri-Monreal, C., and Jizba, T. (2016). Human factors in the design of human–machine interaction: an overview emphasizing V2X Communication. *IEEE Trans. Intell. Vehicles* 1, 302–313.

Olia, A., Abdelgawad, H., Abdulhai, B., and Razavi, S. N. (2016). Assessing the potential impacts of connected vehicles: mobility, environmental, and safety perspectives. *J. Intell. Trans. Syst.* 20, 229–243. doi: 10.1080/15472450.2015.1062728

Osman, O. A., Codjoe, J., and Ishaq, S. (2015). Impact of time-to-collision information on driving behavior in connected vehicle environments using a driving simulator test bed. *J. Traffic Logist. Eng.* 13, 18–24.

Péter, G., Zsolt, S., and Aradi, S. (2017). *Highly Automated Vehicle Systems*. Report No. 978-963-313-173-2. Budapest: Budapest University of Technology and Economics.

Roetting, M., Huang, Y. H., McDevitt, J. R., and Melton, D. (2003). When technology tells you how you drive: truck drivers' attitudes towards feedback by technology. *Trans. Res. Part F* 6, 275–287.

Schwarz, F., and Fastenmeier, W. (2017). Augmented reality warnings in vehicles: effects of modality and specificity on effectiveness. *Accident Anal. Prevent.* 101, 55–66. doi: 10.1016/j.aap.2017.01.019

Sentouh, C., Popieul, J. C., Debernard, S., and Boverie, S. (2014). Human-machine interaction in automated vehicle: The ABV Project. *IFAC Proc. Vol.* 47, 6344–6349. doi: 10.3182/20140824-6-za-1003.01721

Shladover, S. E. (2018). Connected and automated vehicle systems: introduction and overview. *J. Intell. Trans. Syst.* 22, 190–200. doi: 10.1080/15472450.2017.1336053

Song, T. J., Park, S., and Oh, C. (2016). Field experiment for exploring the effects of in-vehicle warning information on driver's responsive behavior. *J. Eng.* 2016, 124–133. doi: 10.1049/joe.2015.0173

Steele, M., and Gillespie, R. B. (2001). Shared control between human and machine: using a haptic steering wheel to aid in land vehicle guidance. *Proc. Hum. Fact. Ergon. Soc. Ann. Meet.* 45, 1671–1675. doi: 10.1177/154193120104502323

Talamonti, W., Tijerina, L., Blommer, Swaminathan, R., Curry, R., and Ellis, R. D. (2017). Mirage events & driver haptic steering alerts in a motion-base driving simulator: a method for selecting an optimal HMI. *Appl. Ergon.* 65, 90–104.

Vaezipour, A., Rakotonirainy, A., Haworth, N., and Delhomme, P. (2018). A simulator evaluation of in-vehicle human machine interfaces for eco-safe driving. *Trans. Res. Part A* 118, 696–713. doi: 10.1016/j.tra.2018.10.022

Vaezipour, A., Rakotonirainy, A., Haworth, N., and Delhomme, P. (2019). A simulator study of the effect of incentive on adoption and effectiveness of an in-vehicle human machine interface. *Trans. Res. Part F* 60, 383–398.

Wang, C., Yu, L., and Hao, Y. (2016). *Automotive Usability: Human Computer Interaction in the Vehicle*. Available online at: http://homepages.rpi.edu/~{}gricer/symposium/papers/Automotive_paper.pdf (accessed July, 2020).

WYDOT, (2017). *Wyoming DOT Connected Vehicle Pilot: Improving Safety and Travel Reliability ON 1-80 in Wyoming*. Cheyenne, WY: Wyoming Department of Transportation.

Zhao, Y., Wagh, A., Hou, Y., Hulme, K., Qiao, C., and Sadek, A. W. (2016). Integrated Traffic-Driving-Networking Simulator for the Design of Connected Vehicle Applications: Eco-Signal Case Study. *J. Intell. Trans. Syst.* 20, 75–87. doi: 10.1080/15472450.2014.889920

Zulkefli, M., Mukherjee, P., Sun, Z., Zheng, J., Liu, H. X., and Huang, P. (2017). Hardware-in-the-loop testbed for evaluating connected vehicle applications. *Trans. Res. Part C* 78, 50–62. doi: 10.1016/j.trc.2017.02.019

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 11 September 2020
doi: 10.3389/fpsyg.2020.02216

Check for updates

Improving Drivers' Hazard Perception and Performance Using a Less Visually-Demanding Interface

Guy Cohen-Lazry and Avinoam Borowsky*

Human Performance Evaluation Lab, Industrial Engineering and Management Department, Ben-Gurion University of the Negev, Beer-Sheva, Israel

In-vehicle devices and infotainment systems occasionally lead to driver distraction, and as a result, increase the risk of missing on-road information. In the current study, a novel multi-touch interface for an in-vehicle infotainment system was evaluated, which potentially requires less visual attention and thus may reduce distraction and increase safety. The interface was compared with a functionally similar control interface in terms of hazard perception metrics and mental workload. Twenty-two participants drove a simulated route once with each system. During each drive, which included eight potentially-hazardous scenarios, participants were instructed to interact with one of the in-vehicle interfaces to perform phone calls or to navigate to specified destinations. Eye-gaze data were collected throughout the drive to evaluate whether participants detected the hazards while interacting with the in-vehicle interface, how much time they needed to identify them, and for how long they engaged with the secondary task. Additionally, after each drive, participants completed a NASA R-TLX questionnaire to evaluate their subjective workload during their engagement with the secondary tasks. Participants using the multi-touch interface needed less time to complete each secondary task and were quicker at identifying potential hazards around them. However, the probability of detecting hazards was similar for both interfaces. Finally, when using the multi-touch interface, participants reported lower subjective workload. The use of a multi-touch interface was found to improve drivers' performance in terms of identifying hazards quicker than the control condition. The road safety and driver distraction implications of this novel interface are discussed.

Keywords: hazard perception, in-vehicle interfaces, interface design, multi-touch interface, mental workload, driver distraction, eye movements

OPEN ACCESS

Edited by:
Oren Musicant,
Ariel University, Israel

Reviewed by:
Iveta Eimontaite,
Cranfield University, United Kingdom
Lee Skrypchuk,
Jaguar Land Rover, United Kingdom

****Correspondence:***
Guy Cohen-Lazry
guy.cohenlazry@gmail.com

Specialty section:
This article was submitted to Performance Science, a section of the journal Frontiers in Psychology

Received: *14 June 2020*
Accepted: *07 August 2020*
Published: *11 September 2020*

Citation:
Cohen-Lazry G and Borowsky A (2020) Improving Drivers' Hazard Perception and Performance Using a Less Visually-Demanding Interface. Front. Psychol. 11:2216. doi: 10.3389/fpsyg.2020.02216

INTRODUCTION

While driving, drivers occasionally engage with secondary tasks and become distracted. These tasks may be activities that relate to safety or performance, like using navigational aids (driving-related activities; Pfleging and Schmidt, 2015), or activities that are not related to driving, like phone conversations or radio tuning (non-driving-related activities; Pfleging and Schmidt, 2015). In the United States alone, nine people are killed daily in crashes related to driver distraction, and more than a 1,000 are injured (National Center for Statistics and Analysis, 2017). The most significant negative effect on drivers' performance is caused by distractions that are both visually and manually demanding (Klauer et al., 2006; Vegega et al., 2013). The diversion of drivers' gaze away from the forward roadway to the in-vehicle device "affects the degree to which drivers are able to perform primary driving tasks, such as event or object detection, and maintain vehicle control" (Vegega et al. 2013, p. 21).

The visual-manual driver distraction guidelines (National Highway Traffic Safety Administration, 2012), adopted and applied by most manufacturers, suggest that any visual-manual task that may be performed on a system, should be designed in such a way that it "can be completed by the driver while driving with glances away from the roadway of 2 s or less and a cumulative time spent glancing away from the roadway of 12 s or less" (National Highway Traffic Safety Administration, 2012, p. 10). One way to follow these guidelines is to use speech-based interfaces. Studies have shown that, for performing certain tasks, speech-based interfaces improve drivers' performance in aspects such as lateral positioning (Itoh et al., 2004), speed management (Gärtner et al., 2001), and hazard detection (Ranney et al., 2002). Also, speech-based interfaces reduce the time required to complete tasks and drivers' subjective workload (Itoh et al., 2004). Nevertheless, other studies have shown contradictory results. Yager (2013), for example, has tested drivers' distraction by asking drivers to engage in secondary tasks and to respond to occasional illuminating lights. Yager has shown that, even though using speech-based interfaces reduce drivers' reaction times to the illuminating light compared with manual-interfaces, they still react slower than drivers who do not engage in a data-entry task at all. In another study (Lee et al., 2001), the use of speech-based interfaces caused a 30% increase in drivers' reaction times to periodic braking of a lead vehicle and introduced a higher workload. In a study regarding cognitive distraction in driving (Strayer et al., 2014), the authors have found that the cognitive demands of speech-based interfaces pose a significant threat to traffic safety, when used for specific tasks such as texting and e-mailing.

In the current research, a novel approach is taken to reduce driver distraction when using an in-vehicle device. A new touch-based interface is evaluated [hereafter multi-touch interface (MTI)], which does not require drivers to gaze toward the screen and thus potentially reduces drivers' distraction. The MTI is designed as such that, in order to use any command, the driver places three fingers anywhere on the screen, and the system detects their absolute and relative locations and adapts to them. Then, by removing two fingers off the screen, the driver initiates one of three menus (functions) that is uniquely assigned to each finger. The menus, starting from the left-most finger, are a radio menu, a phone menu and a navigation menu. When a particular menu is selected, the driver can slide her finger either up, down, left, or right to select one of four predetermined selections from a star-like menu (i.e., favorites). Each phase is accompanied by an appropriate display in case the driver wishes to verify her actions visually. The fact that the MTI identifies the triple-touch wherever the driver places her fingers reduces the driver's need to gaze at the screen to search for specific touch-buttons spatially. This feature addresses a significant disadvantage of other touch-based interfaces, which require users to make almost the same number of glances toward them as tactile interfaces to perform tasks (Bach et al., 2008).

A driving simulator study was conducted to compare the MTI with a typical in-vehicle interface [hereafter control interface (CI)] to test the hypothesis that the MTI will help drivers to complete a predefined secondary task quicker than the CI and that the MTI will lead to better hazard perception performance than the CI. The CI that was chosen for this study was a popular infotainment application that was downloaded from Google Play over half a million times and included both a visual-manual and speech-based interfaces. Three relevant measures were chosen to compare the interfaces. First, the time drivers needed to complete a task was recorded, since minimizing the secondary task's duration is an effective method for reducing driver distraction (National Highway Traffic Safety Administration, 2012) and increasing safety. For the second metric, an eye-tracker was used to measure hazard perception, a measure, which is highly correlated with traffic safety (Horswill and McKenna, 2004; Horswill et al., 2015). Third, whenever a hazard was identified, we measured the time participants needed to identify it, as another indication of hazard perception quality. Finally, the NASA R-TLX was used to test whether the fact that the MTI requires less visual attention also reduces drivers' workload compared with the CI.

Twenty-two participants were asked to drive two simulated routes, once using each system, during which they were instructed by the experimenter to initiate phone-calls or change the destination in the navigation system. The CI was used either in its visual-manual modality or its speech-based modality. Throughout the drive, various scenarios that required drivers' attention (not necessarily their action) were initiated, during which drivers' gaze and task performance were measured. Since the MTI potentially requires fewer number of glances toward the in-vehicle display than the CI, it was expected that:

Hypothesis 1: when using the MTI, participants will complete the predefined secondary tasks faster;
Hypothesis 2: when using the MTI, participants will be more likely to detect hazards;
Hypothesis 3: when using the MTI, participants will identify hazards faster; and
Hypothesis 4: when using the MTI, participants will be experience lower levels of workload than the CI.

MATERIALS AND METHODS

Participants

Twenty-two undergraduate students from the Ben-Gurion University (BGU) of the Negev (12 female, ages 21–28 years, $M = 25.5$, $SD = 2.11$) volunteered to a 1-h session, for which they were compensated by course credit. All participants had normal or corrected to normal visual acuity and normal contrast sensitivity. Participants who had glasses were asked to wear contact lenses for the experiment. Participants reported having a valid driver's license for at least 3 years ($M = 7.34$, $SD = 2.10$), and driving, on average, at least twice a week.

Apparatus

The experiment was conducted using a medium-fidelity desktop driving simulator. Participants were seated on a gaming seat 1.1 m away from three 24" LCDs, providing ~90° of horizontal view.

The driving simulator was controlled *via* a G27 Logitech steering wheel and a set of pedals. The driving environment was generated using a simulator software provided by Realtime Technologies Inc. (RTI; Royal Oak, MI). The experimental route was a 15-min long drive in an urban environment, in which participants were instructed to keep the right lane whenever possible and drive as they would typically do in similar real-world situations.

Participants' point of gaze was monitored using a Dikablis light-weight head-mounted eye-tracker (Ergoneers Inc., Manching, Germany). The eye-tracker's software synchronizes data regarding participants' gaze with the scene displayed on the simulator screen to provide a measure of where participants' point of gaze is located at any given moment of the drive. The two types of interfaces (the MTI and the CI) were installed on a 7" Lenovo tablet. The tablet was positioned to the right of the steering wheel, where an in-vehicle device is commonly located (**Figure 1**).

Driving Scenarios

Participants drove two simulated routes, during which they encountered 12 driving scenarios (eight were hazardous scenarios, and four were filler scenarios that did not include any hazard). Participants were also asked to perform tasks using the in-vehicle tablet 12 times during each drive. Eight out of the twelve tasks were given 4 s before a scenario (target or filler). This resulted in eight tasks for which we could measure participants' hazard perception performance, and four tasks and four scenarios which served as decoys.

The same eight scenarios were used for both drives, but in a randomized order, to allow a direct comparison between the interfaces. Thus, each participant experienced each scenario twice, once while performing a task using the MTI and once while performing the same task using the CI.

In-Vehicle Tasks

During a drive, participants were verbally instructed by the relevant system to complete two types of tasks. Participants were asked either to make a phone call to one of four pre-programmed numbers (four tasks) or to change the destination in the navigation systems to one of four pre-programmed options (four more tasks). When using the MTI, the entry method was always the multi-touch-based interface. When using the CI, four tasks (Set 1) were completed using a visual-manual (touch) interface, and four tasks (Set 2) were accomplished using a speech-based interface. All tasks using the touch interface required three taps on the screen: one tap to choose the required "app" (i.e., navigation or phone), a second tap to enter the "favorites" screen, and a third tap to choose the requested destination or contact. The speech-based interface required only one click to activate the system's "listening mode." **Appendix A** provides a comprehensive description of the eight scenarios, the type of the associated task and the modality used for that task when using the CI.

FIGURE 1 | The simulator setup, with the tablet installed to the right of the steering wheel.

Workload Evaluation

To assess levels of workload, participants filled in the NASA R-TLX. This questionnaire consists of six Likert-style items measuring factors such as "mental demand," "effort," and "frustration" on a scale ranging from one (low) to nine (high). Participants filled in the same questionnaire twice, once after using the MTI and once after using the CI.

Experimental Design

A two (interface type) by eight (scenario) within-subject experiment was designed to minimize the effect of individual differences, and each participant experienced the same eight scenarios using both interfaces. The order of scenarios was randomized so that the two drives did not resemble one another. Additionally, the order of the interfaces that the drivers had to use was counterbalanced between participants. However, for each scenario, the type of task and the modality used (when using the CI) remained the same. For example, during the scenario where a car was pulling into the road from the right shoulder, participants always had to make a phone call, and the modality was always visual-manual (for a complete list, see **Appendix A**).

Procedure

Upon arrival at the lab, participants were briefed about the study and were asked to sign an informed consent form. During a 25-min learning session, participants were introduced to both interfaces and to the presets pre-programmed into them (four phone numbers and four destinations). Participants were allowed to practice and engage with both interfaces and to ask questions if they had any. Participants were also introduced to the simulator, where they were allowed to drive for 10 min without the secondary tasks and for 5 min while performing secondary tasks. Then, the two 15-min experimental drives began. Following each drive, participants were asked to fill in the NASA R-TLX questionnaire. After finishing the two driving sessions and filling the questionnaires, participants were debriefed and were allowed to ask questions about the experiment and the study's goals. The study was reviewed and approved by the internal Human Subjects Research Committee at BGU.

Dependent Variables

Task duration was calculated as the time interval between the initiation of a pre-recorded auditory request to complete a task

(4 s before the beginning of a scenario) and when the task was completed. The end of the task was defined as either releasing the last finger off the screen (MTI), making the last click (CI – visual-manual interface) or finishing the speech command (CI – speech-based interface). A binary variable was used to evaluate hazard identification. Hazard identification was either marked as a success (i.e., participants noticed the hazard in the scenario, "1") or as a failure (i.e., participants did not notice the hazard in the scenario, "0"). A value of "1" was assigned whenever the participant's gaze was fixated at the hazard for more than 100 ms (International Organization for Standardization, 2014). Note that since we used an eye-tracking system to evaluate hazard identification, it can only account for identification using the central vision. It may well be that drivers were able to discern the hazard sooner using their peripheral vision, but this would, in any case, require the shift of the central vision system to the location of the hazard to complete its recognition. The time it took participants to identify a hazard was defined as the time interval between the beginning of a scenario (when the hazard instigator became visible) and the participant's first glance towards it.

Analysis

Analyses were all conducted using SPSS version 23 (IBM Corp., 2015). A repeated measures mixed-model regression was used, using two independent variables as fixed effects: interface type (MTI or CI) and task type (navigation or phone). Two variables (participants and scenario number) were also included in the models as random effects. The interactions between the variables were not relevant for the current study and were left out of the models. Furthermore, since the MTI was used using only one modality, and the CI was used using two different modalities, the analyses had to be separated. Thus, in order to compare between the two types of interfaces in terms of the various dependent measures, each CI modality (visual-manual or speech-based) was compared to the MTI using a separate regression model within the generalized linear mixed model (GLMM). These two separate regression models were applied once for the secondary task duration (log-linear regression), once for hazard identification (logistic regression) and once for the hazard identification time (log-linear regression).

Each regression model included task type and interface type as fixed effects and participants and scenarios as random effects. All second-order interactions were included in the models. Overall, six different models were used in the analyses (**Appendix B** provides an extended description of the analyzed models). The significance level was set to $\alpha = 0.05$. Final models were achieved using a backward elimination procedure, and *post hoc* pairwise comparisons were corrected for multiple comparisons using the sequential Bonferroni procedure.

RESULTS

Task Duration

Two linear regression models included a log transformation of the secondary task's duration as the dependent variable. With regard to the visual-manual modality (Set 1), the final model supported Hypothesis 1 and revealed that the interface type's main effect was significant [$F(1, 96) = 44.6$, $p < 0.01$], as participants were faster to complete the secondary task when using the MTI ($M = 2.95$ s, $SD = 2.69$) than when using the CI ($M = 7.34$ s, $SD = 4.14$). Task type [$F(1, 96) = 0.63$, $p <$ n.s], scenario [$F(2, 96) = 0.02$, $p <$ n.s], and participant [$F(20, 96) = 1.56$, $p <$ n.s] were all insignificant. Similarly, with regard to the speech-based modality (Set 2), Hypothesis 1 was again supported as it was found that interface type had a significant effect on secondary task's duration [$F(1, 87) = 29.8$, $p < 0.01$], with participants performing the task faster when using the MTI ($M = 3.97$ s, $SD = 2.98$) than when using the CI ($M = 7.74$ s, $SD = 3.96$). Task type [$F(1, 87) = 3.55$, $p <$ n.s], scenario [$F(2, 87) = 1.12$, $p <$ n.s], and participant [$F(20, 87) = 1.11$, $p <$ n.s] were all insignificant. Means task durations are presented in **Figure 2**. The averages presented in this figure and every other figure in this paper are based on raw data means and not on the model estimates.

Hazard Identification Probability

Two logistic regression models included hazard identification as the dependent variable. With regard to the visual-manual modality, the final model of the first logistic regression did not support Hypothesis 2, as it revealed that interface type did not significantly affect participants' probability of detecting a hazard, $\chi^2(1) = 0.19$, $p =$ n.s; participants identified 61% of all hazards when using the MTI and 57% of all hazards when using the CI. Among all other variables, scenario was the only significant variable, $\chi^2(3) = 10.93$, $p < 0.01$, whereas task type $\chi^2(1) = 0.01$, $p <$ n.s and participant $\chi^2(20) = 4.51$, $p <$ n.s were both insignificant. Similarly, with regard to the speech-based interface, Hypothesis 2 was again not supported since the final model of the second logistic regression revealed that interface type did not significantly affect participants' probability of detecting a hazard, $\chi^2(1) = 0.57$, $p =$ n.s; participants identified

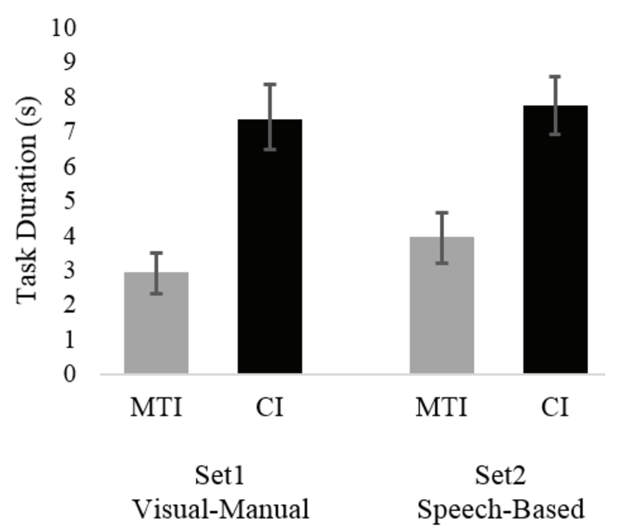


FIGURE 2 | Average task completion duration for the two interfaces and the two scenario sets. Bars represent standard error.

72% of all hazards when using the MTI and 70% of all hazards when using the CI. Again, among all other variables, scenario was the only significant variable, $\chi^2(3) = 14.11$, $p < 0.01$ whereas task type $\chi^2(1) = 4.12$, $p <$ n.s and participant $\chi^2(20) = 7.95$, $p <$ n.s were both insignificant. Hazard detection rates for the different scenarios are presented in **Figure 3**.

Hazard Identification Time

Two linear regression models included a log transformation of hazard identification time as the dependent variable. With regard to the manual-visual modality, the final model of the first linear regression supported Hypothesis 3 and revealed that the effect of interface type was significant [$F(1, 91) = 47.16$, $p < 0.01$], with participants identifying the hazards quicker when using the MTI ($M = 0.65$ s, $SD = 1.70$) than when using the CI ($M = 1.20$ s, $SD = 2.63$). Task type [$F(1, 91) = 0.62$, $p <$ n.s], scenario [$F(2, 91) = 0.74$, $p <$ n.s], and participant [$F(20, 91) = 1.52$, $p <$ n.s] were all insignificant. Similar results, supporting Hypothesis 3, were found for the second linear regression model with regard to the speech-based modality such that interface type had a significant effect on hazard identification time [$F(1, 93) = 31.93$, $p < 0.05$], with participants identifying the hazards quicker when using the MTI ($M = 0.45$ s, $SD = 1.27$) than when using the CI ($M = 0.72$ s, $SD = 0.75$). Task type [$F(1, 93) = 3.77$, $p <$ n.s], scenario [$F(2, 93) = 1.81$, $p <$ n.s], and participant [$F(20, 93) = 2.05$, $p <$ n.s] were all insignificant. Estimated means of the hazard identification times are presented in **Figure 4**.

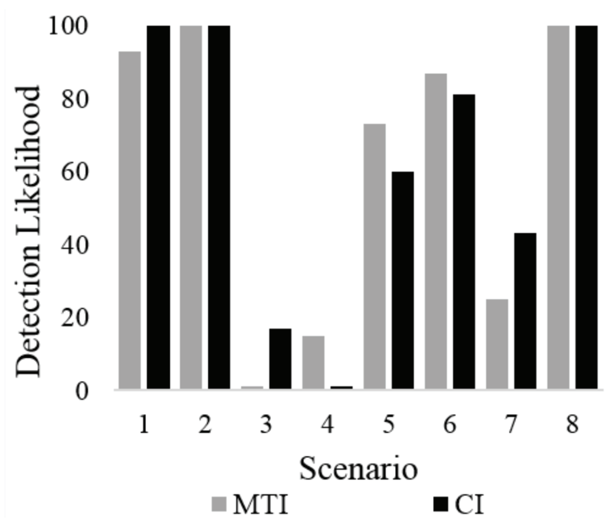


FIGURE 3 | The probability of a participant in either group to detect a hazard, presented per the eight different scenarios.

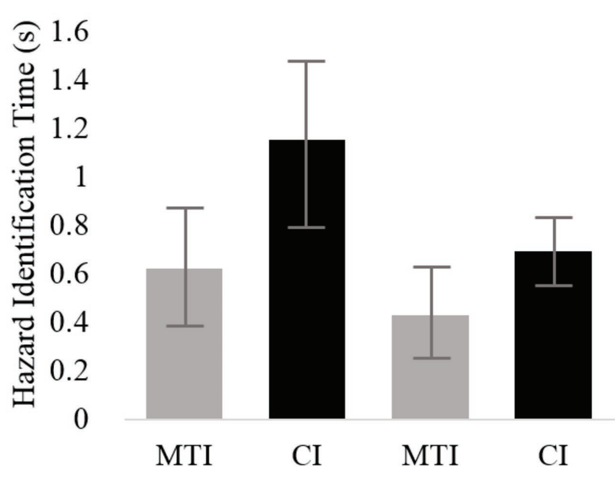


FIGURE 4 | Average hazard-identification durations for the two interfaces and the two scenario sets. Bars represent standard error.

NASA R-TLX

The fourth item in the NASA R-TLX, regarding task performance, is rated on an inverse scale (1 – high performance, 9 – low performance) and was inversed before data analysis. To compare the workload that participants experienced while using each interface, a repeated-measures analysis of variance (ANOVA) was conducted with R-TLX ratings as the dependent variable and interface type as a within-subject fixed factor. As hypothesized (Hypothesis 4), there was a significant effect of interface type on participants' workload ratings, $F(1, 87) = 16.64$, $p < 0.01$. The difference between the questionnaire's items was insignificant, $F(5, 87) = 2.04$, $p <$ n.s. The NASA R-TLX ratings for each item, presented in **Figure 5**, show that the MTI scored lower than the CI across all effort and pressure factors, indicating lower workload. When asked about their task performance using each one of the in-vehicle interfaces (fourth item on

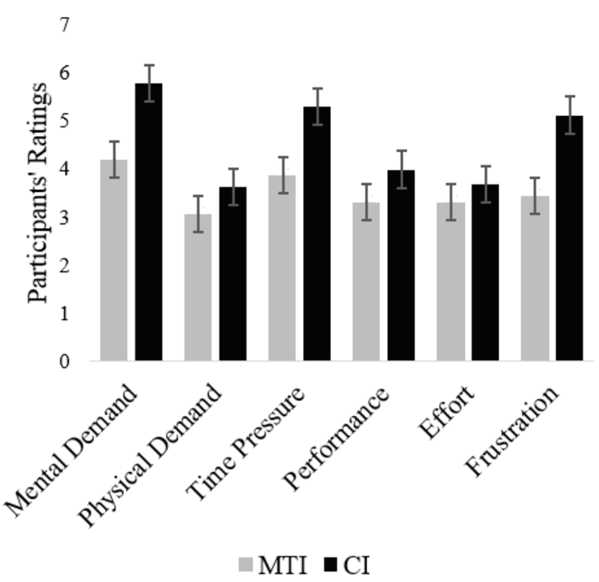


FIGURE 5 | Average NASA R-TLX ratings for the six factors, comparing the multi-touch interface (MTI) and the control interface (CI). Bars represent standard error.

the NASA R-TLX), participants rated the CI significantly higher, meaning they thought that their performance was better when interacting with it than when interacting with the MTI (**Figure 5**). This result is intriguing since the aforementioned objective measures have indicated better performance when using the MTI.

DISCUSSION

This study was aimed at evaluating a novel concept of interfaces for in-vehicle devices that target the reduction of driver distraction and the increase of safety. The interface was compared with a control interface allowing both visual-manual and speech-based interactions. Since all findings were similar for both types of input modalities, from hereon, we will disregard this difference between modalities and only discuss the differences between the two systems. Results point to a significant improvement in three distraction-related measures. First, in-line with Hypothesis 1, when using the MTI, the time participants needed to complete each task was significantly reduced. One possible explanation for the longer time participants needed to complete tasks using the CI is that, to operate it, participants had to visually locate and aim their finger at the right touch-button three times. The MTI, on the other hand, could be operated anywhere on the screen, without glancing towards it even once. Second, as predicted in Hypothesis 3, participants identified hazards faster when using the MTI. Finally, as expected in Hypothesis 4, the workload participants reported was also significantly reduced when the MTI was used. The decrease in task duration time is an essential aspect in designing in-vehicle interfaces, and it has been shown that the longer people glance away from the road to perform a secondary task, the likelihood of a crash increases (e.g., Simons-Morton et al., 2014). Burns et al. (2010) have discussed the negative effect of long task-durations and have suggested that a key manner in which this risk could be decreased is by designing interfaces that support quicker performance. The results of the current study are consistent with their claim, showing that indeed a shorter task-duration may lead to an increase in safety. Despite these essential improvements, Hypothesis 2 was disconfirmed as we did not find any advantage for the MTI concerning drivers' probability of identifying a hazard. One possible explanation regards the overall difficulty of identifying hazards in the various scenarios.

By examining the identification probabilities in **Figure 3**, it is evident that in six out of the eight scenarios, the probability of identifying a hazard was either very high or very low. This suggests that, in most cases, identifying a hazard was either very easy (ceiling effect) or very difficult (floor effect) when using both interfaces, thus reducing the possibility of revealing significant differences between them. Nevertheless, despite this lack of difference between the interfaces, when using the MTI, participants were faster at identifying hazards than when they were using the CI. Possibly, this difference is rooted in the experimental design.

Throughout the experiment, the time between the requirement to complete a task and the initiation of a scenario was fixed at 4 s. Additionally, as seen in **Figure 2**, participants using the MTI needed, on average, less than 4 s to complete a task, whereas participants using the CI needed more than 7 s. Thus, it seems that, on average, participants using the MTI completed their tasks before the initiation of the hazardous scenario. Hence, throughout the entire duration of the scenario, participants were not distracted by a secondary task and could divert all their attention to the road. Conversely, participants using the CI were still engaged in performing the secondary task for a few more seconds when the scenario started. Nevertheless, they were still able to identify the hazard before the end of the hazardous scenario. These task duration differences between the groups explain why the hazard identification times were shorter for the MTI even though the identification probabilities were similar for both interfaces. Participants using the MTI had their full attention allocated to monitoring the environment throughout the entire scenario, whereas participants using the CI had to divide their attention between the secondary task and the road environment, at least for a few seconds. Still, even though participants who were using the CI began monitoring the environment later, the relatively long duration of the scenarios (~10 s) and the aforementioned ceiling and floor effects allowed them to identify the hazards at a similar likelihood to that of participants who were using the MTI. This might explain the similarities in identification probabilities alongside with the differing identification times.

This analysis of task duration may also put into perspective the results regarding the hazard identification times. Since drivers using the MTI completed their tasks before the initiation of the scenario, they did not, in fact, identify hazards while performing secondary tasks. Therefore, their superiority in identifying hazards faster than participants using the CI may be an effect of performing a single task and not two tasks simultaneously, which is a well-known advantage in hazard perception tasks (e.g., Burge and Chaparro, 2018). Further research is required to determine whether the MTI also reduces hazard identification times during the performance of secondary tasks. Nevertheless, even if the faster hazard identification times are the result of performing just one task, this result still denotes an advantage in favor of the MTI.

This study's results suggest an advantage for a multi-touch-only interface over common tactile interfaces. However, several limitations have to be acknowledged. First, the sample size and its homogeneity (undergraduate students) limit the results' generalizability. Second, using the touch-only interface may pose requirements (e.g., a certain level of dexterity) or have implications that were not studied here. Jin et al. (2007), for example, have shown that touch interfaces have to be designed differently when designing for the elderly. This aspect of the interface was not examined in this study and should be a part of future studies. Third, while this study focused on the real-time hazard perception-related effects, the introduction of a new interface probably has long-term effects as well. Specifically, future research should examine people's attitudes towards the interface (e.g., their trust or annoyance with it), and whether they find it useful. Fourth, due to technical limitations, the order of scenarios was not randomized between participants. Although this could have led to a learning effect, an examination of **Figure 3** suggests that even if such an effect existed, it affected both groups similarly, as indicated by their similar detection rates throughout all scenarios. Finally, while the study compared the MTI with a control interface, it did not use a no-task reference condition as a baseline for drivers' non-distracted performance. Therefore, although the MTI showed significant advantages when compared to the CI, it is not possible to tell how distracting the system is when compared to driving without a non-driving-related secondary task.

This study, thus, has shown that the MTI, which is based on a non-visual MTI, has two advantages over a representative in-vehicle touchscreen interface and a speech-based interface. Participants using MTI needed less time to complete phone and navigation tasks and also experienced a lower workload. These two variables are closely related to the concept of hazard perception and thus suggest a significant potential for systems such as the MTI in reducing driver distraction and enhancing safety (e.g., Burns et al., 2010) An analysis of the results regarding hazard identification time and hazard identification probabilities pointed to issues in the experimental design that provide possible alternative explanations for some of the results. Thus, concerning these two variables, we are currently unable to determine whether the MTI does or does not have an advantage over the CI. Further studies will be designed to allow an exploration of these other variables using different scenario designs and timings.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Human Subjects Research Committee of the Ben-Gurion University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

Bach, K. M., Jæger, M. G., Skov, M. B., and Thomassen, N. G. (2008). "You can touch, but you can't look" in *The twenty-sixth annual CHI Conference on Human Factors in Computing Systems—CHI'08*; April 5-10, 2008; 1139–1148.

Burge, R., and Chaparro, A. (2018). An investigation of the effect of texting on hazard perception using fuzzy signal detection theory (fSDT). *Transp. Res. Part F Traffic Psychol. Behav.* 58, 123–132. doi: 10.1016/j.trf.2018.05.033

Burns, P., Harbluk, J., Foley, J. P., and Angell, L. (2010). "The importance of task duration and related measures in assessing the distraction potential of in-vehicle tasks" in *The Second International Conference on Automotive User Interfaces and Interactive Vehicular Applications*; November 11-12, 2010; 12–19.

Gärtner, U., König, W., and Wittig, T. (2001). "Evaluation of manual vs. speech input when using a driver information system in real traffic" in *Proceedings of the First International Driving Symposium on Human Factors In Driver Assessment, Training and Vehicle Design: Driving Assessment*; August 14-17, 2001; 7–14.

Horswill, M. S., Hill, A., and Wetton, M. (2015). Can a video-based hazard perception test used for driver licensing predict crash involvement? *Accid. Anal. Prev.* 82, 213–219. doi: 10.1016/j.aap.2015.05.019

Horswill, M. S., and McKenna, F. P. (2004). "Drivers' hazard perception ability: situation awareness on the road" in *A cognitive approach to situation awareness*, eds. S. Banbury and S. Tremblay (Aldershot, UK: Ashgate), 155–175.

IBM Corp. (2015). *IBM SPSS statistics for windows, version 23.0*. Armonk, NY: IBM Corp.

International Organization for Standardization (2014). Road vehicles—measurement and analysis of driver visual behaviour with respect to transport information and control systems (ISO/FDIS Standard No. 15007).

Itoh, K., Miki, Y., Yoshitsugu, N., Kubo, N., and Mashimo, S. (2004). Evaluation of a voice-activated system using a driving simulator (SAE paper 2004-01-0232). Warrendale, PA: Society of Automotive Engineers.

Jin, Z. X., Piocher, T., and Kiff, L. (2007). "Touch screen user interfaces for older adults: button size and spacing" in *International Conference on Universal Access in Human-Computer Interaction*; July 22-27, 2002; Berlin Heidelberg: Springer, 933–941.

Klauer, S. G., Dingus, T. A., Neale, V. L., Sudweeks, J. D., and Ramsey, D. J. (2006). The impact of driver inattention on near crash/crash risk: an analysis using the 100-car naturalistic driving study data.

Lee, J. D., Caven, B., Haake, S., and Brown, T. L. (2001). Speech-based interaction with in-vehicle computers: the effect of speech-based e-mail on drivers' attention to the roadway. *Hum. Factors J. Hum. Factors Ergon. Soc.* 43, 631–640. doi: 10.1518/001872001775870340

National Center for Statistics and Analysis (2017). Distracted driving 2015. (Traffic Safety Facts Research Note. Report No. DOT HS 812 381). Washington, DC: National Highway Traffic Safety Administration.

National Highway Traffic Safety Administration (2012). Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices.

Pfleging, B., and Schmidt, A. (2015). (Non-) driving-related activities in the car: defining driver activities for manual and automated driving in Chi'15, 5–8.

Ranney, T. A., Harbluk, J. L., and Noy, Y. I. (2002). "Effects of voice technology on test track driving performance: implications for driver distraction" in *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*; September 29 - October 4, 2002; (Santa Monica, CA), 1814–1818.

Simons-Morton, B. G., Guo, F., Klauer, S. G., Ehsani, J. P., and Pradhan, A. K. (2014). Keep your eyes on the road: young driver crash risk increases according to duration of distraction. *J. Adolesc. Health* 54, S61–S67. doi: 10.1016/j.jadohealth.2013.11.021

Strayer, D. L., Turrill, J., Coleman, J. R., Ortiz, E. V., and Cooper, J. M. (2014). *Measuring cognitive distraction in the automobile II: Assessing in-vehicle voice-based interactive technologies*. Washington, DC: AAA Foundation for Traffic Safety.

Vegega, M., Jones, B., and Monk, C. (2013). Understanding the effects of distracted driving and developing strategies to reduce resulting deaths and injuries: A report to congress. (Report No. DOT HS 812 053). Washington, DC: National Highway Traffic Safety Administration.

Yager, C. (2013). An evaluation of the effectiveness of voice-to-text programs at reducing incidences of distracted driving. US Department of Transportation.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX A

Out of the eight hazardous scenarios, four included a materialized hazard, like a car on the shoulders pulling out into the road or a car that fails to maintain its lane position or speed properly. The four other scenarios included latent hazards, like workers standing in a construction area next to the road (that could jump into the participant's lane at any moment) or a playground next to the road (from where a child could run into the road). **Table A1** describes the eight scenarios.

TABLE A1 | Driving scenarios, tasks, and input modalities.

	Scenario	Type of task	Experimental modality	Control system modality
Set 1	**A bus station close to the road (scenario 7)** A detection was defined as looking behind the bus station to make sure no pedestrians walk there	NAV	MT	VM
	A tree hiding the sidewalk during a turn (scenario 6) A detection was defined as looking for a hidden pedestrian behind the tree	NAV	MT	VM
	A construction site (scenario 4) A detection was defined as gazing at the site to make sure no workers barge into the road	PHC	MT	VM
	A car pulling into the road (scenario 2) A detection was defined as observing the car as it pulled out and into the road	PHC	MT	VM
Set 2	**A car unable to maintain lane and speed (scenario 8)** A detection was defined as making more than one observation of that vehicle	NAV	MT	SB
	A playground near the road (scenario 1) A detection was defined as gazing at the site to make sure no workers barge into the road	NAV	MT	SB
	A stopped truck on the right lane (scenario 5) A detection was defined as observing the car as the driver passed by it	PHC	MT	SB
	A phone booth near the road (scenario 3) A detection was defined as looking for a pedestrian to walk from within or behind the booth	PHC	MT	SB

VM, visual manual; MT, multi-touch; SB, speech-based; NAV, navigation; PHC, phone call.

APPENDIX B

Two different general linear mixed models were used to analyze each dependent variable. One regression was used to model the differences between the experimental system and the control system when using the visual-manual modality in both systems, and the other was used to model the differences between the systems when using the speech-based modality in the control system and the multi-touch modality in the experimental system. Notably, participants in the experimental system used only the visual-manual modality. These two separate regression models were applied once for the secondary task completion time (log-linear regression) dependent variable, once for hazard identification (logistic regression) dependent variable, and once for the hazard identification time (log-linear regression) dependent variable. Overall, six models were analyzed, as shown in **Table A2**.

TABLE A2 | A summary of the regression models used for the analyses.

		Dependent variable		
		Task completion time	**Hazard-identification**	**Hazard-identification time**
Control system interface	Visual-manual (Set 1 scenarios)	Regression model 1	Logistic regression model 1	Regression model 3
	Speech-based (Set 2 scenarios)	Regression model 2	Logistic regression model 2	Regression model 4

frontiers in Psychology

ORIGINAL RESEARCH
published: 18 September 2020
doi: 10.3389/fpsyg.2020.571961

Check for updates

Effects of an Unexpected and Expected Event on Older Adults' Autonomic Arousal and Eye Fixations During Autonomous Driving

Alice C. Stephenson[1], Iveta Eimontaite[1], Praminda Caleb-Solly[2], Phillip L. Morgan[3], Tabasum Khatun[1], Joseph Davis[1] and Chris Alford[1]*

[1] Health and Applied Sciences, University of the West of England, Bristol, United Kingdom, [2] Bristol Robotics Laboratory, University of the West of England, Bristol, United Kingdom, [3] Human Factors Excellence (HuFEx) Research Group, and Centre for Artificial Intelligence, Robotics and Human-Machine Systems (IROHMS), School of Psychology, Cardiff University, Cardiff, United Kingdom

OPEN ACCESS

Edited by:
Oren Musicant,
Ariel University, Israel

Reviewed by:
Maud Ranchet,
Institut Français des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux (IFSTTAR), France
Michael Wagner,
Ariel University, Israel
Hannes Devos,
The University of Kansas, United States

***Correspondence:**
Alice C. Stephenson
alice.stephenson@uwe.ac.uk

Specialty section:
This article was submitted to Performance Science, a section of the journal Frontiers in Psychology

Received: 12 June 2020
Accepted: 26 August 2020
Published: 18 September 2020

Citation:
Stephenson AC, Eimontaite I, Caleb-Solly P, Morgan PL, Khatun T, Davis J and Alford C (2020) Effects of an Unexpected and Expected Event on Older Adults' Autonomic Arousal and Eye Fixations During Autonomous Driving. Front. Psychol. 11:571961. doi: 10.3389/fpsyg.2020.571961

Driving cessation for some older adults can exacerbate physical, cognitive, and mental health challenges due to loss of independence and social isolation. Fully autonomous vehicles may offer an alternative transport solution, increasing social contact and encouraging independence. However, there are gaps in understanding the impact of older adults' passive role on safe human–vehicle interaction, and on their well-being. 37 older adults (mean age ± SD = 68.35 ± 8.49 years) participated in an experiment where they experienced fully autonomous journeys consisting of a distinct stop (an unexpected event versus an expected event). The autonomous behavior of the vehicle was achieved using the Wizard of Oz approach. Subjective ratings of trust and reliability, and driver state monitoring including visual attention strategies (fixation duration and count) and physiological arousal (skin conductance and heart rate), were captured during the journeys. Results revealed that subjective trust and reliability ratings were high after journeys for both types of events. During an unexpected stop, overt visual attention was allocated toward the event, whereas during an expected stop, visual attention was directed toward the human–machine interface (HMI) and distributed across the central and peripheral driving environment. Elevated skin conductance level reflecting increased arousal persisted only after the unexpected event. These results suggest that safety-critical events occurring during passive fully automated driving may narrow visual attention and elevate arousal mechanisms. To improve in-vehicle user experience for older adults, a driver state monitoring system could examine such psychophysiological indices to evaluate functional state and well-being. This information could then be used to make informed decisions on vehicle behavior and offer reassurance during elevated arousal during unexpected events.

Keywords: autonomous vehicle, eye tracking, heart rate, human–machine interaction, human–machine interface, older adults, skin conductance level, trust

INTRODUCTION

The private car is a vital element for the mental and physical well-being of older adults. Due to reduced mobility, the public transport system can be inconvenient and inaccessible (Broome et al., 2009). For example, challenges such as walking to a bus stop, or getting on and off a bus, can cause significant problems for adults with mobility issues. As such, driving provides access to local services, social events, and encourages participation in out-of-home activities. As well as the practical benefits to driving, research has indicated several affective advantages such as feelings of sensation, power, and youthfulness (Eisenhandler, 1990; Steg, 2005; Bergstad et al., 2011). However, age-related declines in cognitive, visual capacities, physical disability, and illness, subsequently impact driving ability as it becomes more physically and cognitively demanding. The possibility of becoming a non-driver rises with age (Anstey et al., 2006), and some drivers choose to restrict their driving (Dellinger et al., 2001). Driving cessation can have a negative impact on mobility and well-being, and feelings of isolation can be amplified (Qin et al., 2019). Some older adults find it more difficult to leave the home and stop participating in local or social activities (Marottoli et al., 2000), which in turn leads to a poorer quality of life. Consequently, ceasing driving can rapidly exacerbate physical, cognitive, and mental health challenges, and loss of independence.

Autonomous vehicles (AVs) promise to improve driving safety and efficiency by effectively removing the human from the driving task altogether. The role of the human is dependent on the level of autonomy of the vehicle. The Society of Automotive Engineering illustrated six levels of automation, ranging from 0 "No automation," to 5 "Full automation" (SAE, 2018). While Levels 2 and 3 require a driver to monitor the environment and take back control of the vehicle when requested; Levels 4 and 5 requires little to no input from the driver. As different cognitive and physical demands of the task are replaced by automation elements, AVs may offer an alternative transport solution for the older population. By enabling a viable transportation option, mobility is likely to be restored enabling older adults to lead more independent lives (Smith and Anderson, 2017). In turn, this should promote participation in local and social events, encouraging feelings of social inclusion and satisfaction.

While the advent of AV technology offers many potential advantages for an aging population, the impact of the role as a passive driver on safe human–vehicle interaction and older adults' well-being is not fully understood. Previous research has indicated the negative impact of partially automated vehicles on safe vehicle interaction, where the human is expected to stay 'in-the-loop' and take back control of the vehicle during expected or unexpected situations. Yet during Level 5 autonomous driving, the potentially negative consequences of a takeover request are eliminated due to the fully automatic capabilities of the vehicle. SAE (2018) refers to the in-vehicle user as a 'passenger' rather than a form of driver. Although the negative consequences of a takeover should be eliminated, previous research has demonstrated that full automation still has a significant impact on cognitive and affective functional state. From a cognitive perspective, studies have demonstrated that automation can increase mental underload and promote deficit attentional strategies (Young and Stanton, 2002). Automation has also been shown to encourage complacency and overtrust of a system (Parasuraman and Manzey, 2010), as well as increase frustration levels, particularly when automation cannot be overridden (Comte, 2000). These issues are potentially amplified in an older adult population with aging-related impairments, as they are more likely to rely on automated systems (McBride et al., 2011), find it more difficult to perform two or more tasks simultaneously (Kramer and Madden, 2008), and are more prone to lack understanding of advanced technology (Mann et al., 2007). Moreover, research has indicated that older adults have concerns using AVs due to issues related to trust and confidence, such as not having an operator nearby during failures (Faber and van Lierop, 2020). As such, some autonomous driving situations may initiate feelings of anxiety, and repeated activation of a stress response could be potentially damaging to their health and well-being (Cohen et al., 2007).

Considering the significant impact on a passenger's functional state, a driver state monitoring (DSM) system including cognitive and affective indices to improve safety and well-being has been proposed (Collet and Musicant, 2019). A DSM system continuously monitors a user using a hybrid of measures including biological (e.g., muscle activity) and physical measures (e.g., blink frequency). By synthesizing and classifying functional state, the system can provide feedback to the passenger or adapt vehicle behavior. DSM systems have traditionally been applied during manual driving scenarios to detect fatigue and inattention. Situations such as night-time driving (Phipps-Nelson et al., 2011), prolonged driving (Finkleman, 1994), and extreme temperatures (Xianglong et al., 2018) can induce fatigue; whereas mobile phones (Strayer and Drews, 2007), in-vehicle systems (Arexis et al., 2017), and eating (Tay and Knowles, 2004) can induce inattention. In a manual driving scenario, a DSM system can use remote sensors to monitor fatigue behaviors such as prolonged eyelid closures and yawning. Upon detection of these behaviors, the system can warn the driver, or others, of the potentially dangerous situation.

Detecting fluctuations in cognitive and affective states with a DSM system has many potential benefits for improving passenger well-being and safety during Level 5 driving. The information about a passenger's state could be used to modify in-vehicle information or vehicle behavior. For example, the in-vehicle system could provide reassurance at the appropriate time to reduce stress levels. It could also adapt vehicle behavior to improve comfort, e.g., leaving more headway between the vehicle in front. Alternatively, the system could identify the passenger's cognitive load to present the optimum amount of feedback or information. For example, it could choose between auditory or/and visual feedback modalities depending on what the user is doing or how they are feeling. If the system detects lapses in attention, it could encourage automation monitoring. Manual driving research has provided promise toward the technical development of real-time unobtrusive sensors to detect driver state, however, additional studies are now needed to uncover the impact of *autonomous driving* scenarios on human cognition and arousal.

It is not possible to measure typical performance indicators of functional state during Level 5 autonomous driving as the passenger is not required to carry out manual driving behaviors (i.e., speed or lateral position changes). Capturing the human response in real-time may disentangle functional states during dynamic autonomous driving scenarios. To this end, most studies have utilized continuous measures such as eye gaze and indices of physiological arousal.

Cognitive underload and attentional deficits during automated driving have been demonstrated by measures of visual attention indexed by ocular behaviors. Visual strategy and the distribution of fixation points can provide information about where and when participants are shifting their attention. In general, eye gaze has been shown to be directed away from the driving environment (De Winter et al., 2014), and horizontal gaze dispersion is greater (Louw and Merat, 2017), when compared to manual driving, indicating lower situation awareness and reduced load. However, cognitive load and attentional allocation evolves over time with changing task demands. For example, Strauch et al. (2019) found that participants fixated in safety-critical areas (i.e., the steering wheel and forward roadway) more so during automated versus manual driving.

Several studies have attempted to understand the associations between constructs related to automation monitoring and attention itself. For example, participants with a high level of trust tended to monitor the road less (e.g., Helldin et al., 2013; Hergeth et al., 2016; Körber et al., 2018; Walker et al., 2019); and longer fixation duration and higher fixation count on the driving environment were associated with greater situation awareness (Shinohara et al., 2017). Considering the age-related differences in human-automation interaction, it is not clear whether similar relationships arise in older adult populations during Level 5 driving.

Suboptimal levels of cognitive functioning can also be assessed via psychophysiological measures of autonomic arousal (Lohani et al., 2019). Carsten et al. (2012) found that heart rate was lower during autonomous driving when compared to semi-automated and manual driving, providing further support for cognitive underload during periods of automation. Yet, manual driving is confounded by physical effort (e.g., moving the steering wheel, changing gears) and cardiac activity is likely to be modulated by motor demands (Laborde et al., 2017). Similar to studies measuring eye gaze, research into physiological indices have indicated that safety-critical events impact functional state. For example, Zheng et al. (2015) found that masseter electromyography increased, and self-reported comfort decreased, as the headway between the lead vehicle decreased. During unexpected takeover requests and misleading notifications, Ruscio et al. (2017) demonstrated an increase in sympathetic arousal measured by increased skin-conductance response amplitude. As increases in arousal have been linked to attention narrowing (e.g., Laumann et al., 2003), these results suggest that the breadth of attentional focus is limited during safety-critical events (Meinlschmidt et al., 2019). However, Ruscio et al. (2017) employed semi-automated driving with takeover requests. Therefore, participants were anticipating a takeover. This is distinct to Level 5 AVs where participants will not anticipate having any direct control of the vehicle.

Considering the potential AV benefits for older adults, such as maintaining mobility and independence, and the age-related individual differences related to human-automation collaboration, a comprehensive understanding of older adults' psychophysiological state during periods of automated driving, particularly during safety-critical situations, is needed. Typically, research employs comparisons of autonomous driving to manual driving, but does not consider the distinct physical and cognitive demands. To this end, the aim of the present study was to investigate visual attention and autonomic arousal responses of older adults to a safety-critical event during a Wizard of Oz real-world autonomous journey. Participants experienced two types of stops: (i) one journey with the vehicle executing an unexpected stop due to detection of a 'hazard' (considered the safety-critical event) and, (ii) a different journey with an expected stop due to route set up in a repeated measures design with participants acting as their own controls. We monitored visual attention via fixation duration and fixation count, as well as physiological indices of electrodermal activity and heart rate. We also collected retrospective self-reported trust and reliability ratings in addition to summary qualitative feedback. We predicted the unexpected event would narrow the focus of overt visual attention coupled with an increase in autonomic arousal. This study formed part of the FLOURISH AV research project[1] funded by Innovate UK, which studied older adults' perceptions and interactions with AVs, including the development of an HMI, through co-design in a series of simulator and real-life studies.

MATERIALS AND METHODS

Participants

Thirty-nine adults originally participated in this study. Two participants were excluded from all analyzes due to the AV experiencing technical errors during the journeys, leaving 37 participants (16 females, 21 males, mean age $\pm$ SD = 68.35 $\pm$ 8.49 years, range 48–89 years, two participants under 60 years). Due to recording errors during data collection, only 30 participants' physiological data were subsequently analyzed (12 females, 18 males, mean age $\pm$ SD = 69 $\pm$ 8.75 years, range 48–89 years). Due to vision complications such as cataract (three), technical errors including unsuccessful calibration of the eye tracker (five), and low gaze samples (three), only 26 participants' eye tracking data were subsequently analyzed (12 females, 16 males, mean age $\pm$ SD = 67.19 $\pm$ 7.32 years, range 52–89 years).

Five participants had corrected hearing. 17 participants were educated to degree level and 10 participants were working full- or part-time. All but three participants held a valid driving license, driving, and on average drove 2,500–4,900 miles a year. No participants had any previous experience with highly automated driving. Those with significant health conditions (e.g., epilepsy, neurological impairments, and coronary issues) were

[1] http://www.flourishmobility.com

not permitted to take part. Participants received a £20 voucher as compensation for their participation to cover expenses. All participants gave written informed consent in accordance with the Declaration of Helsinki and were fully debriefed at the end of the study. Ethical approval was obtained by the Faculty of Health and Applied Sciences University of the West of England Research Ethics Committee (HAS.18.09.024).

Apparatus

Autonomous Vehicle

A Pod Zero autonomous pod provided by Aurrigo (RDM Group) was used as the AV (see **Figure 1**). The Pod is a compact research and development vehicle designed to be used in pedestrian areas and shared pedestrian/vehicle routes. It is electrically driven and can be used continuously for a period of 10+ hours of normal operation. It is a four-seater vehicle, with two benches facing each other designed similarly to a four-seater in a train. Due to safety regulations, a safety person was always present in the vehicle observing the environment and had access to an emergency stop button. Four marshals supervised the front and back of the vehicle, and the route was supervised by additional marshals at each intersection to ensure no vehicles or pedestrians caused an obstruction.

The autonomous behavior of the Pod was achieved using the Wizard of Oz approach (Kelley, 1985), whereby the Pod was remotely teleoperated in manual mode using a hand-held wireless control unit by an operator positioned behind the vehicle not in view of the participant. Driving the Pod in the teleoperated mode ensured that its actions were replicable between participants; the Pod could be made to respond similarly to different obstacles and followed the route as planned. At the beginning of the study, participants were told the vehicle was run fully autonomously. During debriefing, participants were told the Pod was operated manually by a teleoperator walking behind and remotely controlling it during the study. As the driving route involved a pedestrian area, the vehicle was controlled at walking speed, approximately 3–5 mph.

Human–Machine Interface (HMI)

The human–machine interface (HMI) was presented on a HannsG HT161HNB 15.6″ Multi Touch Screen connected to a Kodlix GN41 Mini PC (Windows 10, Intel Celeron processor, 8 GB RAM, 64 GB). The design of the HMI was informed by HMI design principles, public engagement workshops with older adults, and feedback from previous iterations of the HMI (Morgan et al., 2018; Eimontaite et al., 2019; Voinescu et al., 2020).

The HMI graphical touch screen displayed the vehicle speed, time remaining until destination, a safe stop button, a journey map, vehicle 'health,' and journey set up/change options (see **Figure 1**). The functionality of the safe stop button was described to the participant at the beginning of the study, emphasizing that pressing this icon would initiate the vehicle to stop. The vehicle 'health' icon provided information about the current working order of the automated system including the tires, brakes, network, and battery level. During the study, the vehicle health was always shown as being in good working order. The HMI presented visual and audio notifications to describe the vehicle's behavior and journey course, such as "*Turning left*" and "*You have arrived at your destination.*"

Journeys

Participants in the study went on six consecutive counterbalanced journeys. Before each journey, participants were provided with a scenario that specified the journey they were required to set up. There were six possible destinations/stops in total: Home, Health Center, Recycling Center, Sports Center, Sports Field, and Post Office. Among the six journeys there was always a journey including an expected stop, and another journey including an unexpected stop due to the 'hazard.' Both journeys were of an equivalent length and lasted for approximately 6 min. Some of the other journeys also included other variables such as picking up a friend. As the main focus of the current paper is to investigate the impact of an unexpected event, other journeys will not be described in detail. All journeys were randomized between participants so that the unexpected stop happened either during journey two or journey five, and the expected stop happened during journey one, journey three, or journey four. Overall, participants experienced approximately 60 min of the automated driving system.

The expected stop was initiated during journey set up and was therefore expected by the participants. A few seconds before

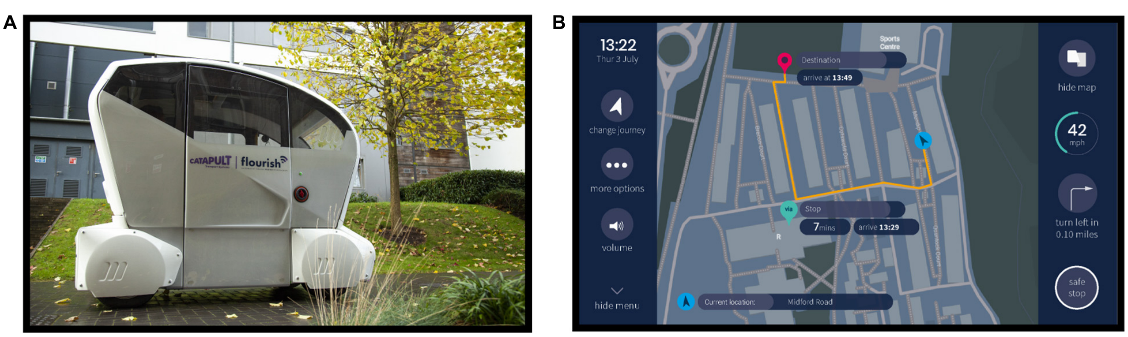


FIGURE 1 | Autonomous vehicle and human–machine interface. **(A)** Autonomous Pod utilized during the study. **(B)** Human–machine interface display during the journey.

the vehicle stopped, an HMI notification "*You are arriving at [Stop]*" was shown. Once the vehicle stopped, a notification "*You have arrived at [Stop]*" was shown. The HMI then displayed an option to either resume or stop the journey. All participants resumed the journey.

The unexpected stop was executed as an emergency stop appearing to the participants as happening suddenly, and as such, was not anticipated by the participants. A marshal was instructed to answer their mobile phone and walk in front of the Pod. The teleoperator of the Pod would then initiate the vehicle to stop. The HMI notification "*The vehicle detected a hazard in the road and has stopped. Your journey will resume shortly*" was presented on the touch screen. Once the marshal had moved safely out the way, the Pod would restart and continue the journey. The participant was not required to do anything.

Procedure

Figure 2 shows a schematic of the experimental procedure. Participants arrived and met the researcher near the student accommodation area on the university campus, where the study took place. Participants were reminded of the content of the information sheet, asked about their well-being, and whether they had any concerns or questions. They were told that the study involved setting up a designated route on the HMI before experiencing AV journeys around the student accommodation area, for a total of six journeys. They signed printed copies of the consent form and filled in the paper pre-trial questionnaires. Next, they were shown images of the HMI and described the overall layout. Once the physiological and eye tracking equipment were set up, participants were taken outside and introduced to the Pod. Participants sat inside the vehicle wearing a seatbelt and facing forward. They were introduced to the safety driver but were advised not to converse with them. Likewise, the safety driver was told not to converse with the participant. Inside the Pod, participants were shown the HMI. At the beginning of each journey, the participant received the journey scenario that specified the journey destination and stop if there was one. Participants were required to set up the journey using the HMI. During the first journey, the researcher assisted them with setting up the journey and answering any questions they had. All participants successfully set up the journeys throughout the trial. Once the journey was set up, the vehicle started, and the journey began. Participants were told they could interact with the HMI as little or as much as they wished to. During the journey, the HMI would present notifications describing the journey process, such as *"Turning left."* The HMI also

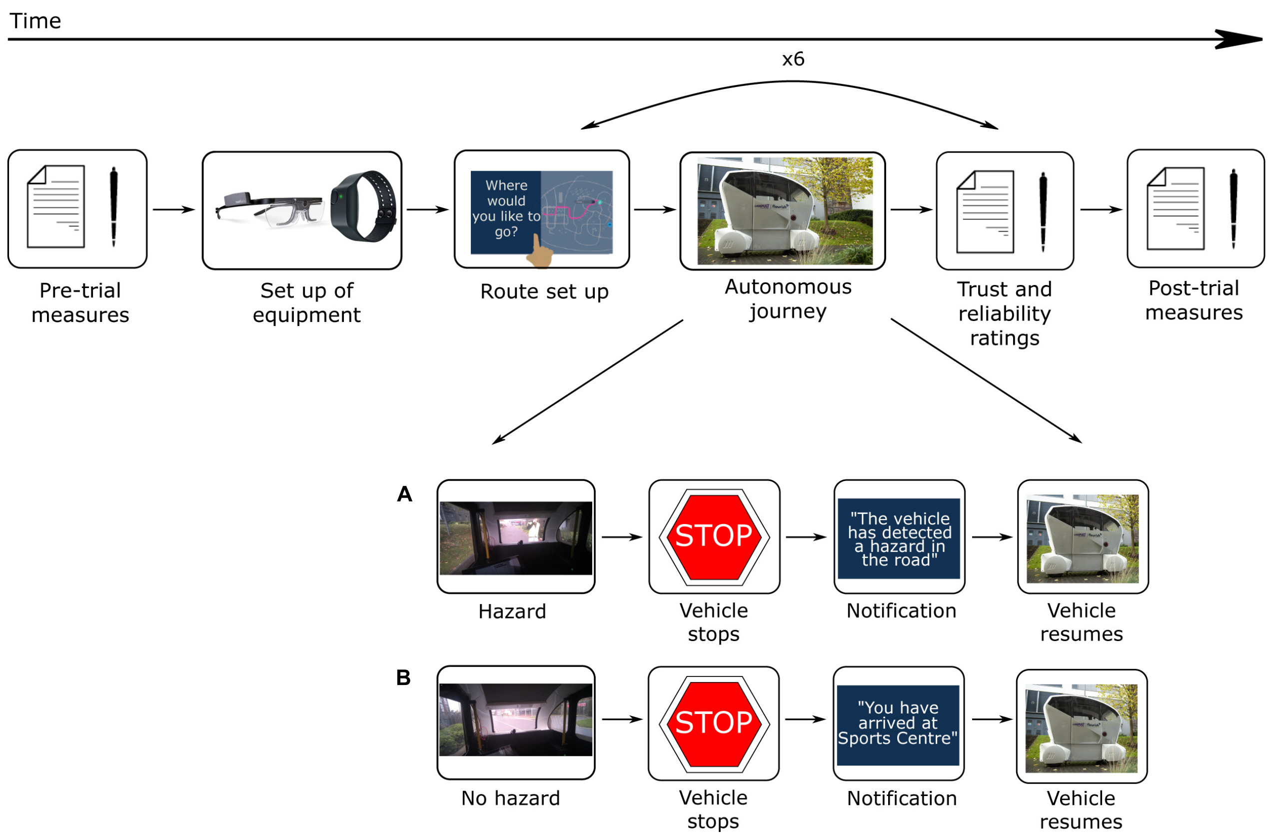


FIGURE 2 | Experimental procedure. **(A)** During the unexpected journey event, a marshal walked in front of the vehicle. The vehicle stopped and displayed a message, *"The vehicle has detected a hazard in the road. The vehicle will resume shortly."* The vehicle resumed once the roadway was clear. **(B)** During the expected journey event, the vehicle came to a stop when it reached a destination. A message displayed, such as, *"You have arrived at Sports Center."* The participant was required to press *"Resume"* on the HMI for the vehicle to resume the journey.

displayed a navigation map showing the vehicle route (see **Figure 1B**). After each journey, participants provided verbal trust and reliability ratings to a researcher, including a reason for their rating. This process was repeated six times and all participants completed six journeys. Afterward, participants left the vehicle and filled in several post-trial questionnaires. The full testing session, including the induction and filling out questionnaires, lasted for approximately 150 min, depending on inter-individual variability. We found that a significant amount of time was required and needed to be scheduled when conducting studies with older participants. It was important to ensure a pace that did not increase fatigue, and enough time to reflect and discuss issues raised and answer questions.

Measures

Trust and Reliability Ratings

We measured trust and reliability with a single-item scale to limit interruptions to the AV journeys. Participants were asked to rate how much they trusted the AV on a scale from 0 "Did not trust" to 10 "Completely trust." They were also asked how reliable the vehicle was on a scale from 0 "Not reliable" to 10 "Completely reliable". They were then asked to provide a reason for their rating. Ratings were taken verbally from participants at the end of every journey. The rating was also taken during the pre- and post-trial questionnaire phase, where participants were asked their current trust and reliability ratings of AVs on a paper questionnaire.

Physiological Signals

Continuous physiological acquisition of heart rate (beats per minute; BPM) and electrodermal activity (skin conductance level; μS) were collected using an Empatica E4 wristband (Empatica Inc., Cambridge, MA, United States and Milan, Italy) to measure levels of autonomic arousal. The sampling frequency for the electrodermal activity sensor was 4 Hz and the photoplethysmography sensor on the Empatica measured blood volume pulse at 64 Hz. The internal Empatica software derived the BPM. The Empatica E4 wristband was placed on participants' non-dominant wrist to reduce the possibility of motion artifacts. The Empatica was fastened tightly as comfortable for the participant, so the wristband did not move around inducing artifacts. The E4 also collected acceleration data from a 3-axis accelerometer, which enabled monitoring of wrist movements. The sampling frequency of the accelerometer was 32 Hz.

An event marking button on the Empatica E4 was pressed in front of a camera, which triggered a LED light to be illuminated on the Empatica, and simultaneously logged a timestamp in the data. This mode of creating a marker was done to aid the later analysis of when events of interest (i.e., the unexpected stop) occurred in the physiological data.

Eye Tracking

Tobii Pro Glasses 2, an eye tracking device, was used to collect fixation metrics (Tobii Glasses Eye Tracker, Tobii Technology, Stockholm, Sweden). The Tobii Glasses are a wearable eye tracker worn like a pair of glasses. The design is lightweight and has no side or bottom frame, preventing any distraction in the participant's visual field. The head unit is comprised of several cameras: a high-definition camera captured the participant's field of view (82° horizontal and 52° vertical), and two eye tracking sensors below each eye captured participants' pupil diameter and movements. To improve the accuracy of the eye tracking sensors, near-infrared lights illuminated the pupil. The sensors have a sampling rate of 100 Hz.

The Tobii Pro Glasses do not work with standard eyeglasses, as glasses can create additional glint that can lead to data corruption. Individuals wearing glasses were asked to remove them, and a suitable prescription lens from a set supplied as part of the Tobii kit was attached to the glasses. Once the participant was wearing the head unit, the manufacturer's calibration procedure was followed which consisted of the participant fixating on a central target. This process typically took less than 30 s. In addition, participants were asked to view specific objects in the environment so that the accuracy and alignment of the system could be checked.

Pre-processing

Physiological Arousal

Data were opened and pre-processed in Microsoft Excel 2016 using Excel's in-built functions. Electrodermal activity and heart rate values, with corresponding timestamps, were pre-processed separately and followed the same procedure. For electrodermal activity (4 Hz sampling rate), every four samples were averaged to produce one value for every second, and similarly, 1 s averages were used to analyze heart rate data. The averaged data were aligned to the appropriate time point, to allow for averaging across time points of interest. Time points of interest were derived from timestamps in a video recording and the Empatica event marker. Z-scores were calculated to standardize the data due to the individual variability of physiological responses (Braithwaite et al., 2013) resulting in z-transformed skin conductance level (zSCL) and z-transformed heart rate (zHR). For data relating to the unexpected and expected stop, data were averaged within two times of interest: 30 s before the stop, and 30 s after the stop. 30 s was chosen as this is a standard epoch length used in vigilance and psychophysiological state monitoring research (e.g., Berry et al., 2015) and in other AV research investigating changes in physiology in response to events (Ruscio et al., 2017). This was also the minimum duration of recorded activity after the specific events that was not affected by other events such as the end of the journey.

Fixation Metrics

Eye tracking analysis was undertaken using Tobii Pro Lab software version 1.138 (Tobii Technology, Stockholm, Sweden). We first assessed the gaze sample percentage across the entire recording. The eye tracking glasses captured a mean of 80% (SD = 18%) of gaze samples.

Events were first logged to indicate the start and end of events in the recording. Times of Interests (TOIs) were defined by selecting the appropriate start and end event markers. This allowed for segmentation of the data into intervals of time relevant to subsequent data analysis. The 'Pre-stop' TOI was considered the 30 s before the presentation of the notification when the vehicle stopped; the 'During' TOI consisted of the time the notification was displayed visually (up to approximately 15 s); the 'Post-stop' TOI was considered the 30 s after the presentation of the notification. Gaze data from the recording were then manually mapped onto an image best depicting the overall visual view of the participant.

Next, Areas of Interests (AOIs) were defined on each mapped image for each TOI (see **Figure 3**). Three AOIs were created representing the HMI, the central view of the driving environment, and the peripheral view of the driving environment. To finish, the 'I-VT Filter (Fixation)' was applied to the data, which set the velocity threshold parameter at 30 degrees/second. If the sample was below this threshold, it was classified as a fixation.

Total fixation duration and total fixation count metrics were exported from Tobii Pro Lab to Microsoft Excel 2016. Because the time of the critical event varied across participants, and to enable standardized comparison which took into account variability within patterns of fixations, it was necessary to calculate fixation count and fixation duration as a proportion of the total number of fixations and fixation durations. Fixation duration was defined as the amount of time spent looking at each AOI divided by the total duration of fixations. Fixation count was defined as the number of fixations toward each AOI divided by the total number of fixations. Averages were then calculated for subsequent analyzes.

RESULTS

All statistical analyzes were performed using IBM SPSS Statistics for Windows, version 26 (IBM Corp., Armonk, NY, United States). Descriptive statistics were performed, and normality was verified using the Shapiro–Wilk test and visualization of QQ plots of the unstandardized residuals. Assumptions of sphericity were tested using Mauchely's test and, if violated, Greenhouse–Geisser estimates were used in the repeated measures calculations. The statistical threshold for significance was set to two-tailed $p < 0.05$. Effect size was reported as eta squared (η^2) for one-way ANOVA significant results and partial eta squared (η_p^2) for two-way ANOVA significant results (Cohen, 1988). *Post hoc* analyzes were run with Bonferroni correction.

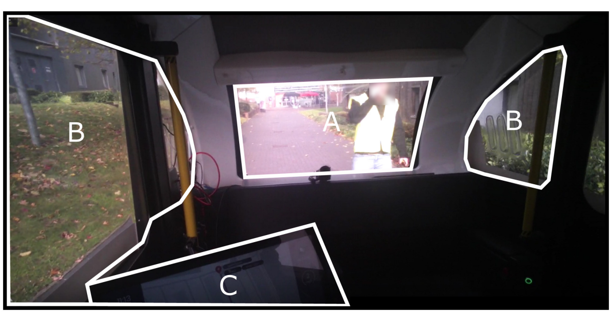


FIGURE 3 | Areas of Interest (AOI) for eye tracking analysis. **(A)** Central environment. **(B)** Peripheral environment. **(C)** Human–machine interface.

For trust and reliability ratings, a one-way repeated measures ANOVA (Journey: pre, unexpected, expected, and post) was undertaken. A 2 (Stop: unexpected and expected) × 2 (TOI: 30 s before and 30 s after) repeated measures ANOVA was performed to understand the impact of an expected and unexpected stop on heart rate and skin conductance level *z* scores. Two two-way repeated measures ANOVA were undertaken on both fixation count and fixation duration measures. The first was a 2 (Stop: unexpected and expected) × 3 (AOI: central, peripheral, and HMI) repeated measures ANOVA to understand the impact of journey type on AOI. The second ANOVA was a 2 (Stop: unexpected and expected) × 3 (TOI: pre-stop, during, and post-stop) repeated measures ANOVA to understand the impact of journey type on time.

Trust and Reliability Ratings

The descriptive statistics are displayed in **Table 1**. For trust ratings, a significant repeated measures ANOVA [$F_{(2.05,73.86)} = 15.05$, $p < 0.001$, $\eta^2 = 0.295$] with *post hoc* comparisons revealed that trust increased significantly from pre-all journeys to post-all journeys [$p < 0.001$], from pre-all journeys to the unexpected stop [$p < 0.001$], and from pre-all journeys to the expected stop [$p < 0.001$]. There was no significant difference in trust ratings between the unexpected and expected stop [$p = 0.100$].

For reliability, the one-way repeated measures ANOVA model showed that the main effect for journey was significant [$F_{(1.44,51.79)} = 25.56$, $p < 0.001$, $\eta^2 = 0.415$] and *post hoc* comparisons revealed that reliability ratings increased from pre-all journeys to post-all journeys [$p < 0.001$], from pre-all journeys to the unexpected stop [$p < 0.001$], and from pre-all journeys to the expected stop [$p < 0.001$]. Again, there was no significant difference in reliability ratings between the unexpected and expected stop [$p = 0.100$]. Overall, these findings indicate that subjective trust and reliability increased after AV experience and were not differentially impacted by the unexpected event.

Heart Rate

A 2 (stop: unexpected and expected) × 2 (time: pre-stop and post-stop) repeated measures ANOVA on zHR yielded no significant main effects of stop, time, or an interaction effect [$F_{(1,29)} \leq 0.56$, $p \geq 0.461$]. Heart rate was similar between the period before the expected stop [$M = -0.15$, $SD = 1.01$] and after the expected stop [$M = -0.12$, $SD = 0.93$]; and between the period before the unexpected stop [$M = -0.13$, $SD = 0.95$], and after the unexpected stop ($M = -0.06$, $SD = 0.92$). As illustrated in **Figure 4**, heart rate increased during the unexpected stop, but this did not reach statistical significance.

Skin Conductance Level

The ANOVA model revealed no significant main effects for stop [$F_{(1,29)} = 0.17$, $p = 0.684$] or time [$F_{(1,29)} = 0.37$, $p = 0.546$]. However, the interaction effect was significant [$F_{(1,29)} = 0.98$, $p = 0.019$, $\eta_p^2 = 0.176$].

Pairwise comparisons revealed that during the unexpected stop, zSCL was greater following the stop [$M = 0.16$, $SD = 0.83$] when compared to zSCL preceding the stop [$M = -0.05$,

TABLE 1 | Mean (*SD*) of trust and reliability ratings over journeys.

Subjective rating(0–10)	Journeys			
	Pre-	Post-	After an unexpected stop	After an expected stop
Trust	7.11 (2.50)	9.22 (1.13)	9.16 (1.43)	9.00 (1.78)
Reliability	7.19 (2.39)	10.00 (0.88)	9.35 (1.18)	9.45 (1.02)

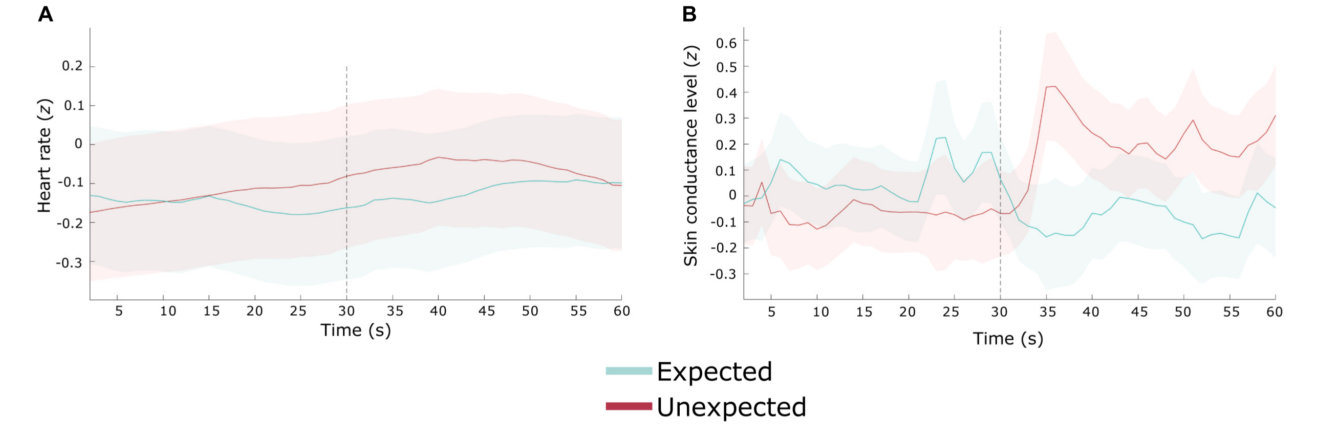


FIGURE 4 | Heart rate and skin conductance level during an unexpected (red) and expected (blue) stop. Gray dashed line represents the time point the vehicle stopped. Shaded areas represent ± the standard error of the mean difference. **(A)** Heart rate (*z*-transformed). **(B)** Skin conductance level (*z*-transformed).

SD = 0.72; $p = 0.043$]. There was no difference in zSCL before [$M = 0.06$, $SD = 0.96$] and after [$M = -0.09$, $SD = 0.89$] the expected stop.

In combination with the heart rate data, these results indicate that sympathetic arousal increased following vehicle cessation during the unexpected stop. See **Figure 4** for a depiction of the skin conductance level response.

Fixation Count

The 2 (stop: expected and unexpected) × 3 (AOI: central view, peripheral view, HMI) repeated measures ANOVA yielded a significant main effect of AOI [$F_{(1.42,35.55)} = 27.74$, $p < 0.001$, $\eta_p^2 = 0.526$], and a significant two-way interaction [$F_{(1.52,38.10)} = 28.47$, $p < 0.001$, $\eta_p^2 = 0.532$]. The main effect of the stop was not significant [$F_{(1,25)} = 0.44$, $p = 0.51$; **Table 2**].

Post hoc comparisons of the two-way interaction revealed a higher number of fixations on the central environment during an unexpected stop compared to an expected stop [$p < 0.001$]; whereas fixation count was greater on the HMI area during the expected stop, compared to the unexpected stop [$p < 0.001$].

Overall, during the unexpected stop, the number of fixations were higher on the HMI area compared to the peripheral environment [$p < 0.001$], and the central environment compared to the peripheral environment [$p < 0.001$]. During the expected stop, the number of fixations were higher on the HMI compared to the central environment [$p < 0.001$], and the HMI compared to the peripheral environment [$p < 0.001$]. In combination, these results reveal that the number of fixations within the central environment was higher during an unexpected stop, whereas the number of fixations within the HMI was higher during an expected stop.

The 2 (stop: unexpected and expected) × 3 (time: pre-stop, during, and post-stop) repeated measures ANOVA revealed no significant main effects of time [$F_{(1.56,39.03)} = 0.88$, $p = 0.399$] or stop [$F_{(1,25)} = 0.44$, $p = 0.513$], nor a significant interaction effect [$F_{(1.21,30.21)} = 1.45$, $p = 0.244$; **Table 3**]. See **Figures 5**, **6** for a depiction of the results.

Fixation Duration

The 2 (stop) × 3 (AOI) ANOVA model yielded a significant main effect for AOI [$F_{(1.59,39.72)} = 29.23$, $p < 0.001$, $\eta_p^2 = 0.539$], and a significant interaction effect [$F_{(1.58,39.37)} = 23.27$, $p < 0.001$, $\eta_p^2 = 0.482$]. The main effect for stop was not significant [$F_{(1,25)} = 0.44$, $p = 0.516$; **Table 2**].

Post hoc comparisons revealed that fixation duration on the HMI was longer during the expected stop compared to the unexpected stop [$p = 0.003$], but longer on the central environment during the unexpected stop compared to the expected stop [$p < 0.001$]. Fixation duration on the peripheral environment was marginally greater during the expected compared to the unexpected stop [$p = 0.055$]. Additionally, for the unexpected stop, fixation duration was shorter for the peripheral environment when compared to the HMI [$p < 0.001$] and the central environment [$p < 0.001$]. For the expected stop, fixation duration was longer on the HMI compared the peripheral environment [$p < 0.001$] and the central environment [$p < 0.001$].

The 2 (stop) × 3 (time) repeated measures ANOVA revealed a significant main effect of time [$F_{(1.63,40.82)} = 6.52$, $p = 0.006$,

TABLE 2 | Mean (*SD*) of fixation metrics count (%) and duration (%) across the human–machine interface (HMI), central environment, and peripheral environment during expected and unexpected stops.

Fixation metric	HMI		Central environment		Peripheral environment	
	Expected	Unexpected	Expected	Unexpected	Expected	Unexpected
Fixation count (%)	52.57(30.07)	30.32(28.23)	20.70(20.02)	46.50(26.35)	17.81(20.85)	12.52(14.76)
Fixation duration (%)	32.36(27.75)	18.86(23.06)	11.99(13.86)	30.03(22.55)	8.44(11.54)	5.07(5.97)

TABLE 3 | Mean (*SD*) of fixation metrics count (%) and duration (%) across pre-, during, and post- expected and unexpected autonomous journeys.

Fixation metric	Pre-		During		Post-	
	Expected	Unexpected	Expected	Unexpected	Expected	Unexpected
Fixation count (%)	31.94(25.91)	28.87(25.68)	30.17(42.32)	30.43(31.74)	28.97(20.38)	30.04(25.03)
Fixation duration (%)	17.36(16.87)	17.58(17.79)	19.34(31.02)	19.51(25.30)	16.08(13.95)	16.88(19.27)

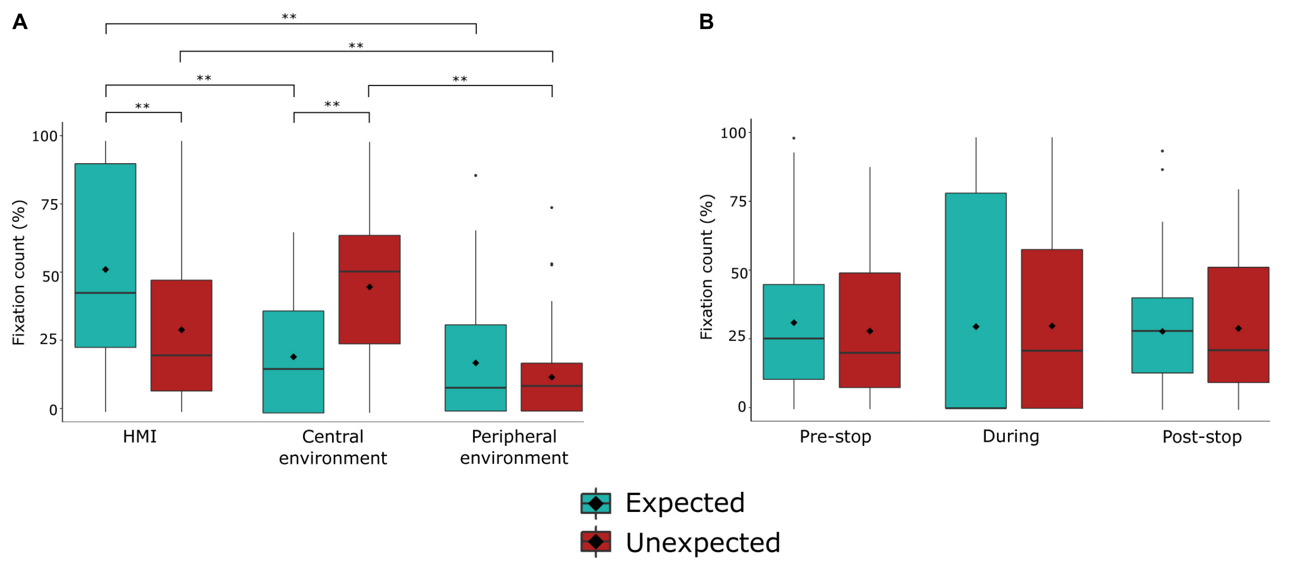


FIGURE 5 | Fixation count (%) during expected and unexpected events (stops). **(A)** Fixation count during unexpected and expected journey events over areas of interest. **(B)** Fixation count during unexpected and expected journey events over times of interest. Bolded line represents the median value. Box represents the interquartile range. Vertical lines represent the lower/upper adjacent values. ◇ represents the mean value. ***p* < 0.001, **p* < 0.05.

$\eta_p^2 = 0.207$]. The main effect for stop [$F_{(1,25)} = 0.44$, $p = 0.516$] and the interaction effect were not significant [$F(2,50) = 0.21$, $p = 0.813$]. Fixation duration was greater during the stop [$M = 19.96$, $SD = 8.42$] compared to after the stop [$M = 16.48$, $SD = 7.72$], regardless of whether it was an expected or unexpected stop [$p = 0.01$]. See **Table 3** for an overview of the means and standard deviations.

In combination, these results suggest that while similar visual demands were afforded to the scene, participants visual attention was distinctly allocated during the unexpected and expected stops. Fixation duration was longer on the central environment during an unexpected stop, whereas fixation duration was longer on the HMI during an expected stop, indicating distinct visual attention resource allocation between the different types of stop. The results are depicted in **Figures 7**, **8**.

DISCUSSION

This study sought to understand the impact of an unexpected event during Level 5 autonomous driving on gaze behavior, autonomic arousal, and associated trust levels. To accomplish this, an experiment was designed where the participants experienced what they thought were autonomous journeys that included two stops on separate journeys: one unexpected stop initiated by a 'hazard,' and one expected stop initiated as part of the planned journey set up. Elevated electrodermal activity persisted after the unexpected stop. Gaze fixation metrics revealed several visual behavior differences. Overall, participants searched the central environment, inclusive of the 'hazard,' for longer during the unexpected stop, whereas during the expected stop, the HMI area captured visual attention, as measured by greater fixation counts and

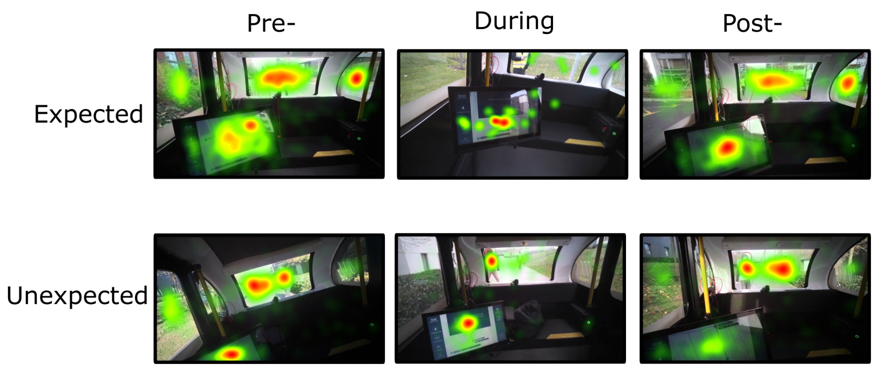


FIGURE 6 | Total fixation count during expected and unexpected journey events (stops). The heat map represents the summary of all gaze points in the visual environment over three time points of interest. Colors indicate the total gaze fixations (fixation count increases from green – yellow – orange – red).

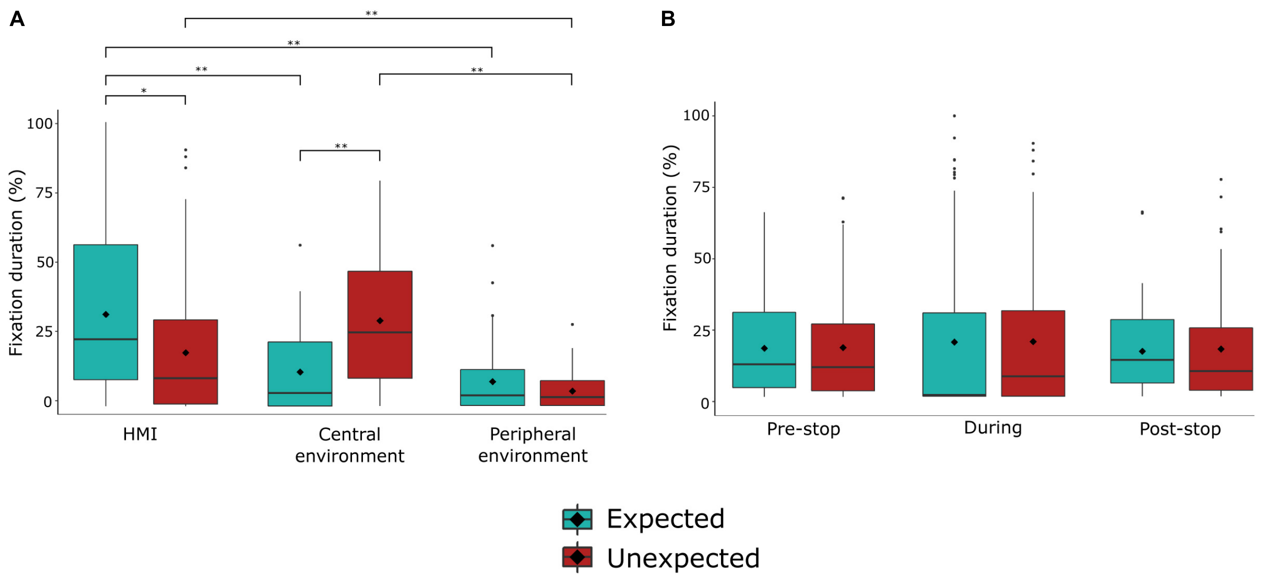


FIGURE 7 | Fixation duration (%) during unexpected and expected journey events (stops). **(A)** Fixation count during unexpected and expected journey events over areas of interest. **(B)** Fixation count during unexpected and expected journey events over times of interest. ◇ represents the mean value. $^{**}p < 0.001$, $^{*}p < 0.05$.

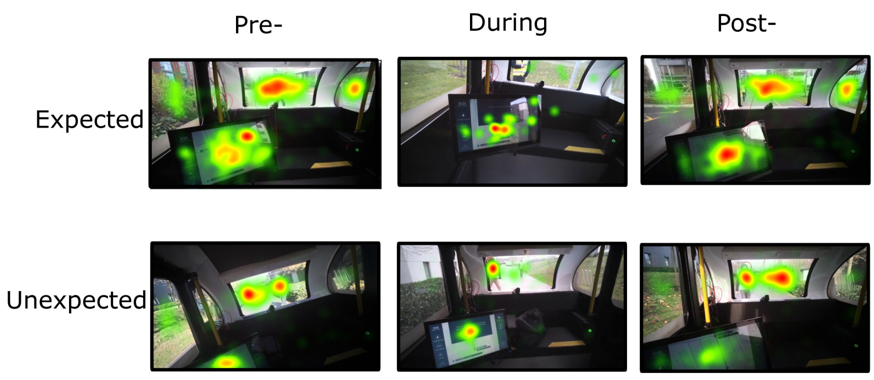


FIGURE 8 | Total fixation duration during expected and unexpected journey events (stops). The heat map represents the summary of all fixations in the visual environment over three time points of interest. Colors indicate the total gaze fixation duration (fixation duration increases from green – yellow – orange – red).

longer fixation durations. Trust and reliability ratings also increased from pre-journey values and remained high after each type of journey.

The distribution and duration of fixations measured with eye tracking outline the differences between the two types of journeys: while in the unexpected stop journey participants had longer fixation durations and greater fixation count toward the central environment (containing the 'hazard'), in the expected stop visual attention was directed toward the HMI. Visual behaviors between stops were similar over time, indicating similar demands on visual attention. General visual scanning behavior can be understood by fixation counts, as more shifts within a scene are associated with a greater frequency of fixations. Fixation duration can provide further insight by indicating visual attention demands. Basic visual processing research has demonstrated that fixation duration increases with visual scene complexity (Pomplun et al., 2013) and cognitive load (Rayner, 1998), and is linked to uncertainty (Brunyé and Gardony, 2017). As such, our patterns of results imply that during an unexpected stop, visual attention was directed toward the central environment containing the 'hazard,' with the participants searching for information, rather than focusing on other aspects of the scene. These results are similar to research studying ocular behavior during manual driving and hazardous situations. The variance of fixations decreased when presented with a critical situation (Chapman and Underwood, 1998). In contrast, fixation duration increased coming up to, and during, a critical situation (e.g., Chapman and Underwood, 1998; Underwood et al., 2005). In addition, a negative relationship between task demands during driving and visual scanning behavior has been demonstrated, i.e., higher task demands reduced the dispersion of visual scanning (e.g., Recarte and Nunes, 2003; Savage et al., 2013). Moreover, Guo et al. (2019) found that fixation frequency and duration increased during accident scenes reflecting increased anxiety.

Searching for information related to the unexpected event might be explained by increased anxiety (Guo et al., 2019): the narrowing of visual attention, focusing on the hazard, is a common feature of increased arousal and stress (Chajut and Algom, 2003; Gable and Harmon-Jones, 2010). Moreover, the physiological results show increased sympathetic arousal, following the unexpected stop. Skin conductance levels increased with vehicle braking due to the unexpected event. This increase persisted up to 30 s, yet there was no significant difference found during the expected stop journey. Driving studies have indicated that high skin conductance levels are modulated by various phenomena such as increased workload (e.g., Mehler et al., 2012), stress (e.g., Affanni et al., 2018), anxiety (Barnard and Chapman, 2018), and lower trust in automation (Morris et al., 2017; see Lohani et al., 2019 for a review). It is therefore difficult to infer specifically why skin conductance levels rose, other than reflecting an overall increase in sympathetic arousal. Trust ratings were high after all journeys, implying trust levels did not modulate sympathetic arousal. However, response bias, particularly following verbal ratings, may have led to an overestimation of self-reported trust. It should be noted that trust was measured retrospectively once the vehicle had successfully completed the journey. Factors such as trust, workload, and anxiety are time-varying, and as such, participants may have experienced lower trust levels during the journey, represented by heightened skin conductance. However, as the vehicle behaved appropriately to the unexpected event (e.g., braking and notifying the participant), and the journey completed successfully, this may have encouraged participants to rate their trust of the vehicle's behavior positively at the end of the journey (Choi and Ji, 2015).

We did not find any statistically significant difference in heart rate although **Figure 4** shows elevated heart rate, similar to skin conductance, for the unexpected stop compared to the expected stop. As skin conductance is regulated by the sympathetic nervous system, and heart rate is modulated by both the activation and suppression of sympathetic and parasympathetic branches of the autonomic nervous system, respectively (Thayer et al., 2010), our results suggest that the vehicle stopping in response to a unexpected event might reflect a mild sympathetic dominance. Ruscio et al. (2017) measured physiological responses to takeover requests following various warnings. During semi-autonomous driving, heart rate decreased relative to manual driving following reliable warnings, misleading warnings, and no warnings. Skin conductance response amplitude increased during misleading warnings and no warnings. They also found that respiratory sinus arrhythmia, an index of parasympathetic activity, increased from manual driving to an unexpected takeover with no warning. Their results reveal an imbalance between the parasympathetic and sympathetic branches during takeovers preceded by a misleading or no warning. The authors suggest that this discrepancy may reduce attentional capacity, resulting in cognitive overload. Although we did not measure specific or non-specific response amplitude changes, but rather changes in skin conductance level, our results are compatible as we found a similar moderate effect of increased skin conductance level following vehicle cessation without any warning (unexpected stop). However, our results are difficult to directly compare to Ruscio et al. (2017) findings as we did not measure physiological responses during manual driving or initiate a takeover request. We also did not separate parasympathetic activity from sympathetic activity; therefore, it is not clear whether a reduction in attentional capacity was associated with an increase in sympathetic activation as measured by an increase in skin conductance level.

A potential limitation in our study was the use of the Empatica E4 for assessing autonomic arousal. Gruden et al. (2019) recently found that manual driving-related movement artifacts impacted heart rate variability and skin conductance level measurements. Reasonable accuracy and reliability have been reported for this device providing wrist movements are low (Pietilä et al., 2017; Ragot et al., 2017), which was the case during our study, as Level 5 driving does not require behaviors such as changing gears. Nevertheless, we conducted additional analyzes and confirmed that accelerometer values did not differ between conditions (included in the **Supplementary Material**). In addition, conventional physiological research measures from the distal or intermediate phalanges of the ring and index fingers where there are a larger number of active eccrine sweat glands (Freedman et al., 1994; Boucsein, 2012). The E4, like many wearables, measures skin conductance via wrist sensors. As the wrist is less responsive to skin conductance, an underestimation of parameters is expected (van Dooren and Janssen, 2012; Payne et al., 2016). Despite this, the Empatica E4 was a relatively unobtrusive measurement device and was sensitive to changes in skin conductance level.

Despite the potential benefits of measuring sympathetic arousal and ocular behavior during Level 5 driving, it is not possible to avoid limitations inherent to skin conductance and eye tracking measurements. Due to a one- to four- second delay, or response latency, following a stimulus presentation (Boucsein, 2012), skin conductance measurements should not be used to detect time-critical events and are therefore not a usable metric on their own for a DSM system. In addition, it should be acknowledged that the skin conductance level values we measured were contaminated by skin conductance responses. If skin conductance responses were triggered by events during the journey, this would increase the underlying skin conductance level. Therefore, the values we report are impacted by both tonic and phasic responses to the events. Furthermore, it is well acknowledged that fixations cannot occur without attention, but attention can occur without fixations (Posner, 1980). Eye tracking is unable to detect the periphery of a participant's visual gaze, but stimuli can be perceived pre-attentively in peripheral vision. Participants may have therefore discerned the notification and inhibited saccadic movement for further processing. Caution is therefore required in directly attributing changes in indirect measures, such as visual attention assessed with eye tracking, to direct measures of central attention. A robust DSM system may consequently benefit from including a variety of measures. The results obtained here do show significant differences in visual gaze behavior, perhaps reflecting changes in visual strategy as a result of reallocation of attentional resources relating to the unexpected event.

Although the current study attempted to produce increased ecological validity compared to laboratory studies, safety restrictions were put in place including the speed of the vehicle, the safety driver, and the marshals surrounding the vehicle. On average, the vehicle went between 3 and 5 mph. The speed of a vehicle has been shown to correlate with self-reported workload measures, i.e., the greater the speed, the greater self-reported workload (Fuller, 2005). However, research has found that this depends on the situation complexity. Low-complexity environments including motorways at faster speeds, or high-complexity situations including town centers at lower speeds, may modulate load in a similar manner (Paxion et al., 2014). In our study, the vehicle drove around a pedestrianized area, where the speed limit was 10mph. The vehicle shared the lane with pedestrians, cyclists, and obstacles such as bollards. Therefore, driving at a greater speed would not have been possible nor realistic or safe, even during manual driving.

Finally, the results imply that the unexpected event placed significant demands on attentional resources. However, eye tracking is an indirect measure of attention, and as the study mimicked Level 5 autonomous journeys, no direct performance measure could be derived to support this view. Yet, all participants were introduced to a "safe stop" button on the HMI, which could be pressed at any time if they wanted the vehicle to stop. None of the participants activated the safe stop. They could also have accessed the "vehicle health" icon, providing information about the overall health of the vehicle. None of the participants accessed this icon during the times of interest. Taken together, these findings suggest that the unexpected stop was not perceived as particularly dangerous as neither subjective reports nor subjective ratings, or all physiological indices reflected an extreme response that may be associated with more imminent or extreme danger. This is supported by summary qualitative analysis where 12 of the participants, representing around a third of the sample, expressed unease (e.g., "nervous it would not stop, would like a horn") when discussing the unexpected stop with the researchers after journey completion. A further 3 expressed ease or confidence (e.g., "had stopped before, would this time") in relation to the event, with the remainder simply noting that the vehicle had spotted the 'hazard' and stopped, performing its intended functions. In addition, as our study only included one unexpected event, further investigations using different types of unexpected events are needed to be able to characterize functional states to specific safety-critical scenarios.

Although our results are supported by the above-mentioned driving studies, the study presented in this paper varies considerably as our participants were not active manual drivers in control of the vehicle. As ocular behavior and motor execution are intrinsically linked both spatially and temporally, active drivers successfully fixate directly at the objects being interacted with or ones that precede the action. Despite these differences, our results are in agreement with Strauch et al. (2019) who investigated eye gaze of passengers during real-world autonomous driving. They found a greater frequency of fixations on safety-relevant AOIs when joining a highway during an autonomous journey when compared to manual driving and to the rest of the route. Visual scanning behavior was therefore affected by safety-critical situations regardless of active involvement in the driving task. Strauch et al. (2019) participants' mean age was 23 years, and so our results extend this earlier research and suggest that older adults display broadly similar ocular behaviors to younger adults during safety-critical situations. However, our study does differ, as passengers could interact with an HMI throughout the journeys. Unlike a passenger in a manually driven car, in a Level 5 vehicle the monitoring of the automated system can take more significance, given the reduced need to pay attention to the road ahead. Despite this, we found that attentional focus narrowed toward the 'hazard' before, during, and after the critical event, which was also accompanied by an increase in skin conductance reflecting increased sympathetic nervous system arousal following the vehicle response.

Repeated stress response activation and consequential negative emotions may have a significant impact on overall health and well-being. Therefore, a DSM system could react in response to detecting increased arousal by either modifying the vehicle's behavior, alerting the driver, or modifying HMI notifications and providing status updates or information about future events. For example, the HMI might adapt safety-related notifications to make them more engaging, multimodal, and alerting, depending on individual characteristics and the attentional level of the passenger (ranging from inattentive to over-alert). Moreover, the vehicle could learn the types of situations that have a negative impact on passenger well-being and adapt vehicle route or driving style to avoid them.

Taken together, these results have several critical implications for the safe implementation of Level 5 AVs for older adults. Our results reveal possible narrowing of visual attention and heightened arousal during an unexpected event as demonstrated by increased sympathetic arousal and a smaller distribution of fixations, coupled with an increase in fixations toward the unexpected event. In combination with consistently high trust ratings, these results suggest that the passive process of automated driving may restrict the focus of visual attention and heighten adverse responses. This study also demonstrates that the physiological indices examined can be useful and practical measures for evaluating passengers' functional state during real-world autonomous driving. As such, a DSM system that includes these measures might be able to detect these behaviors and make an informed decision on vehicle behavior and adapt HMI notifications accordingly. The potential for negative experiences during Level 5 driving, coupled with human limitations in sustained monitoring during low and high arousal situations, suggests that a DSM system may be a necessary adjunct to fully AVs in supporting potentially vulnerable people in unexpected situations.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The study was reviewed and approved by the Faculty of Health and Applied Sciences University of the West of England Research Ethics Committee. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AS, IE, JD, and TK collected and processed the experimental data. IE was responsible for running the study. CA, PC-S, and PM were involved in the planning and design of the experiments. CA and PC-S were involved in the implementation and supervision of the work. AS performed the analysis, drafted the manuscript, and designed the figures. CA, IE, and PC-S aided in interpreting the results and edited the manuscript. All authors commented on the manuscript.

FUNDING

This research was supported by Innovate UK and was undertaken as part of the Flourish connected and autonomous vehicle research project (#62190-456195). The authors declare that AS received funding as a doctoral research candidate provided by UWE RBI and Airbus Groups Innovations (HAS-HSS-16009). Airbus Group Innovations were not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

ACKNOWLEDGMENTS

The authors would like to thank: Aurrigo/RDM (Coventry, United Kingdom) for providing the autonomous Pod Zero, Dave Barnett (Jolly Good Engineering Ltd., Leicestershire, United Kingdom) for his work on the implementation of the HMI used in this research, Jess Ridgers, Hazel Boyd, and Keir Haines, Designability (Bath, United Kingdom) as project partners for their contribution to the design of the HMI, and the Transport Systems Catapult (Milton Keynes, United Kingdom) as project partners for contributing to the development of the HMI software and providing support for the trial.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.571961/full#supplementary-material

REFERENCES

Affanni, A., Bernardini, R., Piras, A., Rinaldo, R., and Zontone, P. (2018). Driver's stress detection using skin potential response signals. *Measurement* 122, 264–274. doi: 10.1016/j.measurement.2018.03.040

Anstey, K. J., Windsor, T. D., Luszcz, M. A., and Andrews, G. R. (2006). Predicting driving cessation over 5 years in older adults: psychological well-being and cognitive competence are stronger predictors than physical health. *J. Am. Geriatr. Soc.* 54, 121–126. doi: 10.1111/j.1532-5415.2005.00471.x

Arexis, M., Maquestiaux, F., Gaspelin, N., Ruthruff, E., and Didierjean, A. (2017). Attentional capture in driving displays. *Br. J. Psychol.* 108, 259–275. doi: 10.1111/bjop.12197

Barnard, M. P., and Chapman, P. (2018). The effects of instruction and environmental demand on state anxiety, driving performance and autonomic activity: are ego-threatening manipulations effective? *Transp. Res. Part F Traffic Psychol. Behav.* 55, 123–135. doi: 10.1016/j.trf.2018.02.04

Bergstad, C. J., Gamble, A., Gärling, T., Hagman, O., Polk, M., Ettema, D., et al. (2011). Subjective well-being related to satisfaction with daily travel. *Transportation* 38, 1–15. doi: 10.1007/s11116-010-9283-z

Berry, R. B., Brooks, R., Gamaldo, C. E., Harding, S. M., Lloyd, R. M., Marcus, C. L., et al. (2015). *For the American Academy of Sleep Medicine. The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. Darien, IL: American Academy of Sleep Medicine.

Boucsein, W. (2012). *Electrodermal Activity*, 2nd Edn. New York: Springer.

Braithwaite, J. J., Watson, D. G., Jones, R., and Rowe, M. (2013). A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology* 49, 1017–1034.

Broome, K., McKenna, K., Fleming, J., and Worrall, L. (2009). Bus use and older people: a literature review applying the person–environment–occupation model in macro practice. *Scand. J. Occup. Ther.* 16, 3–12. doi: 10.1080/11038120802326222

Brunyé, T. T., and Gardony, A. L. (2017). Eye tracking measures of uncertainty during perceptual decision making. *Int. J. Psychophysiol.* 120, 60–68. doi: 10.1016/j.ijpsycho.2017.07.008

Carsten, O., Lai, F. C., Barnard, Y., Jamson, A. H., and Merat, N. (2012). Control task substitution in semiautomated driving: does it matter what aspects are automated? *Hum. Factors* 54, 747–761. doi: 10.1177/0018720812460246

Chajut, E., and Algom, D. (2003). Selective attention improves under stress: implications for theories of social cognition. *J. Pers. Soc. Psychol.* 85, 231–248. doi: 10.1037/0022-3514.85.2.2310018720812460246

Chapman, P. R., and Underwood, G. (1998). Visual search of driving situations: danger and experience. *Perception* 27, 951–964. doi: 10.1068/p270951

Choi, J. K., and Ji, Y. G. (2015). Investigating the importance of trust on adopting an autonomous vehicle. *Int. J. Hum. Comput. Interact.* 31, 692–702. doi: 10.1080/10447318.2015.1070549

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edn. Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Cohen, S., Janicki-Deverts, D., and Miller, G. E. (2007). Psychological stress and disease. *JAMA* 298, 1685–1687. doi: 10.1001/jama.298.14.1685

Collet, C., and Musicant, O. (2019). Associating vehicles automation with drivers functional state assessment systems: a challenge for road safety in the future. *Front. Hum. Neurosci.* 13:131. doi: 10.3389/fnhum.2019.00131

Comte, S. L. (2000). New systems: new behaviour? *Transp. Res. Part F Traffic Psychol. Behav.* 3, 95–111. doi: 10.1016/S1369-8478(00)00019-X

De Winter, J. C., Happee, R., Martens, M. H., and Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: a review of the empirical evidence. *Transp. Res. Part F Traffic Psychol. Behav.* 27, 196–217. doi: 10.1016/j.trf.2014.06.016

Dellinger, A. M., Sehgal, M., Sleet, D. A., and Barrett-Connor, E. (2001). Driving cessation: what older former drivers tell us. *J. Am. Geriatr. Soc.* 49, 431–435. doi: 10.1046/j.1532-5415.2001.49087.x

Eimontaite, I., Voinescu, A., Alford, C., Caleb-Solly, P., and Morgan, P. (2019). "The impact of different human-machine interface feedback modalities on older participants' user experience of CAVs in a simulator environment," in *Advances in Human Factors of Transportation*, ed. N. A. Stanton (Cham: Springer), 120–132. doi: 10.1007/978-3-030-20503-4_11

Eisenhandler, S. (1990). The asphalt identikit: old age and the driver's license. *Int. J. Aging Hum. Dev.* 30, 1–14. doi: 10.2190/0mf5-hq1l-7eby-xnxv

Faber, K., and van Lierop, D. (2020). How will older adults use automated vehicles? Assessing the role of AVs in overcoming perceived mobility barriers. *Transp. Res. Part A Policy Pract.* 133, 353–363. doi: 10.1016/j.tra.2020.01.022

Finkleman, J. M. (1994). A large database study of the factors associated with work-induced fatigue. *Hum. Factors* 36, 232–243. doi: 10.1177/001872089403600205

Freedman, L. W., Scerbo, A. S., Dawson, M. E., Raine, A., McClure, W. O., and Venables, P. H. (1994). The relationship of sweat gland count to electrodermal activity. *Psychophysiology* 31, 196–200. doi: 10.1111/j.1469-8986.1994.tb01040.x

Fuller, R. (2005). Towards a general theory of driver behaviour. *Accid. Anal. Prev.* 37, 461–472. doi: 10.1016/j.aap.2004.11.003

Gable, P., and Harmon-Jones, E. (2010). The blues broaden, but the nasty narrows: attentional consequences of negative affects low and high in motivational intensity. *Psychol. Sci.* 21, 211–215. doi: 10.1177/0956797609359622

Gruden, T., Stojmenova, K., Sodnik, J., and Jakus, G. (2019). Assessing drivers' physiological responses using consumer grade devices. *Appl. Sci.* 9, 5353. doi: 10.3390/app9245353

Guo, Y., Wang, X., Xu, Q., Liu, F., Liu, Y., and Xia, Y. (2019). Change-point analysis of eye movement characteristics for female drivers in anxiety. *Int. J. Environ. Res. Public Health* 16, 1236. doi: 10.3390/ijerph16071236

Helldin, T., Falkman, G., Riveiro, M., and Davidsson, S. (2013). "Presenting system uncertainty in automotive UIs for supporting trust calibration in autonomous driving," in *Proceedings of the 5th International Conference on Automotive user Interfaces and Interactive Vehicular Applications*, Eindhoven, 210–217. doi: 10.1145/2516540.2516554

Hergeth, S., Lorenz, L., Vilimek, R., and Krems, J. F. (2016). Keep your scanners peeled: gaze behavior as a measure of automation trust during highly automated driving. *Hum. Factors* 58, 509–519. doi: 10.1177/0018720815625744

Kelley, J. F. (1985). "CAL – a natural language program developed with the OZ paradigm: implications for supercomputing systems," in *Proceedings of the 1st International Conference on Supercomputing Systems*, St. Petersburg, FL, 238–248.

Körber, M., Baseler, E., and Bengler, K. (2018). Introduction matters: manipulating trust in automation and reliance in automated driving. *Appl. Ergon.* 66, 18–31. doi: 10.1016/j.apergo.2017.07.006

Kramer, A. F., and Madden, D. (2008). "Attention," in *The Handbook of Cognition and Ageing*, eds F. I. M. Craik and T. A. Salthouse (Hillsdale, NJ: Lawrence Erlbaum), 189–249.

Laborde, S., Mosley, E., and Thayer, J. F. (2017). Heart rate variability and cardiac vagal tone in psychophysiological research–recommendations for experiment planning, data analysis, and data reporting. *Front. Psychol.* 8:213. doi: 10.3389/fpsyg.2017.00213

Laumann, K., Gärling, T., and Stormark, K. M. (2003). Selective attention and heart rate responses to natural and urban environments. *J. Environ. Psychol.* 23, 125–134. doi: 10.1016/S0272-4944(02)00110-X

Lohani, M., Payne, B. R., and Strayer, D. L. (2019). A review of psychophysiological measures to assess cognitive states in real-world driving. *Front. Hum. Neurosci.* 13:57. doi: 10.3389/fnhum.2019.00057

Louw, T., and Merat, N. (2017). Are you in the loop? Using gaze dispersion to understand driver visual attention during vehicle automation. *Transp. Res. Part C Emerg. Technol.* 76, 35–50. doi: 10.1016/j.trc.2017.01.001

Mann, W. C., Belchior, P., Tomita, M. R., and Kemp, B. J. (2007). Older adults' perception and use of PDAs, home automation system, and home health monitoring system. *Top. Geriatr. Rehabil.* 23, 35–46. doi: 10.1097/00013614-200701000-00006

Marottoli, R. A., de Leon, C. F. M., Glass, T. A., Williams, C. S., Cooney, L. M. Jr., and Berkman, L. F. (2000). Consequences of driving cessation: decreased out-of-home activity levels. *J. Gerontol. Ser. B Psychol. Sci. Soc. Sci.* 55, S334–S340. doi: 10.1093/geronb/55.6.s334

McBride, S. E., Rogers, W. A., and Fisk, A. D. (2011). Understanding the effect of workload on automation use for younger and older adults. *Hum. Factors* 53, 672–686. doi: 10.1177/0018720811421909

Mehler, B., Reimer, B., and Coughlin, J. F. (2012). Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task: an on-road study across three age groups. *Hum. Factors* 54, 396–412. doi: 10.1177/0018720812442086

Meinlschmidt, G., Stalujanis, E., and Tegethoff, M. (2019). The psychobiology of using automated driving systems: a systematic review and integrative model. *Psychoneuroendocrinology* 105, 51–63. doi: 10.1016/j.psyneuen.2018.09.029

Morgan, P. L., Voinescu, A., Williams, C., Caleb-Solly, P., Alford, C., Shergold, I., et al. (2018). "An emerging framework to inform effective design of human-machine interfaces for older adults using connected autonomous vehicles," in *Advances in Human Aspects of Transportation*, ed. N. A. Stanton (Cham: Springer), 325–334. doi: 10.1007/978-3-319-60441-1_33

Morris, D. M., Erno, J. M., and Pilcher, J. J. (2017). "Electrodermal response and automation trust during simulated self-driving car use," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Los Angeles, CA, 1759–1762. doi: 10.1177/1541931213601921

Parasuraman, R., and Manzey, D. H. (2010). Complacency and bias in human use of automation: an attentional integration. *Hum. Factors* 52, 381–410. doi: 10.1177/0018720810376055

Paxion, J., Galy, E., and Berthelon, C. (2014). Mental workload and driving. *Front. Psychol.* 5:1344. doi: 10.3389/fpsyg.2014.01344

Payne, A. F., Schell, A. M., and Dawson, M. E. (2016). Lapses in skin conductance responding across anatomical sites: comparison of fingers, feet, forehead, and wrist. *Psychophysiology* 53, 1084–1092. doi: 10.1111/psyp.12643

Phipps-Nelson, J. O., Redman, J. R., and Rajaratnam, S. M. (2011). Temporal profile of prolonged, night—time driving performance: breaks from driving temporarily reduce time—on—task fatigue but not sleepiness. *J. Sleep Res.* 20, 404–415. doi: 10.1111/j.1365-2869.2010.00900.x

Pietilä, J., Mehrang, S., Tolonen, J., Helander, E., Jimison, H., Pavel, M., et al. (2017). "Evaluation of the accuracy and reliability for photoplethysmography based heart rate and beat-to-beat detection during daily activities," in *EMBEC & NBC 2017 IFMBE Proceedings*, eds H. Eskola, O. Väisänen, J. Viik, and J. Hyttinen (Singapore: Springer), 145–148. doi: 10.1007/978-981-10-5122-7_37

Pomplun, M., Garaas, T. W., and Carrasco, M. (2013). The effects of task difficulty on visual search strategy in virtual 3D displays. *J. Vis.* 13, 24–24. doi: 10.1167/13.3.24

Posner, M. I. (1980). Orienting of attention. *Q. J. Exp. Psychol.* 32, 3–25.

Qin, W., Xiang, X., and Taylor, H. (2019). Driving cessation and social isolation in older adults. *J. Aging Health* 3(Suppl. 1), S827. doi: 10.1177/0898264319870400

Ragot, M., Martin, N., Em, S., Pallamin, N., and Diverrez, J. M. (2017). Emotion recognition using physiological signals: laboratory vs. wearable sensors. *Adv. Intell. Syst. Comput.* 608, 15–22. doi: 10.1007/978-3-319-60639-2_2

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychol. Bull.* 124, 372–422. doi: 10.1037/0033-2909.124.3.372

Recarte, M. A., and Nunes, L. M. (2003). Mental workload while driving: effects on visual search, discrimination and decision making. *J. Exp. Psychol. Appl.* 9, 119–137. doi: 10.1037/1076-898X.9.2.119

Ruscio, D., Bos, A. J., and Ciceri, M. R. (2017). Distraction or cognitive overload? Using modulations of the autonomic nervous system to discriminate the possible negative effects of advanced assistance system. *Accid. Anal. Prev.* 103, 105–111. doi: 10.1016/j.aap.2017.03.023

SAE (2018). *Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems*. Available online at: https://www.sae.org/standards/content/j3016_201806/ (accessed May 22 2020)

Savage, S. W., Potter, D. D., and Tatler, B. W. (2013). Does preoccupation impair hazard perception? A simultaneous EEG and eye tracking study. *Transp. Res. Part F Traffic Psychol. Behav.* 17, 52–62. doi: 10.1016/j.trf.2012.10.002

Shinohara, Y., Currano, R., Ju, W., and Nishizaki, Y. (2017). "Visual attention during simulated autonomous driving in the US and Japan," in *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, Oldenburg, 144–153. doi: 10.1145/3122986.3122991

Smith, A., and Anderson, M. (2017). *Automation in Everyday Life*. Washington, DC: Pew Research Center.

Steg, L. (2005). Car use: lust and must. Instrumental, symbolic and affective motives for car use. *Transp. Res. Part A Policy Pract.* 39, 147–162. doi: 10.1016/j.tra.2004.07.001

Strauch, C., Mühl, K., Patro, K., Grabmaier, C., Reithinger, S., Baumann, M., et al. (2019). Real autonomous driving from a passenger's perspective: two experimental investigations using gaze behaviour and trust ratings in field and simulator. *Transp. Res. Part F Traffic Psychol. Behav.* 66, 15–28. doi: 10.1016/j.trf.2019.08.013

Strayer, D. L., and Drews, F. A. (2007). Cell-phone–induced driver distraction. *Curr. Dir. Psychol. Sci.* 16, 128–131. doi: 10.1111/j.1467-8721.2007.00489.x

Tay, R., and Knowles, D. (2004). Driver inattention: drivers' perception of risks and compensating behaviours. *IATSS Res.* 28, 89–94. doi: 10.1016/S0386-1112(14)60095-9

Thayer, J. F., Yamamoto, S. S., and Brosschot, J. F. (2010). The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors. *Int. J. Cardiol.* 141, 122–131. doi: 10.1016/j.ijcard.2009.09.543

Underwood, G., Phelps, N., Wright, C., van Loon, E., and Galpin, A. (2005). Eye fixation scanpaths of younger and older drivers in a hazard perception task. *Ophthalmic Physiol. Opt.* 25, 346–356. doi: 10.1111/j.1475-1313.2005.00290.x

van Dooren, M., and Janssen, J. H. (2012). Emotional sweating across the body: comparing 16 different skin conductance measurement locations. *Physiol. Behav.* 106, 298–304. doi: 10.1016/j.physbeh.2012.01.020

Voinescu, A., Alford, C., Caleb-Solly, P., and Morgan, P. (2020). The utility of psychological measures and situational factors in evaluating perceived usability of automated vehicle interfaces – a study with older adults. *Transp. Res. Part F Psychol. Behav.* 72, 244–263. doi: 10.1016/j.trf.2020.05.003

Walker, F., Wang, J., Martens, M. H., and Verwey, W. B. (2019). Gaze behaviour and electrodermal activity: objective measures of drivers' trust in automated vehicles. *Transp. Res. Part F Traffic Psychol. Behav.* 64, 401–412. doi: 10.1016/j.trf.2019.05.021

Xianglong, S., Hu, Z., Shumin, F., and Zhenning, L. (2018). Bus drivers' mood states and reaction abilities at high temperatures. *Transp. Res. Part F Traffic Psychol. Behav.* 59, 436–444. doi: 10.1016/j.trf.2018.09.022

Young, M. S., and Stanton, N. A. (2002). Attention and automation: new perspectives on mental underload and performance. *Theor. Issues Ergon. Sci.* 3, 178–194. doi: 10.1080/14639220210123789

Zheng, R., Yamabe, S., Nakano, K., and Suda, Y. (2015). Biosignal analysis to assess mental stress in automatic driving of trucks: palmar perspiration and masseter electromyography. *Sensors* 15, 5136–5150. doi: 10.3390/s150305136

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

PERSPECTIVE
published: 18 September 2020
doi: 10.3389/fpsyg.2020.574097

Driving Into the Future

*P. A. Hancock**

Department of Psychology and Institute of Simulation and Training, University of Central Florida, Orlando, FL, United States

This work considers the future of driving in terms of both its short- and long-term horizons. It conjectures that human-controlled driving will follow in the footsteps of a wide swath of other, now either residual or abandoned human occupations. Pursuits that have preceded it into oblivion. In this way, driving will dwindle down into only a few niche locales wherein enthusiasts will still persist, much in the way that steam train hobbyists now continue their own aspirational inclinations. Of course, the value of any such prognostication is in direct proportion to the degree that information is conveyed, and prospective uncertainty reduced. In more colloquial terms: the devil is in the details of these coming transitions. It is anticipated that we will see a progressive transformation of the composition of on-road traffic that will be registered most immediately in the realm of professional transportation in which the imperative for optimization exceeds that in virtually all other user segments. The transition from manual control to full automation will be more punctate than gradualist in its evolutionary development. As performance optimization slowly exhausts the commercial sector, it will progressively transition more into the discretionary realm by dint of simple technology transfer alone. The hedonic dimension of everyday driving will be dispersed and pursued by progressively fewer individuals. The traveling window of generational expectation will soon mean that human driving will be largely "forgotten," as each sequential generation matures without this, still presently common experience. Indications of this stage of progress are beginning to be witnessed in the demographic profile of vehicle usage and ownership rates. The purpose of the exposition which follows is to consider and support each of these stated propositions.

Keywords: driving, automation, cognition, human factors, autonomous systems

OPEN ACCESS

Edited by:
Oren Musicant,
Ariel University, Israel

Reviewed by:
Josef Krems,
Technische Universität Chemnitz, Germany
Anat Meir,
Holon Institute of Technology, Israel

Correspondence:
P. A. Hancock
peter.hancock@ucf.edu

Specialty section:
This article was submitted to Performance Science, a section of the journal Frontiers in Psychology

Received: *18 June 2020*
Accepted: *24 August 2020*
Published: *18 September 2020*

Citation:
Hancock PA (2020) Driving Into the Future. Front. Psychol. 11:574097. doi: 10.3389/fpsyg.2020.574097

INTRODUCTION: A SHORT GLANCE BACK – A LONG LOOK FORWARD

There are many and varied forms of human work activities which have, across history, been undertaken. Each of these pursuits would have been considered commonplace, natural, and everyday actions to the contemporaries who witnessed them. In and amongst these, for example, the blacksmith and the peddler were, at one time, almost ubiquitous sights in the world. But now these particular activities, like many other occupations, have largely disappeared from the public landscape. And this to such an extent that we need to access our web-based search engines even to find out who "gas-lighters" and "hop-winders" actually were. These latter pursuits were both common enough and persisted well into the middle twentieth century. How much more recondite are occupations such as night-soilers, town-criers, fletchers, alewives, mudlarks, and gandy-dancers, to name only a few. Not many today could even say what these latter forms of work actually were, or to suggest how the product of these endeavors shaped everyday life at that time. We see that technology changes the functional landscape of work and inventions such as cell phones serve to exterminate jobs such as "tic-tac" man, even to the point that these jobs are now effectively forgotten. The nature of work changes and we, as individuals and society, change along with it (Hancock, 1997). However, none of the aforementioned pursuits, even in their own day, were ever as ubiquitous or as well-recognized, as that of "driver." Indeed, drivers, in their many forms and incarnations (e.g., carters, teamsters, chauffeurs, truckers, pilots, steersmen, bicycle messengers, pyramid stone sled drivers, charioteers, etc.), have persisted now throughout an interval that can be even measured in multiples of millennia. As a result, our collective, social driving habits have been woven into the very fabric of civilized society and this driving enterprise is arguably an integral part of virtually all nominally "civilized" collectives. Few are the people who do not meet and encounter drivers regularly or indeed for that matter participate themselves in driving on a daily basis. But will drivers go the way of typesetters, switchboard-operators, or even more appositely, the human-computer; names which now ring only anachronistically and obscurely in our modern ears? Thus, the aspiration for the present evaluation is to consider and specify what precisely is driving us into the future? Given this ubiquity, the focus of the present work consists of an examination of the following important propositions. (i) What will compose driving in the future? (ii) With the onset of vehicle-control automation, will the profession and skill of driving fade, like others, into memory? and (iii) What of society in a world where no humans drive?

As Yogi Berra is reputed to have observed, "*prediction is hard, especially about the future.*" However, the purpose here, as it has been in other associated works (e.g., Hancock, 2008), is to ruminate upon what, with respect to driving, is to come. The magnitude of change that promises to occur with the widespread penetration of autonomous vehicles (Rajasekhar and Jaswal, 2015), is very much in proportion to the extent of driving's past and persistent longevity as well as its present ubiquity (and see Litman, 2017). Thus, the uncertainty which is associated with this anticipated degree of change is great in proportion (Hancock et al., 2019). To begin, we need first to briefly glance back in time in order to proceed cogently and thoughtfully into the future. Although it is not the purpose here to examine and rehearse the evolution of the specific role of driver in any exhaustive detail, it is enough to assert that we can find evidence of individuals in charge of some form of powered transportation back almost to the edges of recorded history. Mostly, this began by using animals as the power source, and with these capacities, drivers conveyed passengers and material (in all its forms), from origins to destinations; transport being the heart of trade and commerce and so the arteries and lifelines of civilization. Only consider here, for example the "Silk Road," which covered even thousands of miles (Liu, 2010). The demand that was imposed upon early drivers tended, to some degree, to covary with the purported "intelligence" of the pack animals involved. Oxen proved to be sturdy but exhibited relatively little independent intentionality.

FIGURE 1 | The donkey-powered water wheel of Carisbrooke Castle which was used to raise a water-bucket from an extremely deep well (at right). While the donkeys need "training" as to how to walk on the wheel, once that skill had been mastered, little in the way of subsequent human intervention is required.

Mules are hardy but, in human eyes, prove relatively stubborn compared to their close relative, the horse. Some animals, such as sled dogs, are viewed as exhibiting especially high levels of intrinsic intelligence, and so enabled the driver to proceed in a less immediately controlling manner. Regardless of the degree of these inherent levels of animal intelligence, if some rigid constraints could be imposed upon their actions, then selected animals can act almost independently (autonomously) from their human supervisor. If the constraining context is framed with sufficient ingenuity, e.g., a donkey on a wheel, then minimal human intervention may be required (see **Figure 1**). From this point of view, autonomous transportation is not necessarily a recent invention but one which, in differing guises, has been around for quite some time, e.g., "*Open Sesame*." Perhaps the first watershed in driving, at least in terms of ground transportation, was when the source of power morphed from animal to artificial origins. The co-variation here with the Industrial Revolution being no happenstance. But now we stand on the verge of the next, and arguably, more profound watershed in vehicle functionality. This change is qualitatively and quantitatively distinct from any previous incarnation of transition. I frame the understanding of this approaching watershed in light of the recognized, and above referenced step-change, from animal to artificial power.

FROM LIVING MUSCLE TO ARTIFICIAL POWER

It is evident here that I have not featured any substantive discussion of wind-power and the comparable human conquest of the oceans as routes for trade and social interaction (e.g., Revelle, 1962; Campbell, 1995). However, in principle, each of the observations that I make are as applicable to "driving on the sea" as they are to "driving on the land." To an extent, the nature of driving changed radically when artificially powered transportation became widely available. Effectively, the first of these forms of transport came about with the development of the steam-engine although, as I have noted, it can be argued that wind-powered vehicles preceded steam power, again even by millennia. In steam-powered trains, drivers tend largely to exercise control over one essential single degree of freedom (see e.g., Oborne et al., 2014), that being longitudinal velocity. Any degree of lateral control, i.e., route selection, point manipulation etc., tended to be the domain of another human decision-maker within the overall system. Although the train driver's task consists of more than just controlling this one variable, the steam train footplate personnel were primarily involved in this activity either by facilitating or retarding longitudinal velocity on a momentary or more prolonged timescale. Some of these same strictures transferred to ground transportation when steam power was extended to on-road vehicles (Lay, 1992). However, it can be argued, and convincingly so, that horse-riders had already mastered many of these skills and the coachmen of the horse-drawn carriage era would transition fairly seamlessly into powered-vehicle chauffeurs, as history confirms. The inherent driving control skills themselves did not change radically with the transition of the source of power, but the growth of peak vehicle velocity began to impose higher, or more accurately different, cognitive demands upon the person at the controls.

It is not formally known, but can well be suggested, that the cognitive workload imposed upon early powered-vehicle drivers on the road proved to be an increment over that for say steam-train drivers. For, as we know, early steam-trains required a full footplate crew, whereas, effectively, steam-powered cars did not. The invention of the auto-starter also had important effects here since it obviated numerous procedural steps involved with vehicle activation, some of which required quite satisfying quite extensive degrees of physical workload. Here, the combination of physical and procedural constraints proved to be a barrier that technological advances could and did resolve (and see Möser, 2002). Also, we see that the roadway context of driving and the density of traffic began to exert further influences; although a horse-drawn carriage driver in downtown New York of the late nineteenth century would protest that their task had been no sinecure! This assertion about increments in workload can be a polemic one because each of these respectively identified roles were, and had been, composed of multiple tasks, as most professions were then and still are today (Hancock et al., 2020). Regardless of any such disputations, the evident fact is that each of the respective modes of transport were created or evolved so that human controllers could satisfactorily accomplish the task that was then set before them. There is little point in creating a vehicle that is absolutely uncontrollable. In sum, across the ages, driving has represented a satisficed not an optimized task (and see Simon, 1969), and the fabricated road environments vehicles occupy were specifically and intentionally structured to support this form of functionality and thus the associated level of sustainable cognitive demand. These various historical predicates for cognitive workload regulation mean that collisions have proved to be relatively infrequent. For example, the Insurance Institute for Highway Safety (IIHS), provide data that indicates that in the United States in 2018 there were only 1.13 fatalities per 100 million vehicle miles traveled (Insurance Institute for Highway Safety, 2020). This startling lack of fatal collisions becomes especially evident when we begin to consider the number of *potential opportunities* for collision on roadways, a point upon which I elaborate more below. The next threshold, which promises to be that of fully automated driving control is one decked with the "banners" of improvements in safety and efficiency (Kyriakidis et al., 2017; Hancock, 2018a; Hancock et al., 2020a). The empirical question, however, is whether autonomous vehicle collision avoidance capacities can now ubiquitously exceed the proven rates of human-mediated avoidance? In some sense, this question matches the evident degree of success of the human driver in coping with the imposed cognitive workload of driving vs. the equivalent capacity for autonomous vehicles to deal with that same imposed external demand. In the section which follows, I examine this proposition concerning the replacement of the human controller (driver). I aim to do this by throwing a purposive and explicit light on the quite remarkable abilities of humans to adapt to the satisficed demands of everyday driving. In short, my purpose, pro tem, is to emphasize just how good human beings actually are already at driving.

THE MOST PRACTICED SKILL IN THE WORLD

Unlike many of the human professions and pursuits which, as we have seen, have now faded into the mists of time, driving has persisted and grown in proportion to the size of the populace and number of vehicles in circulation. To the present, we have experienced essentially two centuries of powered ground transportation with well over 100 years of individually driven automobiles. It is fair to say that the very landscape of countries such as the United States, Canada, and the like, have been sculpted by the presence and utility of the automobile and its particular needs (see **Figure 2**).

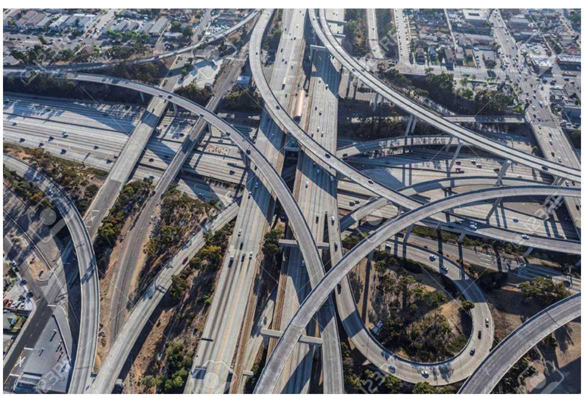


FIGURE 2 | The nature of the urban landscape is massively influenced by the need to cater to powered vehicles and the requirements and ramifications of the overall transportation infrastructure (Ingram and Liu, 1999). It should also be noted that the nature of the rural landscape is also contingent upon this requirement, although materially, this can appear to be less impactful (Image reproduced with permission. iStock.com/Domepitpat).

In many, if not most cities of the world, the service of the automobile is a principal concern. Even in countries, such as those in Europe, where towns and cities were never explicitly or originally designed for the modern automobile, the car's impact exerts considerable and persistent effects. This re-engineering of the cityscape to accommodate the automobile stands, to some degree, in concert with the way that the burgeoning sea-based trade of Amsterdam sculpted its reticulation of waterways and canals. Accommodations for vehicles either motivated or manipulated our modern urban landscapes and if we change vehicles' functionality alongside the inevitable changes in the nature of the infrastructure that supports them, it promises to have a profound social impact, well beyond the confines of the vehicle itself (Stead and Vaddadi, 2019).

The area of urban design, the interaction with roadways and their influence is itself a research domain of vast extent (see e.g., Kenworthy, 2006). It is sufficient here to acknowledge these broad and diverse impacts. However, the present focus here is on individual human behavior in executing the role of driver. It is reasonable to assert that there is, and has been, no human skill more practiced than driving. True, all individuals do not engage in driving and so this is not a truly ubiquitous human practice (Hancock et al., 2009). Rather, the statement is applicable as more of a socially, nomothetic assessment. When we total up the number of commuters, professional drivers, and vehicle owners in general, it can be seen that vast swathes of the human race engage in this one activity and that it exceeds essentially any other singly practiced public skill; certainly in terms of the number of hours involved. It is also reasonable to assert that driving is the last great bastion of "analog" control. This does not mean that digital technologies are not involved in the control functions of modern vehicles, assuredly they are. Rather, it means that driving still requires the momentary exercise of psychomotor skill for continuous tracking whereas, in comparison, the vast majority of our other daily skills are now almost fully digital in nature and largely, or even solely require punctate mouse-clicks or button presses of the user. The exception here being video games which still feature this fulfilling, "tracking" aspect of human experience. As with many skills, and especially psychomotor skills, practice improves capability, even across periods of decades or more on the same task (Crossman, 1959).

With respect to such skill acquisition and its exhibition (Newell and Hancock, 1984; Hancock and Newell, 1985), humans are good at driving and are arguably, on average, even excellent at collision avoidance (cf., Yuris et al., 2019). But how can this be? We are all aware of the carnage on the roads and especially in the driving research community, we have been imbued with fatality and collision rates as the mantra of concern. But what we have never really attacked is the question of the relative rates of these human failures; that is, specification of the elusive denominator; the absolute number of non-events. In reality, how many non-collisions do actually occur per unit time to set against these adverse and life-altering vehicle accidents? It is quite natural for researchers, as well as the public in general, to focus upon the events that did occur and to direct scant, if any, residual attention to the events which did not. It is a human failing of both memory and ratiocination that we are poor at calibrating with respect to all forms of non-event. While there are clear evolutionary reasons why this neglect should be so, it still serves to bias our assessment in multiple areas, especially when the positive events prove as dramatic and life changing as a serious road collision. However, if we take as the unit of "non-collision" the space occupied by a vehicle multiplied by the time it occupies that space, and then reference this value to the frequency of actual collisions, which are represented by two vehicles or objects in transportation research (or more properly, any two material entities) occupying that same unit of space and time. If we were to conduct such a calculation, then human capacities must be well in excess of the 99.9999% reliability level. Arguably, it is even several orders of magnitude better than this. Of course, these levels of performance are not independent of increases in cognitive workload and effort, especially when driving conditions become demanding (Hancock et al., 1990). Automation does make it possible for people to begin to select the level of their participation, but they are not able to do so when automation in driving become obligatory. Nevertheless, this relative degree of human driver reliability makes the grand claims about safety gains for autonomous vehicles difficult to fulfill on both relative and absolute scales. The idea that eliminating driver error can be done by eliminating drivers is therefore rather problematic. As noted elsewhere, any absolute gains in collision reduction is a morally laudable achievement (Hancock, 2019a). However, the full scale of the issue in which both the numerator of collision frequency is juxtaposed to the denominator of non-collision frequency, has still to be even approximately identified, quantitatively speaking.

Claims for greater efficiency, in respect to transit times, may well suffer from similar lacunae in data specification. That is, individual transit times may be collected and plotted as putatively representative samples, but then their expansion and aggregation into assessment across the complete transportation system to hand is fraught with all of the perils of generalization. Obviously, as fully automated vehicles begin to predominate these associated electronic calculations become more tenable. However, in reality, the problem of mixed equipage, consisting of many automated vehicles interspersed with those of differing and nominally "lesser" capacity, serves to inhibit precise transit time specifications. However, the latter metric of change in transportation "efficiency" may perhaps, objectively, be somewhat easier to realize than the more safety critical failure-collision, non-collision ratio as has been mentioned above. Of course, these touted gains in efficiency are exactly what are held out as vitally supportive reasons for embracing automated driving in the first place. The mantra runs that automated vehicles can be "stacked" more efficiently together within the various roadways, traveling mere feet from each other in platoons, convoys, and the like. Technically, this has been shown to be a feasible configuration (see Lee et al., 2018). Yet what remains largely unproven are the outflow influences of these selective groupings on other traffic; presumably at some other juncture "embedded" within the same system. Tantalizing desiderata, such as greater "free time" is also offered to the innocent traveler, so inclined to adopt these commercially motivated innovations. However, as I have argued elsewhere, increasing "time" tends to be vacuumed up by the profit-driven system which then exploits the enhanced opportunity to have individuals now "work" on their way to work (Hancock, 2019c). In other words, this promise will largely serve to simply re-cast the physical location of, mostly, electronic work-based interactions. We have all become witness to this very phenomenon, as involved in response to the recent pandemic.

Critically, of course, lauding human driving abilities does not sell driverless vehicles. Nor does the defense of human capacities involved in driving greatly capture the attention of an information-jaded public. It seems, therefore, that the autonomous vehicle juggernaut will roll on despite any such observations (Daily et al., 2017). However, faced with this almost inevitable line of autonomy's progress, the next consideration has to be one of the transition periods between human and computer control, and what challenges we are facing in the immediate, near-term. This particular challenge is tied to the presumed map of the transition phase bound up with the "levels of automation" taxonomy. This is most evidently articulated in the six Society of Automotive Engineers (SAE) levels conception to which I now turn.

THE LEVELS AUTONOMY DESCRIPTION AND ITS ASSOCIATED FALLACIES AND FAILURES

To this point, the present work has been framed around the general arguments concerning the long arc of driving history. It will in its concluding section, proceed to some prognostications concerning driving's future and some of the associated incarnation (and see also Pettersson and Karlsson, 2015). However, in terms of past, to present to future, it is important to consider the volatility and change embedded, especially within our own present transitional times (Hancock et al., 2013). The focus here is on the now, quite famous, and relatively ubiquitous "levels of automation" taxonomy and some pertinent critiques of it, as well as the path forward that it apparently offers. Criticisms here are somewhat palliated since the SAE construct does have facets of evident utility. Precisely where and how this formulation developed is a task for others to establish. Suffice it to say that I find that much of the thinking underlying these levels of automation formulation can be attributed to one of the luminaries of human-machine interaction; namely Thomas Sheridan (e.g., Sheridan and Verplank, 1978; Sheridan, 2002; and see Sheridan, 1992, p. 358, also Hancock and Sheridan, 2011). His *"ten levels of interaction"* proposition was one that could apply to many operational contexts and domains. Here, via the SAE promulgation, it has been directly applied to the transitional phases of driving control. There are a number of pertinent criticisms that are relevant to its application to the future of driving, whatever form that future driving takes (and see Parkes and Franzen, 1993).

We can begin with the physical form of the SAE scale itself (see **Figures 3A,B**). The description begins at a zero level and provides what appears, putatively, to be a series of equal integer steps. The first impression is that we are looking at these respective steps from 0 and 5 on an apparent ratio scale, although whether this ratio implication was ever actually explicitly intended is probably rather doubtful. This implies an equal-interval structure between each of the discrete steps. This is a false implication and can be extremely misleading. It begins with the assertion that zero provides a no automation baseline state, but in itself this is also simply incorrect. Even for vehicles which well-preceded the modern, large-scale transportation thrusts such as the "*Intelligent Vehicle-Highway Systems*" (IVHS) (and see Hancock and Parasuraman, 1992), there was plenty of automation in cars, and some of it associated with immediate roadway control such as "cruise control." This assumptive foundation of a zero level is evidently in error. A further implication of the SAE taxonomy is that each sequential step, between the respective levels, represents an equivalent change in functionality. As we are experiencing now, this implication is also false; most especially as we consider Stages 2 and 3 and compare them with Stage 4 for example. A further indication that this is an engineering-oriented perspective derives from the fact that there are actually six total levels identified, although the use of a zero anchor tends to convey otherwise. What this means is that there is no true intermediate step between the respective ends of what appears implicitly to be a continuum. The absence of such a "middle" state inhibits conceptualization to a degree that is not immediately obvious to users and/or designers who first encounter this influential representation. Some will chide that these objections as either rather trivial or only the musings of a nit-picking criticaster; but far from it. This representation has been taken as a form of *de facto* design "roadmap," laying out apparently sequential and logical steps to achieve a fully autonomous future; a goal that itself is most often not adequately questioned. But it is the shortcomings in this conceptualization that presently threaten to lead to disruption and dysfunctionality. Of course, since the elaboration of transportation systems in the real-world is largely an empirical exploration anyway, and so no comparable "control" in order to assess the degree of any such dysfunctionality, is ever really feasible.

It is not simply the temptingly pristine representation of each individual step which proves to be problematic. The boundaries between each putative "level," so readily and solidly illustrated in **Figure 3A**, and to a somewhat lesser extent in **Figure 3B**, are themselves frangible thresholds. In actuality, each sequential stage, at least to some degree, bleeds into some of its companions, and that not always to the level that is immediately adjacent to it (i.e., some elements of Stage 2 link directly to Stage 4, etc.). And underlying the whole illustration is the unstated, but highly influential implication, that "progress" necessarily requires us to move ever-upward in this hierarchy of levels (i.e., Stage 3 is two better than Stage 1, etc.). Through further subliminal suggestion, it also implies that full vehicle autonomy "must" be the socially desirable and ultimate design goal that we are aiming for. After all, in general is not anything better than zero? And if 5 is the top, is that not what we are aiming for? Rather than accept this assumption, it is one that we should most severely question (and see Hancock, 2014, 2017a). What precisely is the explicit, pragmatic need for automated vehicles? Do we not presently have enough humans on the face of the Earth in

A

SAE level	Name	Narrative Definition	Execution of Steering and Acceleration/ Deceleration	*Monitoring* of Driving Environment	Fallback Performance of *Dynamic Driving Task*	System Capability *(Driving Modes)*
***Human driver* monitors the driving environment**						
0	No Automation	the full-time performance by the *human driver* of all aspects of the *dynamic driving task*, even when enhanced by warning or intervention systems	Human driver	Human driver	Human driver	n/a
1	Driver Assistance	the *driving mode*-specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	Human driver and system	Human driver	Human driver	Some driving modes
2	Partial Automation	the *driving mode*-specific execution by one or more driver assistance systems of both steering and acceleration/ deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	**System**	Human driver	Human driver	Some driving modes
***Automated driving system* ("system") monitors the driving environment**						
3	Conditional Automation	the *driving mode*-specific performance by an *automated driving system* of all aspects of the dynamic driving task with the expectation that the *human driver* will respond appropriately to a *request to intervene*	System	**System**	Human driver	Some driving modes
4	High Automation	the *driving mode*-specific performance by an automated driving system of all aspects of the *dynamic driving task*, even if a *human driver* does not respond appropriately to a *request to intervene*	System	System	**System**	Some driving modes
5	Full Automation	the full-time performance by an *automated driving system* of all aspects of the *dynamic driving task* under all roadway and environmental conditions that can be managed by a *human driver*	System	System	System	**All driving modes**

B

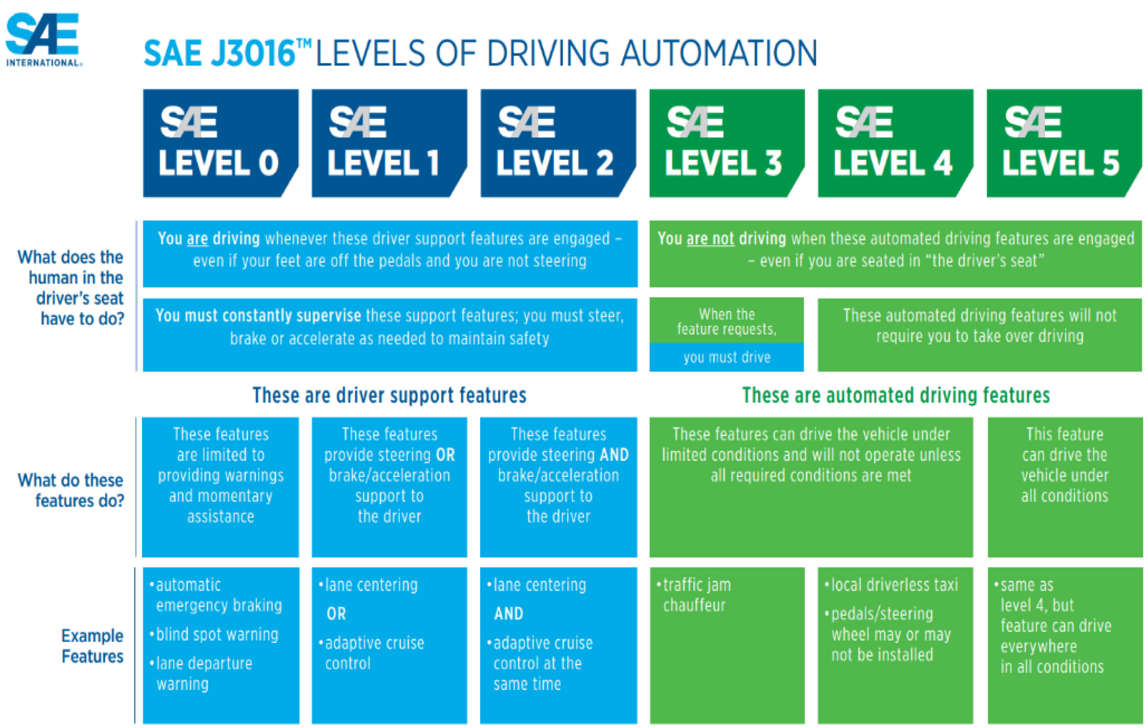


FIGURE 3 | (A,B) SAE specified transition phases between no automation and full automation. The upper version is from 2014, the lower is the most recent revised version and is designated: SAE J3016 Standard: Levels of Driving Automation," and is reproduced here with permission of SAE International. Propagated as a form of descriptive taxonomy, it promises to become a design ontology. Therein lies a number of debatable and potentially misleading assumptions as articulated in the text.

order to provide a sufficient number of controlling "drivers?" And when the evolving shibboleth of improved safety is once again paraded before a somewhat naive public and even trooped before professional scientists, let us ask expressly, where are the data to confirm this assertion? And most especially, where is the data that directly compares human safety records with automated control performance in exactly the same operational conditions? (and see Hancock, 2018a). Until this information is produced and analyzed, if indeed it is actually being recorded in such fully comparable instances, we will not know whether the whole process is indeed "safer," or whether safety is once again being used as a "weasel" word to mean whatever its protagonists choose it to mean in order to convince others of their case (and see Hancock and Volante, 2020). There are no necessary reasons why many, if not all, of the assumptions embedded in the SAE levels description may not be flawed or simply incorrect.

With respect to the perspective promulgated in the five levels conception, there is another assumption, perhaps even more insidious for its unstated presence. This is that control is, in essence, a zero-sum function. In this concept, what any automated and autonomous systems gains, the human must necessarily lose. There have been numerous recent and cogent challenges to this assumption (see e.g., Shneiderman, 2020). There is no necessary reason why an overall expansion of mutual degrees of control could not be enacted i.e., a greater than zero-sum. This requires us to think of technologies more as team-mates than simple surrogates or direct replacements. The fallacy of the zero-sum of control persists only if we think in the constrained terms of monetary vehicle control. However, if we elevate our argument to a more macro-level concern for transportation and its diverse demands, this underlying restrictive premise is fractured, and the expansive vision of mutuality and sharing emerges (and see Gadsden and Habibi, 2009). In the various points discussed above, the "levels of automation" have, to a degree, morphed from an initial descriptive taxonomy to an eventual design ontology. As representative of our approaching transition into a new incarnation of driving, what is in essence the next watershed of driving itself, we need now to discuss the most immediate, problematic element of the SAE depiction as representing a form of an immediate future driving roadmap.

As has been discussed and elucidated elsewhere (e.g., Hancock, 2019a), one of the most pivotal issues of today concerns whether it is even feasible for a driver to recover full active control of their vehicle in the Stage 2 or Stage 3 conceptions of these proposed automation levels. While this might possibly be feasible in other realms of transportation (e.g., large container-ship control, and to a lesser extent commercial aircraft control; Scallen et al., 1996), the time-horizon limitations for successful resolution to momentary on-road challenges and/or automation failures seems sufficient to defeat the advocated and advertised resumption of manual control in these ground vehicles. It should also be noted that any innovative change, even in the other, arguably less temporally challenging transport circumstances (i.e., air, ocean), has been accompanied by transient increases in failure rates. We must be prepared for these spikes in adverse outcomes, as is discussed in more detail below (and see Hancock, 2011, 2019a). Most especially, this seems to apply to vehicles in high-density traffic situations, and/or on high-velocity roadways. It is not that we cannot conceive of these control transitioning technologies, but rather whether it is practicable, feasible, and even advisable to pursue these forms of control return strategy (Desmond et al., 1998). For, in these respective stages, the "driver's" role is translated into one of passive monitoring, which is a role that we understand already that humans can be extremely poor at (Hancock and Warm, 1989; Hancock, 2013b, 2017b). Response latencies increase across time in such vigilance situations (Hancock, 2017b), as do the frequency of missed signals as epitomized in the well-recognized "*vigilance decrement function*."

Although Stages 2 and 3 are more than difficult to manage in terms of human-vehicle interfaces and recovery response, if we do choose to adopt these forms of transition, it may be especially pertinent to consider a "driver command by negation" architecture (and see Hancock and Verwey, 1997). Here, the automation temporarily assumes complete command and so communicates that state through differing perceptual modes (e.g., voice warnings, visual icons, etc.). The only response required of the driver at this juncture is a no statement. This form of interaction does not require any form of an affirmation, i.e., "I agree." Rather, the human requirement here is only an interruptive "command by negation," i.e., "I disagree." Pilot command by negation (PCN) is quite a well-known as a human-machine communication strategy, especially in aviation (Hancock, 2007). This command structure follows an old maxim in law being *qui tacet consentit* or, "he who is silent agrees," or even more simply, "silence gives consent." However, it is almost without doubt that the time constants involved in ground-vehicle control takeovers will eventually defeat any form of human interaction after putatively regaining such control (Hancock, 2019a). In terms of driving, the temporal paradox might be expressed as follows. The Stage 2 and 3 takeover policies almost inevitably require that we know the future beyond the time horizon that normative perception-response provides. That being so, we would need a degree of prospection that we do not currently possess. Indeed, if we did possess prospection to this degree, we would already be able to anticipate that future event successfully enough that take-over would be unnecessary. The temporal constraints of the human response system itself prevents this from happening (Hancock and Weaver, 2005; Francis et al., 2020).

One further unwritten sub-text here is one that features the issues of control and legal liability. For, if we employ the traditional vision of liability, then the individual human driver is both responsible for control and thus at fault when failure occurs. However, if we employ a systems-based perspective, the skein of responsibility becomes much more complicated and potentially exposes the vehicle's manufacturers, and their constituent sub-contractors, in a way inimical to their own best interests. This is one of the touchpoints where the radical differing "magesteria" of technology (science) meets that of the law (Hancock, 2020). It will be especially edifying to witness how the two competing visions of driving, i.e., the commitment to driverless vehicles vs. the shared control/driver assist strategies play out in the coming decade, especially in light of this liability issue. Given the tide of technological progress, it appears the long-term winner is basically already decided in the ground transport realm. However, the race is still in progress and the forces which continue to favor a human driver-centered approach are by no means negligible.

A final issue upon which we can reflect concerning human interaction with autonomous vehicles is the concern for attribution error (Hancock et al., 2020b; Stanton et al., 2020). Much of the interaction on our roadways is mediated currently by implicit communication between human drivers. Such behaviors are, for example, evident at unregulated intersections, where eye-contact can mediate arrival and departure priority. Attribution is also contingent upon facets of behavior such as courtesy and etiquette. These self-same strictures are in action when drivers interact with other road users such as pedestrians, bicyclists, etc. Here, shared common assumptions and expectations mean that it is not only the formal rules of the road which guide action but informal, social ones also. This is also why driving in differing countries with different cultures and varying implicit assumptions can prove to be rather stressful. The central point here being that courtesy, empathy, and implicit knowledge are not yet built into automated vehicles nor their controlling software. Nor does there appear to be any great customer clamor for manufacturers to do so. The question that we can ask is whether such attributional dissonance between human and automated vehicle will necessarily lead to conflict, confusion, and collision (Hancock, 2018b). This concern is, of course, one that is greater than automated terrestrial vehicles, for it asks questions about our physical interaction with all other objects, and most especially advanced forms of technology. For example, how are robots supposed to react to the presence and motion of people, both their users and others in their ambient environment? Here, the principles of biomimesis gives us a lead. For, in nature, we often implicitly understand our role in gatherings such as crowds, or in unusual situations such as cattle stampedes or during the running of the bulls (in the latter situations, getting out of the path of the animals being a recommended strategy). These sorts of principles of self-organization and self-separation are now being codified into advanced commercial aircraft. It is almost certain that they will be incorporated into driverless vehicles also. In conclusion, with respect to the present stage of control transition, the SAE description has proved to be a provocative and probing proposition. While not without some element of value, it has served to both frame and constrain the avenues of progress in the driverless vehicle world to arguably, a disproportionate degree of influence.

Although not engaging in prescriptive designations *per se*, it still remains possible to consider the relative advantages and disadvantages of both human-centered and automation-focused driving propositions, and this I have presented in **Table 1** which follows. I am very aware of the propensity to couch these forms of observation in the comparative terms of so-called MABA-MABA (men are better at; machines are better at) types of juxtaposition. Indeed, the merits of such contrasts have previously been discussed and debated, and that rather extensively so (see e.g., Sheridan et al., 1998; Dekker and Woods, 2002; Hancock, 2007; De Winter and Hancock, 2015). As a result, the observations given in **Table 1** serve more as points of discussion for greater scientific and social consideration rather than hard and fast rules for function allocation between either human, automation, or human and automation together working in some form of concert.

In general, we have become well aware of the capabilities and flaws associated with human behavior (e.g., see Hancock and Matthews, 2015). In particular, drivers can become fatigued, stressed, and/or distracted (Hancock et al., 2003). In part reaction, the promulgated advocacy for automation is that automated vehicles would not be vulnerable to these influences. Similarly, human drivers can only tolerate certain absolute levels of task-related workload, and theories of driving have even been predicated upon each individual's management of this, their own dynamic levels of regulated task demand (see e.g., Fuller, 2005). Further, although human drivers can, in general, satisfy the demands placed upon them, there remain large differences in individual capacity which means a lack of uniformity of competence of the human drivers on our roadways.

Yet the promised alleviation of these concerns for human shortfalls by aspiringly autonomous vehicles does not itself come without costs. Members of the driving community, in giving up control, also give up some degree of personal privacy. They certainly give up some aspects of personal autonomy and in so doing they must interact with machines whose full spectrum of functions they need not necessarily either understand or trust (see Hancock et al., 2011a,b). Such machines are, like all computer systems, to a degree, vulnerable to cyber-attack. Rational decisions as to the strategy adopted for future transportation ought to be founded on factually supported assessments of the veracity of purported gains of automation, as compared to what is frequently advertised by those wishing to sell advanced vehicles (Hancock et al., 2019). In respect of both the promised gains in safety, in terms of reduced frequency of collision, and in terms of gains in trip time efficiency, the data should be determinative over advocational publicity. Sadly, as with many such social policies, this does not promise to be so.

Many of the present concerns with the approaching incursion of automated vehicles are ones that are necessarily embedded in the period of transition that we currently inhabit. It is also important to note that driving is only one example of this transition, as the same issues task many other operational domains. This general debate is articulated in further detail in the section which immediately follows. In a number of ways, this global transition from human to computer is epitomized in the description of SAE Stages 2–4 in which the emigration of control from the human is taken up, in a form of zero-sum way, by the attendant automation. However, as noted earlier, progress through each of the discrete stages is neither necessary nor obligatory. However, as with comparable innovations in advanced aviation operations, we will witness various transient effects during this epoch of transition. These transients have to be recognized so that collectively, we can be cognizant of the degree that the outcome pattern of performance witnessed is reflective of the novel elements associated with such transitions, as opposed to any fundamental flaws in any particular line of

TABLE 1 | Side-by-side descriptions of a series of advantages and disadvantages for human control juxtaposed with automated control.

Driver-controlled		Automation-controlled	
Advantages	**Disadvantages**	**Advantages**	**Disadvantages**
Significant pool of accumulated skill	Obligatory task focus for effective Performance	Advertised superiority in transit efficiency	Potential violation of personal privacy
Often able to respond to unexpected events	Human controller suffers from progressive fatigue	Advertised reduction in collision frequency/intensity	Vulnerable to cyber attack
Proven low relative error rate	Vulnerable to stress and workload disruptions	Readily scalable for widespread utilization	Difficulty in dealing with bespoken challenges
Control fosters individual self-efficacy	Poor at extensive monitoring activity	Ready inter-operability with other automated systems	Contemporary lack of all affect
Capable of subtle forms of pattern-recognition	Relatively slow rates of skill accumulation	Little performance degradation across time	Imposes constraints on personal human freedom
Uses most complex control mechanism currently known	At risk for distraction during active control	On-line, remote operational improvements possible	Non-transparent operational states
Able to experience joy and fulfillment	Large individual differences across user population	Currently perceived as "inevitable."	Vulnerable to rider distrust and neglect of use

The vulnerabilities of this structure to misinterpretation as deterministic compositions between the relative capacities of human and machine are articulated in the associated text. Descriptions of the team advantages and disadvantages have not been presented in the Table but are discussed in the text.

technical development. The latter concern emerges when failure rates spike and companies and institutions, whose personnel are attuned to such acute changes in event numbers, tend to react accordingly. Much here depends upon the business cases made for the innovations offered and the way that the market responds to these technological offerings. There is also, of course, an overall propensity to reject retrenchment such that when a new system is initiated, it proves hard if not impossible to return to the older approach, should the new one fail to deliver on its advertised promises. This form of antipathy to putatively "backward" steps in technology is a powerful force, and one to be reckoned with.

In ground transportation, the period of transition is liable to be quite a prolonged one. More formally, the percentage prevalence of on-road legacy systems is liable to be high. This is in part because, like the function of transportation itself, vehicles are not exclusively utilitarian in nature. Indeed, a non-trivial proportion of travel is undertaken for purely hedonic reasons and the ownership of vehicles is not solely for pragmatic transit, but often for the pleasure of ownership. Like other high legacy systems, such as firearms, there will then persist in use a broad mixture of age and capability of on-road vehicles. The efforts of infrastructure designers to parse these various segments of the traveling inventory into differing regions of spatial operations, or temporal distinctions in terms of permitted hours of operation, will be motivated by the imperatives of efficiency. However, these forces will be pitted against the persistent desire for freedom of operation. It is liable to be a complex and polemic trade-off involving such opposing forces. And across this checkered landscape is emerging the observed trend for reduction in vehicle ownership rates, especially among the cadre of younger drivers (cf., Shaheen and Cohen, 2013; Knittel and Murphy, 2019). As to the stability of these trends and their social, cultural, and national variations across the globe, the trends are rather regionally contingent. However, there is little doubt that the injection of permanent circulating autonomous vehicles for hire and their immediate availability via smartphone linkages will further serve to influence vehicle purchase decisions as future generations face their own mobility challenges. In sum, transition is liable to be a motif of transport systems for some time yet to come. However, transport exerts wider influences than simple passage between origin and destination alone. It is to these impacts that I next turn.

DRIVING AS THE "WEDGE" ISSUE: TO CHANGE DRIVING IS TO CHANGE SOCIETY

In many ways, driving very much represents the "thick end of the wedge." It is, what I have thus termed the modal, "wedge" issue concerning the penetration of ever more autonomous systems into human life (Hancock, 2015, 2019b). In the steps from an analog to a digital world, driving retains at least the vestiges of a past and passing era. Lives across the world have already been vastly transformed by this revolution, but now the tide of such change is attacking perhaps its last and most formidable bastion. As such, we are not simply looking at the future of driving, we are surveying what promises to be a differing way of social organization and human existence. The issue of momentary control is embedded within a much wider concern for personal autonomy. That is, when we relinquish the effective momentary control of the vehicle, we are also abdicating from the full expression of freedom that it represents. Future vehicles will come pre-programmed with multiple origins and destinations; and even some present vehicles have these resident capacities. Departures from these everyday journeys linking home, work, grocery store etc., will be exception processed. But just as no man is an island, so no autonomous vehicle will actually be completely autonomous. That is, such vehicles will necessarily be embedded into a systems-wide integrated transport system that will require to know about each componential element; where it is coming from and where it is going to, and when. While this requirement might well make it harder to accomplish "getaways" after robberies, a constraint that socially we might approve of, it will also interfere with many other more provocative dimensions of privacy. For example, how do we organize a surprise party, when the data are necessarily available to let inquirers know exactly where one's friends and colleagues are? This might be a fairly puerile example, but it does serve to make the point; there are occasions upon which people do not wish to let anyone or anything know where they are going. Changing the nature of momentary vehicle control thus has outflows into society that need be neither immediately obvious nor easily anticipatable. As mentioned in the summary here, we also neglect to consider the pleasures of driving at our peril.

As the proponents and pilgrims of higher, automated "safety" wend their unhindered ways through the lines of social discourse, what of those who still personally want to drive? Must their pursuit be limited to out of the way facilities, fit only for enthusiastic hobbyists? Will we not lose something more than a symbol when the steering wheel is finally abandoned? Driving is not merely the simple act of vehicle control; it is a declaration of personal expression. In a world where the natural propensity of the dominant consumerist system is to curtail such elaborative forms of human behavior, automated control of one's personal vehicle represents another step along the straight-jacketed road to obligatory social conformity. Let us then beware of what is here driving us into the future. For, it may not be the panoramic promise of autopia that we are being taught to visualize, but something potentially much more disturbing.

ARE WE THERE YET?

How far off is the future? This is always a thorny conundrum. No one disputes the fact that semi-automated vehicles have already been "let loose" in the world to conduct an informal empirical exploration of their capacities in the wild as it were. And these more recent incarnations are inevitably introduced into worlds which, for the foreseeable future will, as has been noted, still contain various and traditional forms of human vehicle control. With temporary transference of human driver skills across vehicles, such as that experienced in renting a car, it may be that drivers will traverse the differing levels of tactical and strategic control, as represented by Stages 2 and 4 of the SAE specification. This might even be envisaged within truly short periods of time, such as the transfer that occurs when one rents a totally different model and generation of vehicle. In this general sense the introduction of more automated vehicles, is no different from offering new iPhones, Tablets, and other advanced forms of computational consumer systems which are designed but never exhaustively tested before deployment. The issue here is that these new technologies control a one-ton vehicle proceeding at 60 mph and glitches, faults, bugs, and errors are not merely frustrating, they can be fatal (and see Templeton, 2020). But is this same process not as true for other, equally safety-critical systems such as advanced fly-by-wire aircraft whose similar failure we have witnessed in recent months? It might be suspected that in rather the same way that an aircraft crash draws widespread news coverage and single vehicle fatalities much less public attention, so the respective flaws in automated ground transportation will draw neither the same level of social disapprobation, nor the same level of regulatory scrutiny as is visited upon commercial aircraft and their functioning. We might well hope this is not the case, but precedent militates against this aspiration. Here, again, time will tell the tale.

At one and the same time that we are about to radically change the nature of physical transportation and its control, we are witnessing a forced proliferation of the "electronic" transportation of information. Here data is, to the greatest degree possible, substituted for material, and the transport of data is so much more easily achieved. Why "port" material goods if, for example, 3-D printers can take remote instruction and create a need item on the spot? The pandemic, which began essentially at the start of 2020, now engulfing our world has forced a re-casting of movement imperatives. Here, much discretionary, and elective travel has been curtailed, or at least stultified (and see Ellwood, 2020). Ways are now being envisaged to extend that propensity to what has previously been viewed as obligatory transportation when, assumedly, the pandemic subsides (cf., Freedman, 2017). At the same time, the whole demographic of the driving public is itself in a state of flux. Today's younger generation no longer see vehicle ownership as obligatory or even perhaps even preferable. Services such as Uber, Lyft, etc., tend to mean that personal travel is an on-demand requirement that does not necessarily entail a vehicle of one's own. Timeshared vehicles and on-demand transport are even more likely to burgeon if pure driverless (and perhaps individually ownerless) vehicles circulate in local environments. If a number of these trends are sustained, and if they are underwritten by greater profit, and there is no reason to believe they will not be, then motivations for goals such as improved roadway capacity may begin to dissipate as the absolute level of vehicle numbers on roadways are themselves reduced. This is another of the promises held out to the public to persuade them toward collective acceptance.

What were once thought to be rather fanciful notions, e.g., the delivery of small packages by purpose-directed drones, now look much more realistic in light of the times in which we live. Rather paradoxically, the collision rates, which assumedly ought, at least to a degree co-vary with the number of vehicles on the road (Hancock, 2013a) does not seem to follow any such simple relationship. Such patterns challenge us to understand what safety levels will be like in an increasingly mixed equipage situation (Sivak and Schoettle, 2015). There might, for example, be the opportunity for low-level flying personal transport to be substituted for on-road vehicles. All this is to say that we are now fully engulfed in perhaps the next great wave of powered transportations evolution. If the first wave was characterized by the replacement of human muscle with animal power, and the second the replacement of animal strength by artificial power, then the third wave is most certainly the one where human movement control is abrogated to a computational surrogate. As each of these steps were accompanied by fairly radical changes in the organization and structure of human society, so we cannot but expect that the latter will exert both anticipated, but also unknown effects in a similar manner. Driving is not merely the act of vehicle control; it is a relationship between a person and the world in which they live (Singer, 2014). For good or bad, we are changing driving's rules and with them the very role of vehicles in society itself (Howey, 2012). The step-change that we are in the middle of will recast the world as we understand it.

A SUMMARY NOT YET CONCLUDED

As we conclude with a vision of a slowly diminishing, then dying, then virtually an extinct activity of driving, we must spare a final thought for those who actually love it as a desired pursuit. Here, we not only include those that actually rely on driving for their profession, such as taxi drivers, truck drivers and the like, but also for those various ancillary enthusiasts. What of those whose profession is a Formula One driver, NASCAR racer, dragster, and those who enjoy watching these forms of vehicle-related entertainment? Will crowds turn out to watch autonomous stock car racing in the same way we now have a niche for "battling" robots? Perhaps, but one senses not. How will Formula One look when an fully designed, tested, and evaluated optimal control algorithm can easily exceed even the greatest human exponent of the art? And what of the simple, plain experience of satisfaction in controlling one's own destiny on the open highway? Will books such as "*Zen and the Art of Motorcycle Maintenance*" (Pirsig, 1974) be inspired by autonomous motorcycles? There is therefore an important hedonomic dimension to driving (Hancock et al., 2005; Hagman, 2010) which extends well beyond the mere utilitarian necessity to relocate persons and material from origin to destination. Of course, these pleasurable and hedonic values have to be set against the traditional concerns in driving for downsides such as pollution, driving's subsequent contribution to global warming, the problems of over-crowding and time-wasting in queues and gridlock etc. However, if driving is the wedge issue that I have identified, then it is possible to postulate that many presently gratifying, associated human activities will be submerged and then extirpated by the insensible tide of spreading autonomy. It will not be too long before our children may ask, in all naivete: "what was a driver?" And after that what? The changes to society will extend well beyond only these vanishing hedonomic dimensions. Most disturbingly, the growth of independent, autonomous vehicles will, without careful political legislation, serve very much to limit human freedom. In the same way that even presently existing video surveillance systems curtail and constrain people's actions, the inability to "*see the USA in your Chevrolet*," is something more than just transferring momentary vehicle control from a human to an AI-based autonomous substitute. Rather, it promises to represent a profound change in the human condition. When driving into the future, we need to be much more wary about the coming roads, whether they be ones either more or less traveled.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Campbell, I. C. (1995). The lateen sail in world history. *J. World History* 6, 1–23. doi: 10.1080/09503110.2012.655580

Crossman, E. R. F. W. (1959). A theory of the acquisition of speed-skill∗. *Ergonomics* 2, 153–166. doi: 10.1080/00140135908930419

Daily, M., Medasani, S., Behringer, R., and Trivedi, M. (2017). Self-driving cars. *Computer* 50, 18–23.

De Winter, J. C. F., and Hancock, P. A. (2015). Reflections on the 1951 Fitts list: do humans believe now that machines surpass them? *Proc. Manuf.* 3, 5334–5341. doi: 10.1016/j.promfg.2015.07.641

Dekker, S. W., and Woods, D. D. (2002). MABA-MABA or abracadabra? Progress on human–automation co-ordination. *Cogn. Technol. Work* 4, 240–244. doi: 10.1007/s101110200022

Desmond, P., Hancock, P. A., and Monette, J. (1998). Fatigue and automation-induced impairments in simulated driving performance. *Transp. Res. Rec.* 1628, 8–14. doi: 10.3141/1628-02

Ellwood, M. (2020). *Coronavirus Air Travel: These Numbers Show the Massive Impact of the Pandemic. Conde Nast (CN) Traveler*. Available online at: https://www.cntraveler.com/story/coronavirus-air-travel-these-numbers-show-the-massive-impact-of-the-pandemic (accessed August 7, 2020).

Francis, E. L., Tyrrell, R. A., and Owens, D. A. (2020). Perception response time and its misapplication: an historical and forensic perspective. *Theor. Issues Ergonom. Sci.* 21, 327–346. doi: 10.1080/1463922x.2020.1736204

Freedman, D. H. (2017). Self-driving trucks. *MIT Technol. Rev.* 120, 62–71.

Fuller, R. (2005). Towards a general theory of driver behaviour. *Accid. Anal. Prev.* 37, 461–472. doi: 10.1016/j.aap.2004.11.003

Gadsden, S. A., and Habibi, S. R. (2009). "The future of automobiles: cognitive cars," in *Proceedings of 22nd Canadian Congress of Applied Mechanics* (CANCAM), Halifax, 111–112.

Hagman, O. (2010). Driving pleasure: a key concept in Swedish car culture. *Mobilities* 5, 25–39. doi: 10.1080/17450100903435037

Hancock, G. M., Longo, L., Young, M., and Hancock, P. A. (2020). "Mental workload," in *Handbook of Human Factors and Ergonomics*, eds W. Karwowski and G. Salvendy (New York, NY: Wiley).

Hancock, P. A. (1997). On the future of work. *Ergon. Design* 5, 25–29.

Hancock, P. A. (2007). On the process of automation transition in multi-task human-machine systems. *Transact. IEEE Syst. Man Cybernetics Part A Hum. Syst.* 37, 586–598. doi: 10.1109/tsmca.2007.897610

Hancock, P. A. (2008). Frederic Bartlett: Through the lens of prediction. *Ergonomics* 51, 30–34. doi: 10.1080/00140130701802692

Hancock, P. A. (2011). *Stress and Workload in Aviation.* Invited Plenary Address to the 5th International Summer School on Aviation Psychology. Graz: International Summer School on Aviation Psychology.

Hancock, P. A. (2013a). "Driven to distraction and back again," in *Driver Distraction and Inattention: Advances in Research and Countermeasures*, eds M. A. Regan, T. Victor, and J. Lee (Chichester: Ashgate), 9–25. doi: 10.1201/9781315578156-2

Hancock, P. A. (2013b). In search of vigilance: the problem of iatrogenically created psychological phenomena. *Am. Psychol.* 68, 97–109. doi: 10.1037/a0030214

Hancock, P. A. (2014). Automation: how much is too much? *Ergonomics* 57, 449–454. doi: 10.1080/00140139.2013.816375

Hancock, P. A. (2015). "Automobility: the coming use of fully automated on-road vehicles," in *Proceedings of the IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support* (COGSIMA), (Piscataway, NJ: IEEE), 137–139.

Hancock, P. A. (2017a). Imposing limits on autonomous systems. *Ergonomics* 60, 284–291. doi: 10.1080/00140139.2016.1190035

Hancock, P. A. (2017b). On the nature of vigilance. *Human Factors* 59, 35–43.

Hancock, P. A. (2018a). *Are Autonomous Cars Really Safer than Human Drivers? The Conversation*. Available online at: https://theconversation.com/are-autonomous-cars-really-safer-than-human-drivers-90202 (accessed February 2, 2018).

Hancock, P. A. (2018b). *Self-driving Cars and Humans Face Inevitable Collisions. The Conversation*. Available online at: https://theconversation.com/self-driving-cars-and-humans-face-inevitable-collisions-93774 (accessed April 24, 2018).

Hancock, P. A. (2019a). Some pitfalls in the promises of automated and autonomous vehicles. *Ergonomics* 62, 479–495. doi: 10.1080/00140139.2018.1498136

Hancock, P. A. (2019b). Some promises in the pitfalls of automated and autonomous vehicles: a reply to commentators. *Ergonomics* 62, 514–520. doi: 10.1080/00140139.2019.1586103

Hancock, P. A. (2019c). The humane use of human beings. *Appl. Ergon.* 79, 91–97. doi: 10.1016/j.apergo.2018.07.009

Hancock, P. A. (2020). Science in court. *Theor. Issues Ergon. Sci.* 21, 266–284.

Hancock, P. A., Billings, D. R., and Oleson, K. E. (2011a). Can you trust your robot? *Ergon. Design* 19, 24–29.

Hancock, P. A., Billings, D. R., Olsen, K., Chen, J. Y. C., de Visser, E. J., and Parasuraman, R. (2011b). A meta-analysis of factors impacting trust in human-robot interaction. *Human Factors* 53, 517–527. doi: 10.1177/0018720811417254

Hancock, P. A., Hancock, G. M., and Warm, J. S. (2009). Individuation: the N=1 revolution. *Theor. Issues Ergon. Sci.* 10, 481–488.

Hancock, P. A., Jagacinski, R., Parasuraman, R., Wickens, C. D., Wilson, G., and Kaber, D. (2013). Human-automation interaction research: past, present, and future. *Ergon. Design* 21, 9–14.

Hancock, P. A., Kajaks, T., Caird, J. K., Chignell, M. H., Mizobuchi, S., Burns, P. C., et al. (2020a). Challenges to human drivers in increasingly automated vehicles. *Human Factors* 62, 310–328. doi: 10.1177/0018720819900402

Hancock, P. A., Lee, J. D., and Senders, J. W. (2020b). *Attribution Error in Humans and Machines*. Submitted.

Hancock, P. A., Lesch, M., and Simmons, L. (2003). The distraction effects of phone use during a crucial driving maneuver. *Accid. Anal. Prev.* 35, 501–514. doi: 10.1016/s0001-4575(02)00028-3

Hancock, P. A., and Matthews, G. (2015). "Stress and attention," in *The Handbook of Attention*, eds J. M. Fawcett, E. Risko, and A. Kingstone (Boston, MA: MIT Press), 547–568.

Hancock, P. A., and Newell, K. M. (1985). "The movement speed-accuracy relationship in space-time. Invited Chapter," in *Motor Behavior: Programming, Control, and Acquisition*, eds H. Heuer, U. Kleinbeck, and K. H. Schmidt (Berlin: Springer), 153–188. doi: 10.1007/978-3-642-69749-4_5

Hancock, P. A., Nourbakhsh, I., and Stewart, J. (2019). On the future of transportation in an era of automated and autonomous vehicles. *Proc. Natl. Acad. Sci. U.S.A.* 116, 7684–7691.

Hancock, P. A., and Parasuraman, R. (1992). Human factors and safety in the deign of intelligent vehicle-highway systems (IVHS). *J. Safety Res.* 23, 181–198. doi: 10.1016/0022-4375(92)90001-p

Hancock, P. A., Pepe, A., and Murphy, L. L. (2005). Hedonomics: the power of positive and pleasurable ergonomics. *Ergon. Design* 13, 8–14. doi: 10.1177/106480460501300104

Hancock, P. A., and Sheridan, T. B. (2011). "The future of driving simulation," in *Handbook of Driving Simulation for Engineering, Medicine and Psychology*, eds D. L. Fisher, M. Rizzo, J. K. Caird, and J. D. Lee (Boca Raton, FL: CRC Press).

Hancock, P. A., and Verwey, W. B. (1997). Fatigue, workload, and adaptive driver systems. *Accid. Anal. Prev.* 29, 495–506. doi: 10.1016/s0001-4575(97)00029-8

Hancock, P. A., and Volante, W. V. (2020). Quantifying the qualities of language. *PLoS One* 15:e0232198. doi: 10.1371/journal.pone.0232198

Hancock, P. A., and Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors* 31, 519–537. doi: 10.1177/001872088903100503

Hancock, P. A., and Weaver, J. L. (2005). On time distortion under stress. *Theor. Issues Ergon. Sci.* 6, 193–211. doi: 10.1080/14639220512331325747

Hancock, P. A., Wulf, G., Thom, D., and Fassnacht, P. (1990). Driver workload during differing driving maneuvers. *Accid. Anal. Prev.* 22, 281–290. doi: 10.1016/0001-4575(90)90019-h

Howey, D. A. (2012). A challenging future for cars. *Nature Clim. Change* 2, 28–29. doi: 10.1038/nclimate1336

Ingram, G. K., and Liu, Z. (1999). *Vehicles, Roads, and Road Use: Alternative Empirical Specifications*. Washington, DC: The World Bank.

Insurance Institute for Highway Safety (2020). Available online at: https://www.iihs.org/topics/fatality-statistics/detail/state-by-state (accessed July 23, 2020).

Kenworthy, J. R. (2006). The eco-city: ten key transport and planning dimensions for sustainable city development. *Environ. Urban.* 18, 67–85. doi: 10.1177/0956247806063947

Knittel, C. R., and Murphy, E. (2019). *Generational Trends in Vehicle Ownership and Use: Are Millennials any Different?* (No. w25674). Cambridge, MA: National Bureau of Economic Research.

Kyriakidis, M., de Winter, J. C. F., Stanton, N., Bellet, T., van Arem, B., Brookhuis, K., et al. (2017). A human factors perspective on automated driving. *Theor. Issues Ergon. Sci.* 20, 223–249. doi: 10.1080/1463922X.2017.1293187

Lay, M. G. (1992). *Ways of the world: A History of the World's Roads and of the Vehicles that Used Them*. New Brunswick, NJ: Rutgers University Press.

Lee, S., Oh, C., and Hong, S. (2018). Exploring lane change safety issues for manually driven vehicles in vehicle platooning environments. *IET Intell. Transp. Syst.* 12, 1142–1147. doi: 10.1049/iet-its.2018.5167

Litman, T. (2017). *Autonomous Vehicle Implementation Predictions*. Victoria: Victoria Transport Policy Institute.

Liu, X. (2010). *The Silk Road in World History*. Oxford: Oxford University Press.

Möser, K. (2002). *Geschichte Des Autos*. Frankfurt: Campus.

Newell, K. M., and Hancock, P. A. (1984). Forgotten moments: skewness and kurtosis are influential factors in inferences extrapolated from response distributions. *J. Motor Behav.* 16, 320–335. doi: 10.1080/00222895.1984.10735324

Oborne, D. J., Leal, F., Saran, R., Shipley, P., and Stewart, T. (eds). (2014). *Person-Centered Ergonomics: a Brantonian View of Human Factors*. Boca Raton, FL: CRC Press.

Parkes, A. M., and Franzen, S. (1993). *Driving Future Vehicles*. Boca Raton, FL: CRC Press.

Pettersson, I., and Karlsson, I. M. (2015). Setting the stage for autonomous cars: a pilot study of future autonomous driving experiences. *IET Intell. Transp. Syst.* 9, 694–701. doi: 10.1049/iet-its.2014.0168

Pirsig, R. (1974). *Zen and the Art of Motorcycle Maintenance*. New York, NY: Morrow.

Rajasekhar, M. V., and Jaswal, A. K. (2015). "Autonomous vehicles: the future of automobiles," in *Proceedings of the IEEE International Transportation Electrification Conference (ITEC)*, (Piscataway, NJ: IEEE), 1–6. doi: 10.1016/b978-0-12-817696-2.00001-9

Revelle, R. (1962). Sailing in new and old oceans. *AIBS Bull.* 12, 45–47. doi: 10.2307/1293007

Scallen, S. F., Hancock, P. A., and Duley, J. A. (1996). Pilot performance and preference for short cycles of automation in adaptive function allocation. *Appl. Ergon.* 26, 397–403. doi: 10.1016/0003-6870(95)00054-2

Shaheen, S. A., and Cohen, A. P. (2013). Carsharing and personal vehicle services: worldwide market developments and emerging trends. *Int. J. Sustain. Transp.* 7, 5–34. doi: 10.1080/15568318.2012.660103

Sheridan, T., Hancock, P. A., Pew, R., Van Cott, H., and Woods, D. (1998). Can the allocation of function between humans and machines ever be done on a rational basis? *Ergon. Design* 6, 20–25.

Sheridan, T. B. (1992). *Telerobotics, Automation, and Human Supervisory Control*. Cambridge: MIT Press.

Sheridan, T. B. (2002). *Humans and Automation: System Design and Research Issues*. Santa Monica, CA: Human Factors and Ergonomics Society.

Sheridan, T. B., and Verplank, W. L. (1978). *Human and Computer Control of Undersea Teleoperators*. Cambridge, MA: Massachusetts Institute of Technology.

Shneiderman, B. (2020). Human-centered artificial intelligence: reliable, safe, and trustworthy. *Int. J. Hum. Comp. Interact.* 36, 495–504. doi: 10.1080/10447318.2020.1741118

Simon, H. A. (1969). *Sciences of the Artificial*. Cambridge, MA: MIT Press.

Singer, A. (2014). *Why We Drive: The Past, Present, and Future of Automobiles in America*. Portland, OR: Microcosm Publishing.

Sivak, M., and Schoettle, B. (2015). *Road Safety with Self-driving Vehicles: GENERAL limitations and Road Sharing with Conventional Vehicles*. Ann Arbor, MI: University of Michigan.

Stanton, N. A., Eriksson, A., Banks, V. A., and Hancock, P. A. (2020). Turing in the driver's seat: can people distinguish between automated and manually driven vehicles? *Hum. Factors Ergon. Manuf. Serv. Industries* (in press).

Stead, D., and Vaddadi, B. (2019). Automated vehicles and how they may affect urban form: a review of recent scenario studies. *Cities* 92, 125–133. doi: 10.1016/j.cities.2019.03.020

Templeton, B. (2020). Tesla in Taiwan crashes directly into overturned truck, ignores pedestrian, with autopilot on. *Forbes*

Yuris, N., Sturman, D., Auton, J., Giacon, L., and Wiggins, M. W. (2019). Higher cue utilisation in driving supports improved driving performance and more effective visual search behaviours. *J. Safety Res.* 71, 59–66. doi: 10.1016/j.jsr.2019.09.008

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 20 October 2020
doi: 10.3389/fpsyg.2020.574262

Searching for Street Parking: Effects on Driver Vehicle Control, Workload, Physiology, and Glances

*Canmanie Teresa Ponnambalam and Birsen Donmez**

Human Factors and Applied Statistics Laboratory, Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON, Canada

Urban areas that allow street parking exhibit a heightened crash risk that is often attributed to factors such as reduced road width, decreased visibility, and interruptions to traffic flow. No previous on-road studies have investigated how the demands of searching for parking affect driving performance, physiology, and visual attention allocation. We are interested in these effects on the driver and their possible influence on the safety of the environment. While simulator studies offer several benefits, the physical, mental and social pressures incurred by searching for parking in an urban streetscape cannot be emulated in a simulator. We conducted an on-road instrumented vehicle study with 28 participants driving in downtown Toronto, Canada to explore the effect of searching for street parking on drivers. During the experiment, participants drove two routes in a counterbalanced order: one route with a parking search task, and the other route as a baseline. Speed and lane position were measured via vehicle instrumentation, heart rate and galvanic skin response were measured through physiological sensors, and gaze position was collected through a head-mounted eye-tracker. Participants completed the NASA Task Load Index after each route. It was found that while searching for parking, participants drove slower and closer to the curb, and perceived higher workload. While there were no statistically significant effects in physiological measures, there was a rise in heart rate approaching statistical significance. A detailed analysis of eye-tracking data revealed a clear change in glance behavior while searching for parking, with an increase in long off-road glances (>2 s) and decrease in shorter off-road glances (<1.6 s). Some exhibited behaviors (e.g., slowing down) may be seen to compensate for the potentially negative effects of increased demands associated with parking search, while others (e.g., increase in long off-road glances) have the potential to increase crash risk. This study acts as an important first step in revealing changes in driving performance, physiology and glance behavior brought on by searching for parking in a real-world urban environment.

Keywords: driver behavior, distraction, on-road study, instrumented vehicle, visual attention, traffic safety, parking

OPEN ACCESS

Edited by:
Oren Musicant,
Ariel University, Israel

Reviewed by:
Guy Cohen-Lazry,
Ben-Gurion University of the Negev, Israel
Frank P. McKenna,
University of Reading, United Kingdom
Bruce Mehler,
Massachusetts Institute of Technology, United States

****Correspondence:***
Birsen Donmez
donmez@mie.utoronto.ca

Specialty section:
This article was submitted to Performance Science, a section of the journal Frontiers in Psychology

Received: *19 June 2020*
Accepted: *09 September 2020*
Published: *20 October 2020*

Citation:
Ponnambalam CT and Donmez B (2020) Searching for Street Parking: Effects on Driver Vehicle Control, Workload, Physiology, and Glances. Front. Psychol. 11:574262. doi: 10.3389/fpsyg.2020.574262

1. INTRODUCTION

The convenience and often limited availability of street parking makes it a coveted resource in many downtown areas. Hampshire and Shoup (2018) highlighted 22 studies in 15 different cities from 1927 to 2015 examining proportions of traffic cruising for parking. The latest (2005–2015) of these studies found an average of just under 40% of traffic cruising for parking across 3 cities. More recent studies found the number of vehicles cruising for parking to be 15% in Stuttgart, Germany (Hampshire and Shoup, 2018), 5–6% in San Francisco, CA, and 3–4% in Ann Arbor, MI (where cruising refers only to *excess* travel due to the search for parking) (Weinberger and Millard-Ball, 2017). Regardless of the variability in statistics, measurement methods and definitions, it is widely regarded that street parking can be difficult to find on demand (especially in urban areas during busy times of the day). At these times drivers may be forced to search for parking on or off their intended route. As presented in detail below, many studies document the heightened crash risk evident in areas that allow street parking. However, to the best of our knowledge, no studies have attempted to measure the effect on drivers of engaging in the parking search task, nor how drivers actively searching for parking while driving affect the safety of the road environment. In this paper, we present an on-road instrumented vehicle study investigating how drivers' vehicle control, workload, physiology, and glances are affected by searching for street parking.

There has been significant research on the safety effects of the presence of street parking. A review of street parking in the U.S. estimated that it was associated with 15% of crashes (Sisiopiku, 2001). Similarly, a 1971 report concluded that street parking was directly or indirectly responsible for 20% of all urban crashes in the U.S. (Highway Research Board, 1971). The report named five primary reasons for why street parking increases crash risk: increased obstacles (i.e., parked vehicles), disruption of traffic flow by cars leaving parking spaces, disruption of traffic flow by cars entering parking spaces, drivers or passengers exiting parked vehicles, and reduced sight distance of pedestrians. Decreased road width and sight restrictions due to parked vehicles have also been cited as an issue (Greibe, 2003; Box and Levinson, 2004; Cao et al., 2017). A study conducted by Edquist et al. (2012) showed that the visual complexity of an urban environment in the presence of parked cars can increase the workload of drivers and influence their driving behavior. In that study, there was little difference between an environment that did not allow street parking and one with empty bays, suggesting that the presence of parked cars was the most significant contributor to workload. Some have argued that searching for parking results in drivers slowing down to safer speeds, reducing crash severity (Lerner-lam et al., 1992; Daisa and Peers, 1997; Marshall et al., 2008), or that parked vehicles can protect pedestrians by separating moving traffic from the sidewalk (Lerner-lam et al., 1992). Despite these arguments and supposed safety benefits of street parking, crash risk appears to be elevated in areas that allow it.

While no studies were found that assessed the task of searching for parking, tasks that generally visually engage drivers have been shown to affect many measures of driving, such as lane position and lane position variability, speed and speed variability, reaction time to external events, and subjective workload (Regan et al., 2008). The tasks most commonly studied are voluntary in-vehicle tasks (e.g., mobile phone use) that often have a manual component in addition to visual. Few studies were found that examine visual secondary tasks without a manual component or that concern distractions outside of the vehicle. A recent analysis of the largest naturalistic driving study to date found that some observable type of distraction was involved in 68% of crashes and that extended glances to external objects were associated with a crash risk 7.1 times that of normal driving (Dingus et al., 2016). Visual search is the premier component of searching for parking and requires drivers to scan the environment on the side of the road to locate and confirm vacant spots in tandem with reading posted parking restrictions and road markings. The existence of parked cars provides an obstacle to getting within reading distance of roadside signage and contributes to the complexity of the road environment. While searching for parking, it is expected that drivers spend more time glancing off-road and exhibit an increased number of off-road glances, an effect that we aim to verify and quantify in this study. Regarding driving behavior, the addition of a visual task has been shown to result in reduced speeds and increased lane keeping variability (Dingus et al., 1995; Engström et al., 2005; Zhang et al., 2006). The same behavior was found when drivers drove in an environment with higher visual complexity (Edquist et al., 2012). Similar results are anticipated for speed and lane keeping variability when drivers are tasked with searching for parking. In addition, because sign reading is assumed to be a significant aspect of finding street parking, it is possible that drivers drive closer to the curb when searching to allow them to read posted parking restrictions, particularly due to the potentially small letter sizing (see: example in **Figure 4**). In order to quantify how drivers are affected while searching for parking, both vehicle control and glance behavior require investigation.

In addition to visual demand, searching for parking may increase cognitive demand and stress. Drivers searching for parking in an urban center could be further burdened with the task of navigating while searching for parking. In addition, the time drivers spend locating a parking space has been shown to be a major influencing factor in choosing a parking spot (Brooke et al., 2014). Time spent looking for parking is time removed from the driver's ultimate destination, making the search for parking a task best done as quickly as possible. Time urgency has been shown to relate to driver stress and affect driving behavior (Hennessy and Wiesenthal, 1999). In addition, as drivers search and obstruct traffic in busy areas, they are likely to find themselves under the pressure of following vehicles, who may honk or keep close distances. This social pressure can further contribute to the stress of the driver and pressure them to maintain a speed that makes the parking search more difficult. These cognitive load and stress effects can be assessed through various measures including self-reports (Hart, 2006) as well as heart rate and skin conductance, which are known to rise under increased stress and cognitive load (Healey and Picard, 2005; Mehler et al., 2012). These measures have been used in other on-road studies as quantifiers of stress levels in drivers; such a study found an increase in heart rate, indicating a rise in stress level, exhibited by drivers when parallel parking manually compared to parking with assistive technology (Reimer et al., 2016).

From existing research, it is unclear what (if any) the effects of searching for parking while driving are on drivers and, in turn, the road environment. To investigate this, we conducted an on-road instrumented vehicle study in downtown Toronto, Canada, to explore how drivers' vehicle control, perceived workload, physiology, and glance behaviors change while searching for parking in a busy urban area. As an inaugural step into investigating the parking task, we focus only on the search itself and not the task of parking the vehicle. To the best of our knowledge, no other research has investigated the effects of searching for parking at the driver level from any of the perspectives of vehicle control, perceived workload, physiology, and visual attention allocation.

2. EXPERIMENTAL METHOD

To study the effects of searching for parking at the driver level, an experiment must adequately simulate the parking search task in a controlled manner and allow for relevant measures to be recorded under representative driving scenarios. Despite the limitations of on-road studies in regards to experimental control, driving simulators have other limitations that render them less effective to study parking search. For example, it is hypothesized that the social pressure of blocking traffic is a contributor to the demands on the driver while searching for parking; this pressure cannot be induced in a simulator. In general, the perception and influence of risk is limited in a simulated environment. In addition, sign reading and visual scanning are key components of searching for parking, and simulators are limited in the resolution and visual detail they can provide, making them less effective in studies that focus on visual scanning (Kaptein et al., 1996). We therefore chose to conduct our study on the road in an instrumented vehicle. The study was approved by the University of Toronto Research Ethics Board (protocol number 32795).

As this experiment was the first to examine the effects of searching for parking, we preferred to focus on roads with attributes that pose the highest demands on drivers: complex visual environment, erratic traffic flow, and high occupancy of pedestrians and cyclists. The busy environment also ensures that others (i.e., drivers in following vehicles) are affected by changes in driving behavior, such as potential reductions of speed. This maintains the social pressure expected when searching for parking while driving. However, as this is a first step, we chose not to explicitly investigate this influence nor the influence of time pressure, though both are expected to play a role in the searching for parking task. A simulator study by Edquist et al. (2012), though they did not study the parking search itself, showed that areas with many empty parking bays did not create as high a visual demand on drivers as when there were many parked cars (90% of bays occupied), thus we also conducted the study on roads with a high occupancy of parked cars. There was one (within-subject) independent variable in this study with two levels: driving with a parking search task and driving with no-task (baseline). Participants completed two 15- to 20-min routes under the parking-search and baseline conditions. While both routes were selected to be similar in length and complexity, the order of the routes and the conditions were fully counterbalanced across the participants to remove the potential effects of route and order confounds. During the analysis, it was validated that route and order did not significantly affect the results. The experiment was run between July 2017 and October 2017, on Saturdays or Sundays, starting at either 10:30 a.m. or 1:30 p.m. Running experiments during the summer and only on weekends offered some level of control over the weather and traffic density as well as the number of pedestrians in the area, and ensured that there would be no road work or waste collection interruptions during the experiment.

2.1. Participants

Participants were recruited via posters placed around the university campus and on online forums. Due to insurance and Research Ethics Board constraints, participants were required to be between the ages of 35 and 54 and have a full driver's license for at least 3 years. Therefore, our sample represented a low-crash risk group (Cooper, 1990; McGwin and Brown, 1999). Participants could not wear glasses during the experiment as this affected the quality of data gathered by the head-mounted eye tracker. Therefore, only drivers who can legally drive without glasses (contacts were allowed) could participate in the study. Twenty-eight participants (14 male and 14 female, mean age 41.9, st. dev. of age 5.7) completed the experiment, however due to equipment malfunctions not all participants had full sets of data (discussed in more detail in the section 4). Of those that answered (23 participants), 80% reported that they drive a vehicle a few days a week or more. When asked how frequently they drive in the downtown location of the study, 38% of participants reported a few days a week or more, 44% reported a few days a month, and 17% a few days a year. Participants were compensated at CAN $15/h.

2.2. Apparatus

The instrumented vehicle was a 2014 Toyota RAV4 equipped with a MobilEye device to sample data from the Controller Area Network (CAN bus) connection. The MobilEye also provided measures calculated through image processing techniques applied to video from its internal camera (such as lane position). Another camera mounted on the dashboard provided video of the front-view of the vehicle. Both the MobilEye and front-facing camera can be seen in **Figure 1**.

FIGURE 1 | MobilEye (black unit) and front-facing camera mounted on windshield.

Electrocardiogram (ECG) and galvanic skin response (GSR) sensors produced by Becker Meditec were used to measure heart rate and skin conductance, and recorded data at 240 Hz; three electrodes were placed on the chest to read the ECG signal and two electrodes were placed on the bottom of the left foot to obtain the GSR signal. Both the hand and foot are popular placements for GSR measurement in driving studies; one study investigated both placements and found both to be feasible (Avcı et al., 2014). We opted for the foot placement as we presumed the wires would then be less disturbing to the driver while driving. Gaze position was captured using the head-mounted Dikablis Eye-Tracking Glasses (**Figure 2**), produced by Ergoneers. When calibrated, this device uses two cameras pointed toward the eyes to determine gaze position (tracked at 50 Hz) and overlays the gaze position on video data captured by its front-view camera.

The data from all devices was synced with vehicle data during data collection. A computer and monitor in the back seat allowed for real-time monitoring of data (**Figure 3**).

FIGURE 2 | Driver outfitted with Dikablis eye-tracking glasses.

FIGURE 3 | Data collection computer and monitor in the backseat of the instrumented vehicle.

2.3. Parking Search Task

The task was designed to induce only the loads of visually searching for parking while driving rather than the task of parking itself. Participants were asked to identify legal, vacant parking spaces that were in their direction of travel and on the same street they were on. In the parking-search condition, there were four predetermined sections of each route where participants were asked to search for parking continuously (these sections ranged from approximately 400–800 m). They were told to verbally announce each space they encountered that they understood to be vacant and legal. After each announcement, they were told whether the parking space was indeed legal or why it was not. They then continued to search for parking. Participants were not asked to stop nor park. They were given turn-by-turn directions during the experiment (in both conditions) to eliminate the navigation component of the parking search; although not investigated in our experiment, this component is expected to further distress drivers in real-life scenarios.

2.4. Procedure

The experiment contained both an off-road and an on-road component and took an average of 2 h. Participants first read and signed the informed consent form. They provided their driver's license to verify their age and license type and were aware that a scanned record was made for insurance purposes. After signing the informed consent document, participants filled out questionnaires to gain insights into their driving behavior and history. Participants were then given an instructional booklet that provided a brief overview of parking rules in Toronto and explained the restrictions described by parking signage found along the experimental driving routes (**Figure 4**). After going through the booklet on their own, participants were administered a five-question quiz and were assured that their performance on the quiz would not affect their participation in the rest of the experiment. The multiple-choice quiz tested their understanding of parking signage; after each question, if they chose an incorrect response, the investigator discussed the correct answer with the participants. The purpose of the booklet and quiz was to ensure that all participants had the same minimum level of exposure and understanding of the parking restrictions and signs in the area.

Participants were then taken to the instrumented vehicle and seated in the driver seat. They were given time to adjust the seat position and mirrors. The lead investigator sat in the passenger seat and a research assistant was seated behind the investigator, operating the data collection computer. Participants first completed a 5- to 10-min familiarity drive to allow them to get used to the vehicle. They were told that the investigator would provide them with turn-by-turn directions and that they should ask questions about operating the vehicle during the familiarity drive as talking during the experiment would be discouraged. The familiarity drive was on roads similar to the experimental routes. After the familiarity drive, the participants were outfitted with the physiological sensors and the head-mounted eye-tracking device, which was calibrated before each experimental route. They then completed the two experimental routes, one with the searching for parking task and the other serving as the baseline.

FIGURE 4 | Example of sign explanation in parking instructions booklet provided to the participants.

As previously stated, the order of the routes and the conditions were fully counterbalanced across the participants to remove the potential route and order confounds. After each experimental route, participants completed the NASA Task Load Index (Hart and Staveland, 1988). As part of the questionnaire, they were required to complete a pairwise comparison of six types of workload based on which they felt contributed more to their workload. This calibration was done only once per participant after the first drive. For both drives, drivers rated the extent to which they felt six different types of demand (e.g., mental, physical) during the drive. After the second route, participants exchanged seats with the investigator and were driven back to campus and given their compensation.

3. ANALYSIS

The experiment produced vehicle, physiological, eye-tracking, and subjective data from 28 participants each driving two 15- to 20-min routes under the parking-search and baseline conditions. While both routes were similar in length and complexity, they contained a variety of different streets and intersections. To mitigate these differences, each route included the same 540 m stretch of Bloor St. driven west to east by all participants, once under the baseline condition and once while searching for parking. Participants had already completed at least two sections of parking search before reaching the region, regardless of whether it was their first or second drive. For these reasons, this region was the focus of analysis for vehicle control, physiology, and glances. However, perceived workload was assessed at the end of each route, and therefore its analysis did not focus on the Bloor St. stretch. The Bloor St. stretch contains a single lane in each direction, each with a separated bike lane (**Figure 5A**). There is paid street parking allowed at parking bays indicated by pavement markings, signs, and bollards; parking on the right side of the street was observed to be almost fully-occupied at the times when the experiment took place (by reviewing videos post-experiment), with 3 or 4 spaces free out of 25 on average in the region.

3.1. Vehicle Control

We captured vehicle control through speed and lane position as well as their variability, measured as standard deviation (within participants) and all provided by the MobilEye device. Traffic flow, signal status, and pedestrian behavior could not be controlled during the experiment, thus there were many instances where participants were forced to stop or slow down, regardless of the speed they would normally choose. It was observed that on the Bloor St. stretch, speeds under 15 km/h were driven when participants were either slowing or stopping due to interruptions in the road, such as a red light, vehicles parking, pedestrians crossing, or congested traffic. Therefore, when calculating average speed and standard deviation of speed, only data recorded for speeds above 15 km/h was considered; drivers drove above 15 km/h 68.8% of the time (on average) when driving the Bloor St. stretch.

For all other vehicle measures, the entire set of data from the stretch was used. Lane position was recorded via the Distance to the Left Lane value provided by the MobilEye System (located 6 cm right of the center of the front windshield, **Figure 1**); this is calculated by the device using lane marking detection on video captured by its internal camera. Vehicle measures between the two experimental conditions were compared by paired *t*-tests. Time spent driving the Bloor St. stretch and time spent driving the stretch above 15 km/h were also compared between conditions using paired *t*-tests and applied as offset variables where appropriate in the statistical analysis of glance metrics.

3.2. Subjective Workload

The NASA TLX was administered after both experimental drives, with participants completing the pairwise comparison section only after the first drive. This pairwise comparison of six workload types produced a weighting for each participant, 5 being the type of workload they felt most contributed to their drive and 0 being the least. An average workload score (from 0 to

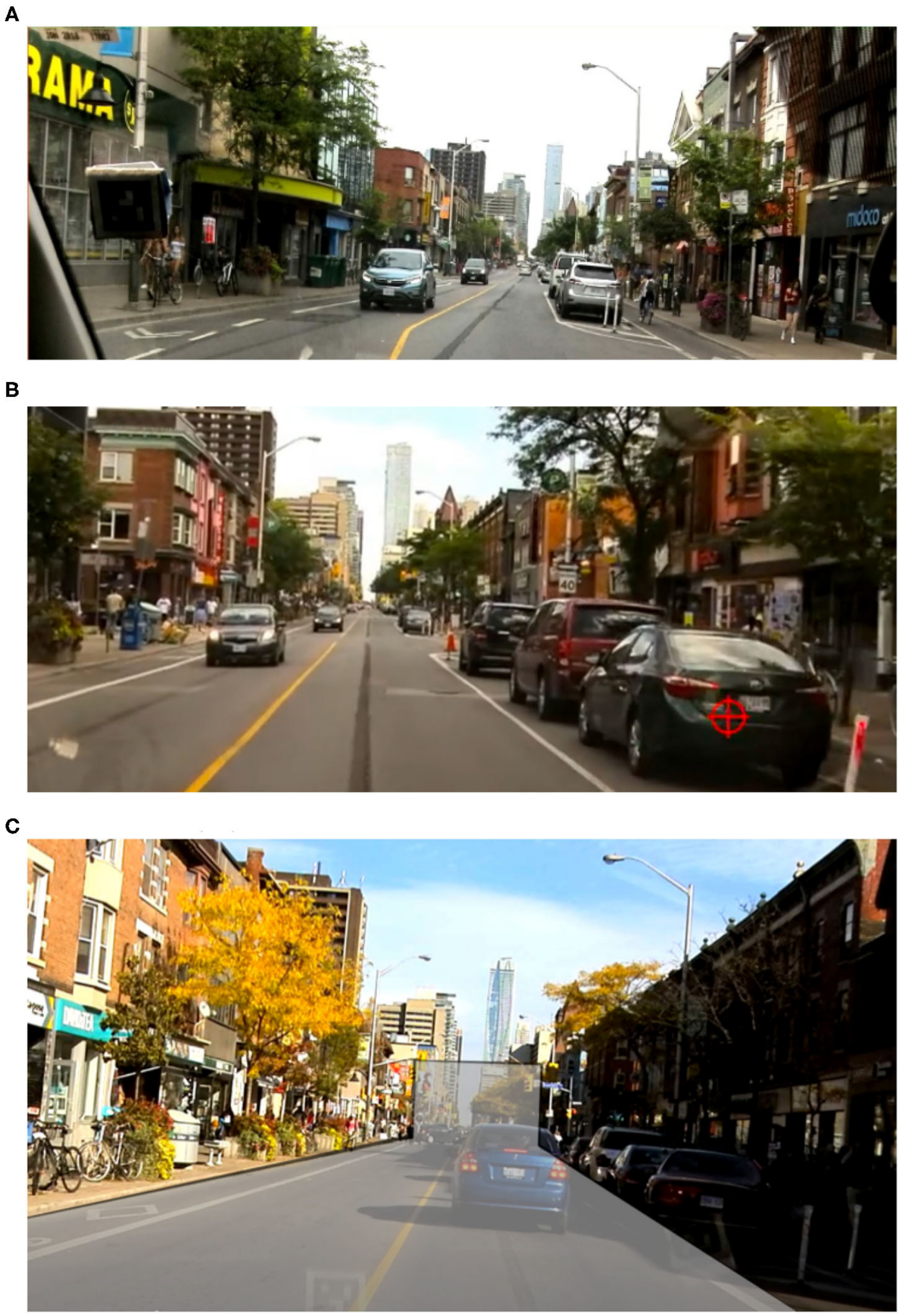


FIGURE 5 | Views from the head-mounted camera on the Dikablis eye-tracker used for glance analysis. **(A)** Snapshot of Bloor St. **(B)** Gaze position indicated by the red cross-hair. **(C)** Bounded region considered "on-road".

20) was calculated for each condition with these weights and the participant ratings (from 0 to 20) of the amount of each type of workload they experienced. The overall self-reported workload, and its components, were compared between the two routes via paired t-tests.

3.3. Physiological Measures

Physiological measures included average heart rate (calculated from EKG signals) and average galvanic skin conductance. Paired t-tests were carried out to compare the two experimental conditions with regard to physiological measures.

3.4. Glance Measures

Glances were coded by reviewing eye tracking videos (**Figure 5B**) and determining the periods when the driver had an off-road glance, i.e., was looking outside of the perimeter deemed on-road (**Figure 5C**). The length of a glance included both the fixation on the area of interest as well as the saccade to the area before the fixation, as defined by the International Organization for Standardization (ISO 15007-1). Glances less than 100 ms were removed from analysis as they may not represent meaningful fixations (Crundall and Underwood, 2011). In addition, because our interest was on searching for parking while driving, glances during periods where drivers were slowing to a stop, stopped (generally at a red traffic signal or to allow pedestrians to cross the street, not because they were searching for a parking spot), or following a very slow-moving vehicle were not of interest. Drivers tended to scan the environment far more during these periods, greatly skewing results. Therefore, glances were filtered to include only those made while the vehicle was moving above 15 km/h.

Based on an on-road study using an eye-tracker, it was reported that drivers rarely glance off the road for longer than 1.6 s (Sodhi et al., 2002); in addition, through a naturalistic driving study, it has been shown that glances off the forward roadway of over 2 s double the risk of a crash (Klauer et al., 2006). These two thresholds (1.6 and 2 s) are used widely in the study of driver distraction (e.g., Sodhi et al., 2002; Horrey and Wickens, 2007; Hallihan et al., 2011; Reimer et al., 2014). Thus, we also used these thresholds in our analysis. It should be noted that Klauer et al. (2006) utilized video recordings of the participants' face to assess gaze direction. Thus, their method is likely not as precise as our study's assessment of glance duration (hence the label "off the forward roadway" as opposed to "off-road"); however, the naturalistic nature of the study entails a high level of ecological validity.

Our glance measures included percentage of time looking away from the road, average off-road glance duration, rate of off-road glances per minute, and rate of shorter ($<$1.6 s) and long ($>$2 s) off-road glances per minute. Percent time and glance duration measures were analyzed with paired t-tests. Number of glances data were non-normal, thus were modeled through generalized linear models with the Poisson distribution and log link function, and with task condition (baseline or parking-search) as the predictor variable. The time spent above 15 km/h in minutes was used as an offset variable; therefore, the models predicted rate of glances (/min). Repeated measures were accounted for using generalized estimating equations.

4. RESULTS

As mentioned earlier, some participants had to be dropped from analysis of some measures due to equipment malfunctions. **Table 1** summarizes the number of participants whose data were analyzed for each measure, as well as their gender and age information.

TABLE 1 | Number of participants with usable data for each measure.

Measure	*N*	Age mean, standard deviation
NASA TLX	14 male, 14 female	41.9, 5.7
Driving duration, speed	13 male, 13 female	42.2, 5.7
Lane position	12 male, 13 female	42.5, 5.7
Galvanic skin response	12 male, 12 female	42.4, 5.9
Heart rate	11 male, 12 female	42.6, 6.0
Glance measures	9 male, 7 female	42.8, 5.7

4.1. Vehicle Control

The time spent driving the Bloor St. stretch was not expected to significantly differ between conditions, as the stretch contained 2 traffic signals and various traffic conditions that greatly affect this measure regardless of whether the searching task was performed. Indeed, there was no significant difference; it took participants an average of 77.2 s to complete the stretch while searching for parking (SD = 25.2 s) and 73.0 s in the baseline (SD = 17.1 s); $t_{(25)} = 0.66$, $p = 0.51$. When considering only speeds above 15 km/h (**Figure 6**), the average speed was found to be significantly higher in the baseline condition (M = 28.8 km/h, SD = 4.0 km/h) than when drivers were searching for parking (M = 26.3 km/h, SD = 3.2 km/h), $t_{(25)} = 2.34$, $p = 0.03$, Cohen's d = 0.68. The standard deviation of speed above 15 km/h was also, on average, significantly higher in the baseline (M = 5.5 km/h, SD = 1.8 km/h) than in the parking-search condition (M = 4.4 km/h, SD = 1.0 km/h), $t_{(25)} = 2.27$, $p = 0.03$, Cohen's d = 0.67. The average distance to the left lane significantly differed between conditions (**Figure 6**); drivers drove further from the left lane when searching for parking (M = 1.62 m, SD = 0.13 m) than in the baseline (M = 1.57 m, SD = 0.13 m), $t_{(24)} = 2.11$, $p = 0.045$, Cohen's d = 0.37. The standard deviation of distance to the left lane did not differ significantly when participants searched for parking (M = 0.33 m, SD = 0.09 m) compared to their baseline (M = 0.37 m, SD = 0.14 m), $t_{(24)} = 1.54$, $p = 0.14$.

4.2. Subjective Workload

The overall NASA TLX score was significantly higher during the parking-search route (M = 54.63, SD = 14.53) than in the baseline route (M = 39.02, SD = 11.27), $t_{(27)} = 5.36$, $p < 0.001$, Cohen's d = 1.19. **Figure 7** displays the raw (un-adjusted by weighting, for comparison) rating for each specific type of workload by task conditions. These box plots as well as the ones presented later depict the minimum, maximum, 1st and 3rd quartiles, and the median, as well as the mean overlaid on the box as triangles along with its value. The difference between parking-search and baseline routes for the individual workload components were all significant: physical [1.86; $t_{(27)} = 2.6$, $p = 0.02$, Cohen's d = 0.43], mental [3.28; $t_{(26)} = 3.7$, $p = 0.001$, Cohen's d = 0.84], temporal [4.78; $t_{(27)} = 5.5$, $p < 0.001$, Cohen's d = 1.03], performance [1.9; $t_{(27)} = 4.4$, $p < 0.001$, Cohen's d = 0.69], effort [3.46; $t_{(27)} = 3.4$, $p = 0.002$, Cohen's d = 0.87] and frustration [2.40; $t_{(27)} = 2.5$, $p = 0.02$, Cohen's d = 0.51].

4.3. Physiological Measures

The average skin conductance did not differ significantly between conditions, $t_{(23)} = 1.51$, $p = 0.14$. The difference in average heart rate approached significance, with participants exhibiting

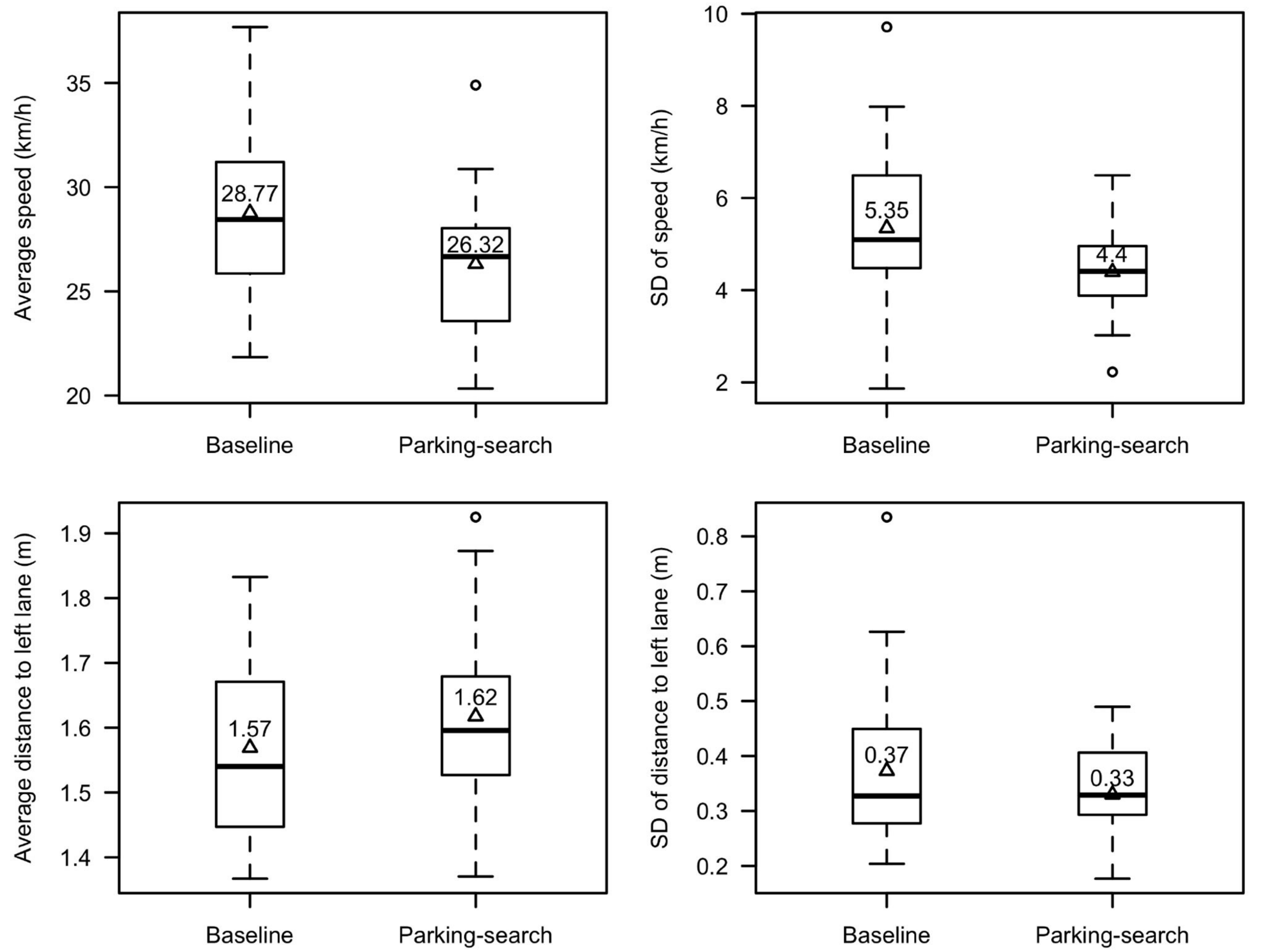


FIGURE 6 | Average and standard deviation of speed and lane position. The mean is reported in the boxplot and indicated with a triangle.

a potential increase when searching for parking (M = 78.99 beats/minute (bpm), SD = 12.34 bpm) over the baseline (M = 77.89 bpm, SD = 12.55 bpm), $t_{(22)} = 1.76$, $p = 0.09$.

4.4. Glance Measures

As shown in **Figure 8**, participants spent more time looking off-road (driving over 15km/h) when searching for parking (M = 53%, SD = 17%) than in the baseline condition (M = 39%, SD = 15%), $t_{(15)} = 2.80$, $p = 0.01$, Cohen's d = 0.70. Further, considering only when driving over 15 km/h, participants had longer off-road glances while searching for parking (M = 1.1 s, SD = 0.4 s) than in the baseline (M = 0.7 s, SD = 0.2 s); $t_{(15)} = 3.5$, $p = 0.003$, Cohen's d = 0.87. Rate of all off-road glances was not significant, $\chi^2(1) = 1.36$, $p = 0.24$; however, rate of off-road glances under 1.6 s, $\chi^2(1) = 10.94$, $p < 0.001$, and the rate of off-road glances over 2 s, $\chi^2(1) = 22.11$, $p < 0.001$, were significant. The rate of off-road glances under 1.6 s was 28% higher in the baseline condition compared to the parking-search condition, 95% CI = (9, 47%), whereas the rate of off-road glances over 2 s was 235% higher in the parking-search condition than in the baseline, 95% CI = (81, 520%). The plots for significant findings in glance rates are found in **Figure 9**.

5. DISCUSSION

This is the first known study that attempts to quantify the effect searching for parking has on drivers through an on-road experiment. We aimed this study to act as a first step into understanding how the necessity for searching for street parking affects the safety of the road environment. The use of an instrumented vehicle and head-mounted eye tracker allowed for relatively precise data collection in a real-world environment, compared to simulator and naturalistic studies. The parking-search task designed for this experiment was a simplified version of the search for parking drivers normally experience. Participants were not required to navigate, did not have any time pressures enforced on them, and were only required to search for parking that was in the same direction and on the same street as they were going. We found evidence that searching for

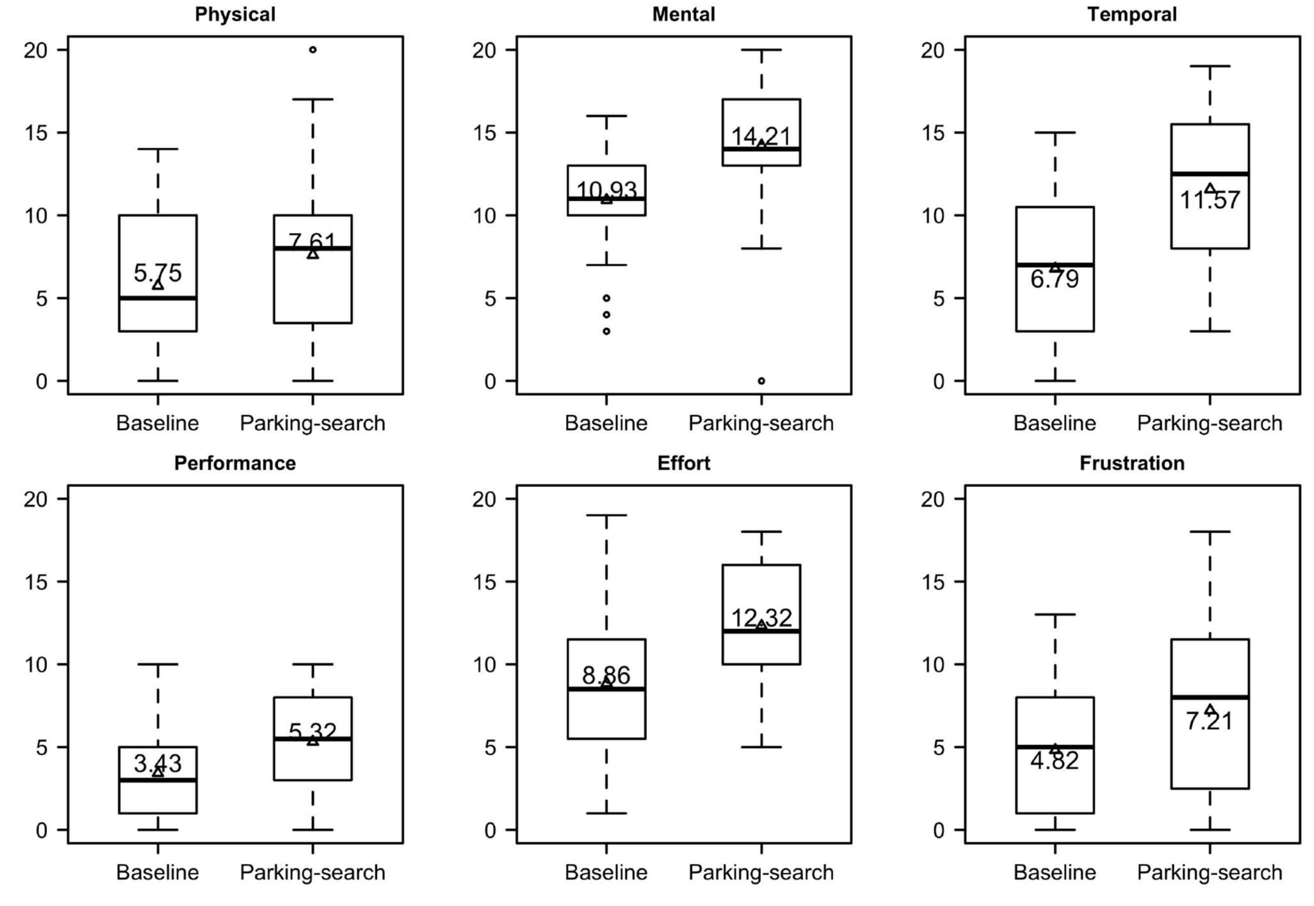


FIGURE 7 | NASA TLX un-weighted ratings for each workload category and task condition.

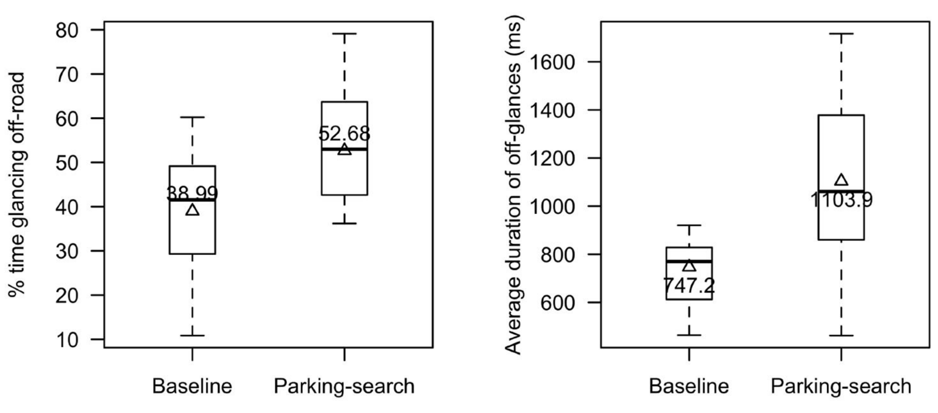


FIGURE 8 | Percent time glancing off-road and average duration of off-road glances.

parking has a measurable effect on drivers, particularly on their perceived workload, vehicle speed and lane position and glance behavior. Drivers reported an increase in workload and were found to drive slower and closer to the curb when searching for parking. They also exhibited longer off-road glances and more frequent long off-road glances. Given the simplification of the parking-search task in this experiment, it is expected that drivers in a similarly complex environment are affected even more so in natural conditions.

Under the condition of searching for parking in which they experience increased perceived workload, participants drove slower on average. Lowering speed is often seen as a

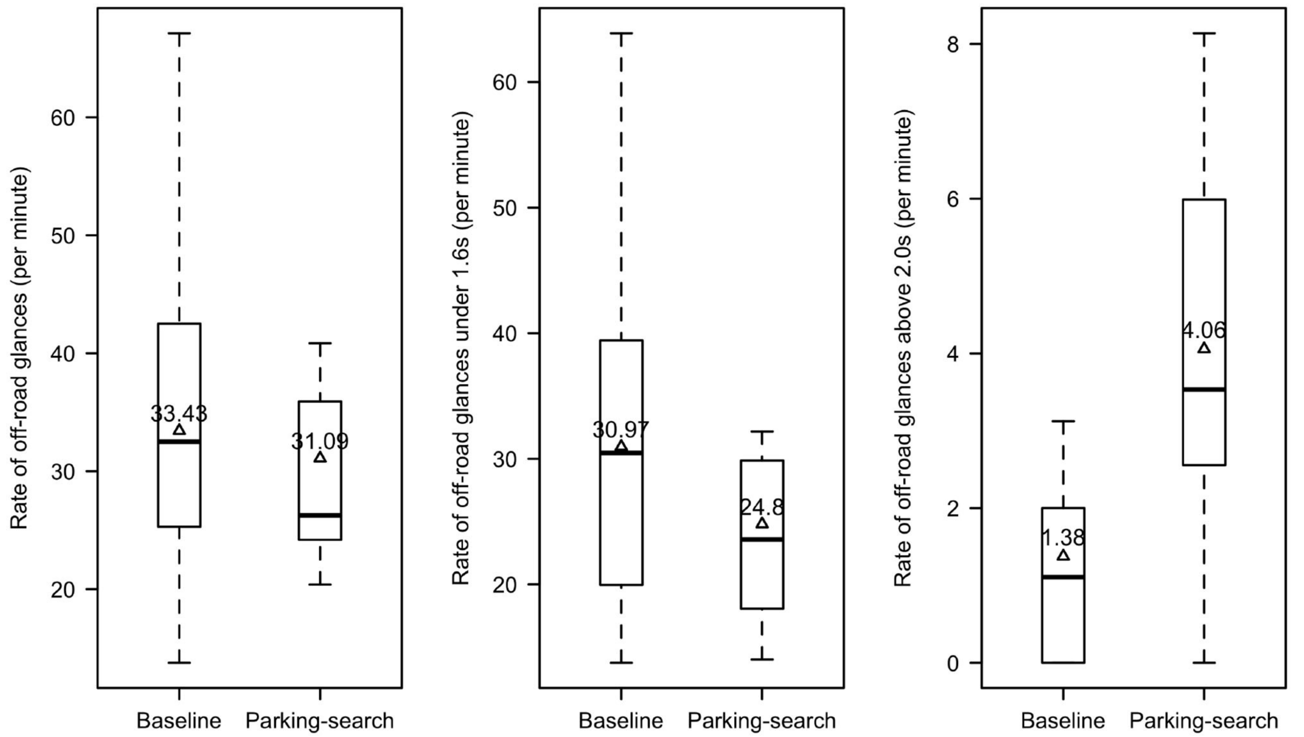


FIGURE 9 | Rates of off-road glances (total, under 1.6 s, above 2 s) per minute.

compensatory strategy that has been observed when drivers experience high visual workload (Engström et al., 2005). A decrease in the standard deviation of average speed was also observed, however the lower speed variability may be a statistical artifact of the generally lower speeds exhibited when searching for parking. The Edquist et al. (2012) simulator study found that drivers drove further from the curb in similar conditions to the Bloor St. stretch (i.e., urban environment, single lane in each direction, fully occupied parking bays along both sides) analyzed in our study; it was suggested that this may be dangerous as vehicles were positioned closer to oncoming traffic. Worth noting is that the influence of this potential danger is likely minimized when driving in a simulator study, possibly leading to riskier behavior than found in a real environment. Our analysis revealed that when tasked with searching for parking in similarly visually complex conditions, participants drove closer to the curb. Drivers may have purposefully kept themselves farther from oncoming traffic while they engaged in a potentially distracting task. However, it is also possible that they drifted nearer to the parking bays as they visually inspected them for vacancy, or to better read parking signs with small letter sizing. Interestingly, it was observed that participants received many of their cues regarding the legality of parking spaces by the presence of other parked cars, rather than by reading the parking signage thoroughly. For this reason, it was difficult to investigate how their understanding of the parking rules may have affected the task, and modifications to the experimental methodology would be needed to do so. In addition, further research is needed to comment on the role reading signs plays in the search for parking. There was no increase in lane keeping variability observed, despite the hypothesis that it would increase under heightened visual load as reported in another study (Engström et al., 2005). The lack of significance in our study may be due to the generally low speeds of the driving area which allowed participants to maintain their course with minimal deviation; it is suggested that further research be done in an area where speeds average above 40 km/h. A lack of statistical power may also explain these and other non-significant findings.

Participants self-reported a clear increase in workload between driving under the baseline condition and driving when periodically searching for parking. While not explicitly studied, we found evidence that time pressure could be induced by the social aspects of driving (i.e., slowing traffic) and not only by the driver's own motivation to complete the task as quickly as possible. The results of the NASA TLX questionnaire revealed that the largest average difference in demand reported was in temporal demand; this indicates that drivers did feel time pressure when searching for parking even though there was no deadline to reach a destination. It seems then that the rate at which they performed the task was at least partially imposed on them by external pressures, social or otherwise.

This study revealed that searching for parking brought on measurable differences in glance behavior. Participants exhibited fewer off-road glances under 1.6 s but more glances over 1.6 s (as suggested by the minimal change in overall rate of glances) when searching for parking compared to the baseline. This suggests

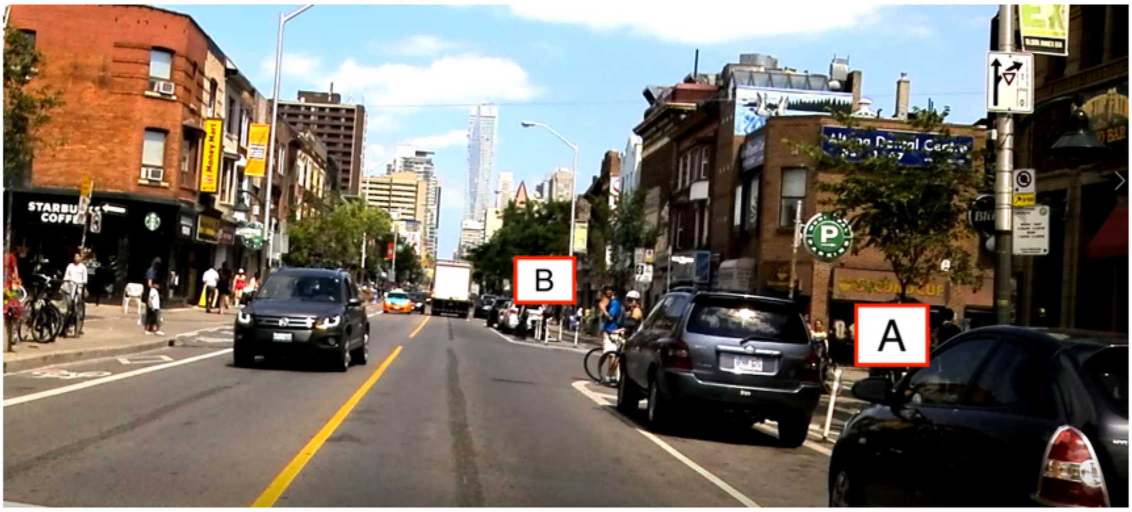


FIGURE 10 | Fixation on point B allows more of the road ahead to be maintained in the driver's central vision than fixation on point A.

an adjustment of visual scanning behavior when searching for parking by lessening the number of short glances they perform to allow more long off-road glances. Long off-road glances can reduce the driver's ability to respond to unexpected events on the road (Liang et al., 2012). Drivers exhibiting more frequent long off-road glances may contribute to a more dangerous driving environment. Even though participants were found to decrease the number of short off-road glances when searching for parking, the total percentage of off-road glance duration was still higher in this condition compared to the baseline. It is important to note that not all off-road glances are equivalent. Glances made off-road far ahead of the vehicle (**Figure 10**, point B) still allow the driver to maintain the environment ahead in their field of view, while fixations on points closer to the vehicle with a higher angular velocity (**Figure 10**, point A) have a reduced portion of the road ahead held in their view. Further analysis on the angular velocity of fixation points is needed to determine how unsafe long off-road glances are. In addition, some off-road glances are necessary to ensure a safe environment, such as glances to a pedestrian about to cross the street.

Physiological signals (heart rate and galvanic skin response, GSR) were expected to reflect an increase in workload. Although heart rate variability is another measure of workload, it was not analyzed given that our study did not provide the recommended 5-min minimum of baseline signal to properly assess any change in HRV between task conditions (Shaffer and Ginsberg, 2017). However, average heart rate showed only a slight increase approaching significance when drivers were searching for parking, and average GSR did not show any significant difference between task conditions. The lack of significance may again be due to a lack of power resulting from our limited sample size or from the variability introduced from the driving environment encompassing uncontrolled factors (e.g., pedestrian jay-walking, traffic signal status, behavior of other traffic) that may have impacted the driver's physiological state more than the searching for parking task. Another possible factor, given that participants did self-report a clear increase in workload, is that increased sensory information taken in when searching for parking caused a decrease in heart rate counter to the rise experienced due to stress. This phenomenon, known as "*sensory intake*," has been suggested to occur when drivers are intently focused on absorbing sensory information (e.g. visually searching for an open parking space) (Mehler et al., 2008). The GSR sensors used in the study could also be unreliable due to the noise in the signal brought on by the vibrations in the vehicle and the movement of the participant; a more robust placement than the bottom of the foot may have achieved better results.

Our work adds to the growing body of on-road experimental studies that aim to quantify driver behaviors in real-world environments. We found that, when searching for parking, drivers exhibited some compensatory behaviors which are conducive to a safer driving environment, such as reduced speed. They also exhibited behaviors which can be considered unsafe, such as increased off-road glances over 2 s. It is recognized that, though statistically significant, the differences in speed and lane position between task conditions are relatively small. Further investigation in different types of road environments is needed to conclude whether such differences can contribute to the heightened crash risk that has been exhibited in areas that allow street parking. This work serves as an important initial step in investigating how searching for parking affects drivers, and acts as an invitation to continue researching a common, often taxing task for drivers that may contribute to crash risk in busy urban areas. Such findings would justify the development of measures, such as changes in road design or parking search assistance via mobile applications, that provide additional benefits to drivers beyond reducing traffic congestion and parking payment efficiency.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the raw data contains confidential information as defined in the agreement with the University of Toronto Research Ethics Board. Requests to access the datasets should be directed to Birsen Donmez.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Toronto Research Ethics Board. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

CP was the graduate student who designed this experiment, collected, and analyzed the data under the supervision of and with feedback from BD. CP drafted this manuscript with feedback from and revisions by BD. All authors contributed to the article and approved the submitted version.

FUNDING

Funding for this research was provided by the Natural Sciences and Engineering Research Council (NSERC) of Canada; NSERC Discovery Grant: RGPIN-2016-05580.

ACKNOWLEDGMENTS

We acknowledge Dr. Justin Hollands and Dr. Mark Chignell for their valuable feedback regarding this work. We also thank Ryan Cheng and Elzat Yilizhati for their support in running the experiments. Preliminary findings from this study were previously published in conference proceedings (Ponnambalam et al., 2018).

REFERENCES

Avcı, C., Akbaş, A., and Yüksel, Y. (2014). "Evaluation of statistical metrics by using physiological data to identify the stress level of drivers," in *Proceedings of the 3rd International Conference on Environment, Chemistry and Biology IPCBEE*, Vol. 78 (Singapore: IACSIT Press), 124–128. doi: 10.7763/IPCBEE

Box, P. C., and Levinson, H. S. (2004). Curb-parking problems: overview. *J. Transport. Eng*. 130, 1–5. doi: 10.1061/(ASCE)0733-947X(2004)130:1(1)

Brooke, S., Ison, S., and Quddus, M. (2014). On-street parking search. *Transport. Res. Rec. J. Transport. Res. Board* 2469, 65–75. doi: 10.3141/2469-08

Cao, Y., Yang, Z. Z., and Zuo, Z. Y. (2017). The effect of curb parking on road capacity and traffic safety. *Eur. Transport Res. Rev*. 9:4. doi: 10.1007/s12544-016-0219-3

Cooper, P. J. (1990). Differences in accident characteristics among elderly drivers and between elderly and middle-aged drivers. *Accident Anal. Prevent*. 22, 499–508. doi: 10.1016/0001-4575(90)90044-L

Crundall, D., and Underwood, G. (2011). "Visual attention while driving: measures of eye movements used in driving research," in *Handbook of Traffic Psychology*, ed B. E. Porter (San Diego, CA: Academic Press), 137–148. doi: 10.1016/B978-0-12-381984-0.10011-6

Daisa, J. M., and Peers, J. B. (1997). "Narrow residential streets: do they really slow down speeds?" in *Proceedings of the Institute of Transportation Engineers 67th Annual Meeting* (Tampa, FL: Institute of Transportation Engineers).

Dingus, T., McGehee, D., Hulse, M., Jahns, S., and Manakkal, N. (1995). *Travtrek Evaluation Task C3–Camera Car Study*. Technical report, Turner-Fairbank Highway Research Centre, McLean, VA.

Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., et al. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proc. Natl. Acad. Sci. U.S.A*. 113, 2636–2641. doi: 10.1073/pnas.1513271113

Edquist, J., Rudin-Brown, C. M., and Lenné, M. G. (2012). The effects of on-street parking and road environment visual complexity on travel speed and reaction time. *Accident Anal. Prevent*. 45, 759–765. doi: 10.1016/j.aap.2011.10.001

Engström, J., Johansson, E., and Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transport. Res. F Traffic Psychol. Behav*. 8, 97–120. doi: 10.1016/j.trf.2005.04.012

Greibe, P. (2003). Accident prediction models for urban roads. *Accident Anal. Prevent*. 35, 273–285. doi: 10.1016/S0001-4575(02)00005-2

Hallihan, G., Mayer, A., Caird, J., and Milloy, S. (2011). Effects of hybrid interface on ecodriving and driver distraction. *Transport. Res. Rec. J. Transport. Res. Board* 2248, 74–80. doi: 10.3141/2248-10

Hampshire, R., and Shoup, D. (2018). What share of traffic is cruising for parking? *J. Transport Econ. Policy* 52, 184–201.

Hart, S. G. (2006). "NASA-Task Load Index (NASA-TLX); 20 years later. *Proc. Hum. Fact. Ergon. Soc. Annu. Meet*. 50, 904–908. doi: 10.1177/154193120605000909

Hart, S. G., and Staveland, L. E. (1988). Development of NASA-TLX (task load index): results of empirical and theoretical research. *Adv. Psychol*. 52, 139–183. doi: 10.1016/S0166-4115(08)62386-9

Healey, J., and Picard, R. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Trans. Intell. Transport. Syst*. 6, 156–166. doi: 10.1109/TITS.2005.848368

Hennessy, D. A., and Wiesenthal, D. L. (1999). Traffic congestion, driver stress, and driver aggression. *Aggress. Behav*. 25, 409–423. doi: 10.1002/(SICI)1098-2337(1999)25:6<409::AID-AB2>3.0.CO;2-0

Highway Research Board (1971). *Parking Principles*. Technical report, National Academy of Sciences, Washington, DC.

Horrey, W., and Wickens, C. (2007). In-vehicle glance duration: distributions, tails, and model of crash risk. *Transport. Res. Rec. J. Transport. Res. Board* 2018, 22–28. doi: 10.3141/2018-04

Kaptein, N., Theeuwes, J., and Van Der Horst, R. (1996). Driving simulator validity: some considerations. *Transport. Res. Rec. J. Transport. Res. Board* 1550, 30–36. doi: 10.1177/0361198196155000105

Klauer, S., Dingus, T., Neale, V., Sudweeks, J., and Ramsey, D. (2006). *The Impact of Driver Inattention On Near Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data*. Technical report. Virginia Tech Transportation Institute. doi: 10.1037/e729262011-001

Lerner-lam, E., Celniker, S. R., Halbert, G. W., Chellman, C., and Ryan, S. (1992). Neo-traditional neighborhood design and its implications for traffic engineering. *Instit. Transport. Eng. J*. 62, 17–25.

Liang, Y., Lee, J. D., and Yekhshatyan, L. (2012). How dangerous is looking away from the road? Algorithms predict crash risk from glance patterns in naturalistic driving. *Hum. Fact*. 54, 1104–1116. doi: 10.1177/0018720812446965

Marshall, W., Garrick, N., and Hansen, G. (2008). Reassessing on-street parking. *Transport. Res. Rec. J. Transport. Res. Board* 2046, 45–52. doi: 10.3141/2046-06

McGwin, G. Jr., and Brown, D. B. (1999). Characteristics of traffic crashes among young, middle-aged, and older drivers. *Accident Anal. Prevent*. 31, 181–198. doi: 10.1016/S0001-4575(98)00061-X

Mehler, B., Reimer, B., and Coughlin, J. F. (2012). Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task: an on-road study across three age groups. *Hum. Fact*. 54, 396–412. doi: 10.1177/0018720812442086

Mehler, B., Reimer, B., Pohlmeyer, A. E., and Coughlin, J. F. (2008). The association between heart rate reactivity and driving performance under dual task demands in late middle age drivers. *Adv. Transport. Stud*. 53–70. doi: 10.4399/97888548259876

Ponnambalam, C. T., Cheng, R., and Donmez, B. (2018). Effects of searching for street parking on driver behaviour and physiology: results from an on-road instrumented vehicle study. *Proc. Hum. Fact. Ergon. Soc. Annu. Meet*. 62, 1404–1408. doi: 10.1177/1541931218621320

Regan, M. A., Lee, J. D., and Young, K. L. (2008). *Driver Distraction: Theory, Effects, and Mitigation*. Boca Raton, FL: CRC Press; Taylor & Francis Group. doi: 10.1201/9781420007497

Reimer, B., Mehler, B., and Coughlin, J. F. (2016). Reductions in self-reported stress and anticipatory heart rate with the use of a semi-automated parallel parking system. *Appl. Ergon*. 52, 120–127. doi: 10.1016/j.apergo.2015.07.008

Reimer, B., Mehler, B., and Donmez, B. (2014). A study of young adults examining phone dialing while driving using a touchscreen vs. a button style flip-phone. *Transport. Res. F Traffic Psychol. Behav*. 23, 57–68. doi: 10.1016/j.trf.2013.12.017

Shaffer, F., and Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Front. Public Health* 5:258. doi: 10.3389/fpubh.2017.00258

Sisiopiku, V. P. (2001). "On-street parking on state roads," in *Proceedings of ITE 2001 Annual Meeting and Exhibit* (Chicago, IL).

Sodhi, M., Reimer, B., and Llamazares, I. (2002). Glance analysis of driver eye movements to evaluate distraction. *Behav. Res. Methods Instrum. Comput*. 34, 529–538. doi: 10.3758/BF03195482

Weinberger, R., and Millard-Ball, A. (2017). "Parking search caused congestion : where's all the fuss?" in *Transportation Research Board 96th Annual Meeting* (Washington, DC).

Zhang, H., Smith, M. R. H., and Witt, G. J. (2006). Identification of real-time diagnostic measures of visual distraction with an automatic eye-tracking system. *Hum. Fact. J. Hum. Fact. Ergon. Soc*. 48, 805–821. doi: 10.1518/001872006779166307

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 27 October 2020
doi: 10.3389/fpsyg.2020.596257

Evaluation of Naturalistic Driving Behavior Using In-Vehicle Monitoring Technology in Preclinical and Early Alzheimer's Disease

Jennifer D. Davis[1,2†], Ganesh M. Babulal[3*†], George D. Papandonatos[4], Erin M. Burke[2], Christopher B. Rosnick[3], Brian R. Ott[5] and Catherine M. Roe[3]*

[1] Department of Psychiatry and Human Behavior, Warren Alpert Medical School of Brown University, Providence, RI, United States, [2] Neuropsychology Program, Department of Psychiatry, Rhode Island Hospital, Providence, RI, United States, [3] Department of Neurology, Charles F. and Joanne Knight Alzheimer's Disease Research Center, Washington University School of Medicine, St. Louis, MO, United States, [4] Department of Biostatistics, Brown University, Providence, RI, United States, [5] Department of Neurology, Rhode Island Hospital, Warren Alpert Medical School of Brown University, Providence, RI, United States

OPEN ACCESS

Edited by:
Haneen Farah, Delft University of Technology, Netherlands

Reviewed by:
Silvia Ceccacci, Marche Polytechnic University, Italy
Jordi Solé-Casals, Universitat de Vic – Universitat Central de Catalunya, Spain

***Correspondence:**
Jennifer D. Davis
Jennifer_Davis@Brown.edu
Ganesh M. Babulal
babulalg@wustl.edu

†These authors share first authorship

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Psychology

Received: *18 August 2020*
Accepted: *05 October 2020*
Published: *27 October 2020*

Citation:
Davis JD, Babulal GM, Papandonatos GD, Burke EM, Rosnick CB, Ott BR and Roe CM (2020) Evaluation of Naturalistic Driving Behavior Using In-Vehicle Monitoring Technology in Preclinical and Early Alzheimer's Disease. Front. Psychol. 11:596257. doi: 10.3389/fpsyg.2020.596257

Cognitive impairment is a significant risk factor for hazardous driving among older drivers with Alzheimer's dementia, but little is known about how the driving behavior of mildly symptomatic compares with those in the preclinical, asymptomatic phase of Alzheimer's disease (AD). This study utilized two in-car technologies to characterize driving behavior in symptomatic and preclinical AD. The goals of this pilot study were to (1) describe unsafe driving behaviors in individuals with symptomatic early AD using G-force triggered video capture and (2) compare the driving habits of these symptomatic AD drivers to two groups of cognitively normal drivers, those with and those without evidence of cerebral amyloidosis (CN/A+ and CN/A−) using a global positioning system (GPS) datalogger. Thirty-three drivers (aged 60+ years) were studied over 3 months. G-force triggered video events captured instances of near-misses/collisions, traffic violations, risky driver conduct, and driving fundamentals. GPS data were sampled every 30 s and all instances of speeding, hard braking, and sudden acceleration were recorded. For the early AD participants, video capture identified driving unbelted, late response, driving too fast for conditions, traffic violations, poor judgment, and not scanning intersections as the most frequently occurring safety errors. When evaluating driving using the GPS datalogger, hard breaking events occurred most frequently on a per trip basis across all three groups. The CN/A+ group had the lowest event rate across all three event types with lower instances of speeding. Slower psychomotor speed (Trail Making Part A) was associated with fewer speeding events, more hard acceleration events, and more overall events. GPS tracked instances of speeding were correlated with total number of video-captured near-collisions/collisions and driving fundamentals. Results demonstrate the utility of electronic monitoring to identify potentially unsafe driving events in symptomatic and preclinical AD. Results suggest that drivers with preclinical AD may compensate for early, subtle cognitive changes by driving more slowly and cautiously than healthy older drivers or those with cognitive impairment. Self-regulatory changes in driving behavior appear to occur in the preclinical phase of AD, but safety concerns may not arise until symptoms of cognitive impairment emerge and the ability to self-monitor declines.

Keywords: driving, Alzheimer's disease, naturalistic, technology, preclinical Alzheimer's disease, driving mobility

INTRODUCTION

Older drivers with cognitive impairment are at high risk for unsafe driving, carrying a relative crash risk of 2–5 times higher compared to matched controls (Marshall, 2008). Alzheimer's disease (AD), one of the most prevalent diseases affecting cognition in older adults (Alzheimer's Association, 2019), adversely impacts driving. Individuals with AD are at increased risk for failing a road test with disease progression (Duchek et al., 2003; Ott et al., 2008), make more safety errors when driving in their own environment compared to cognitively normal older adults (Davis et al., 2012), and are at greater risk for crashes (Drachman and Swearer, 1993). Despite their increased risk, drivers with AD continue to drive during their disease course but may modify their behavior by reducing driving in complex situations (Festa et al., 2013; Molnar et al., 2013). This includes driving without passengers, during daytime hours, good weather, light traffic, and residential rather than commercial environments. Festa et al. (2013) Despite behavioral modification, however, most drivers with AD eventually need to cease driving due to progressive cognitive and functional decline (Connors et al., 2018).

Little is known about how driving is affected early in the disease process or the pathological process underlying that decline. AD begins decades (∼2–3) before overt expression of cognitive symptoms as beta amyloid, the pathological marker of AD, begins to accumulate in the brain. The presence of cerebral amyloid has been directly associated with driving. Specifically, several postmortem studies of the brains of older drivers who were killed in motor vehicle crash (MVCs) have found that many had the neuropathologic changes of AD but had never been diagnosed (Johansson et al., 1997; Viitanen et al., 1998; Kibayashi and Shojo, 2002; Gorrie et al., 2007).

The development of biomarkers for AD has allowed for *in vivo* studies of amyloid deposition and driving behavior. Traffic violations and accidents over the 3 years prior to brain imaging was strongly related to accumulating amyloid on positron emission tomography [PET] scans, even in individuals not yet displaying measurable cognitive impairments resulting from the disease (i.e., preclinical AD) (Ott et al., 2017b). More abnormal levels of cerebral amyloid detected using Pittsburgh compound B radiotracer via PET predicted poorer performance on a standardized road test among cognitively normal older adults (Roe et al., 2017). Using naturalistic methodology, real world driving behavior appears to change even in the preclinical phase of the disease with amyloid positive older adults driving to fewer places/unique destinations, traveling fewer days, and taking fewer trips compared to amyloid negative same-aged peers. Furthermore, those with preclinical AD had fewer trips with any aggressive behaviors and showed a greater decline across a 2.5-year follow-up period in the number of days driving per month and number of trips taken (Roe et al., 2019).

CURRENT STUDY

To date, there is little data examining the spectrum of age-related driving behavior ranging from normal cognition to preclinical AD to symptomatic AD. Using a convenience sample of older drivers, the goal of this pilot study was to describe naturalistic driving behavior among these three groups using in-vehicle video and global positioning system (GPS) technologies. The first aim was to describe the types of hazardous driving errors captured by video technology in a subset of the sample of older adult drivers with early AD. The second aim was to compare the driving behaviors of drivers with early AD to two groups of cognitively normal (CN) drivers, those with evidence of brain amyloid (preclinical AD; CN/A+) to healthy adults without evidence of brain (CN/A−) over 3 months of naturalistic driving.

MATERIALS AND METHODS

Participants

Symptomatic AD drivers (n = 11) were recruited from a multidisciplinary outpatient memory clinic in Rhode Island. All participants underwent a diagnostic evaluation by a neurologist at the Center. Neurological examination results were judged to be normal for age or consistent with AD. For inclusion, Mini-Mental State Examination (MMSE) (Folstein et al., 1975) scores were <28 and Clinical Dementia Rating (Morris, 1993) (CDR) scores were categorized as CDR = 0.5 or 1, indicating questionable or mild dementia. It is well established that CDR 0.5 is equivalent to mild cognitive impairment (Morris et al., 2001). Participants met diagnostic criteria for possible or probable AD based on NINCDS-ADRDA criteria (McKhann et al., 1984). Patients were on a stable dose of a cholinesterase inhibitor for 6 weeks, if prescribed. The amyloid status was known only for a subset of the cognitively impaired participants, as amyloid imaging was not a standard part of the original study protocol. Amyloid imaging was obtained within 18 months of study entry (M = 333 days; range = 56–511 days). Amyloid imaging results are presented in **Table 1** to add to the clinical characterization of the early AD group. This subset of participants underwent amyloid PET imaging using the radiotracer ^{18}F-Florbetapir (Clark et al., 2011) as part of their participation in other clinical research studies. An established standardized uptake value ratio (SUVR) threshold of $\geq$1.19 was used to indicate amyloid PET positivity (Clark et al.,

TABLE 1 | Demographic characteristics of participants.

	CDR 0 (CN/A−) (*n* = 11) (%) or M (SD)	CDR 0 (CN/A+) (*n* = 11) N (%) or M (SD)	CDR.5/1 (AD) (*n* = 11) N (%) or M (SD)	Statistic	*p*
Women, *N*	5 (45%)	5 (45%)	5 (45%)		
White, *N*	10 (91%)	11 (100%)	11 (100%)		
MCI/CDR.5, N	0 (0%)	0 (0%)	9 (82%)		
MMSE (total)	29.18 (1.67)	29.09 (1.30)	25.18 (3.84)	*F* = 9.66	0.001*
Age, years	73.33 (5.21)	73.71 (5.12)	72.88 (6.84)	*F* = 0.56	0.95
Education (years)	17.45 (1.97)	16.55 (1.86)	15.45 (3.36)	*F* = 1.78	0.19

**Implies difference between CDR 0 vs. CDR.5/1.CN/A−, cognitively normal/amyloid negative; CN/A+, cognitively normal/amyloid positive; AD, Alzheimer's disease; CDR, Clinical Dementia Rating Scale; MCI, mild cognitive impairment.*

2012; Johnson et al., 2013). All scans were read by two clinical neuroradiologists who gave also gave a clinical read of the scan. One participant had a SUVR threshold of 1.16 but had a positive clinical read. That participant was considered to be amyloid positive in this study. The sample was further characterized by apolipoprotein (ApoE) genotype, a known risk factor for AD if the ε4 allele is present. Of the eight participants with ApoE genotyping completed, 50% possessed the ε4 allele.

All participants were $\geq$60 years of age, English speaking with a valid driver's license. Exclusion criteria included ophthalmologic, physical, or neurologic disorders other than dementia that impair their driving abilities, visual acuity worse than 20/40 in best eye using distance vision measured by wall chart, homonymous hemianopia or bitemporal hemianopia, musculoskeletal disorders causing major physical handicaps, history of alcohol or substance abuse by DSM V criteria within the past year, had used sedating medications that impair level of consciousness or attention, had a language impairment that would interfere with the ability to participate in the study, or had a previous road test evaluation or opinion of caregiver or health professional that they were unsafe to drive. Study protocols were approved by the Rhode Island Hospital Institutional Review Board, and all participants provided written informed consent that was also signed by a study partner.

Since beta-amyloid is the primary driver and earliest marker of AD pathogenesis and cascade, it was selected as the main biomarker (Jack et al., 2018). A group of cognitively normal drivers were selected from participants enrolled in a longitudinal study assessing preclinical AD and driving performance (R01 AG043434) at Washington University School of Medicine in St. Louis and matched for age and gender to the early AD group. All participants were cognitively normal (CDR = 0), $\geq$65 years old, had a valid driver's license, drove at least once per week, and had *in vivo* imaging of amyloid using PET with either Pittsburgh compound B (PIB) or florbetapir AV45 radiotracer to confirm group membership. PET imaging was selected if it occurred 2 years before or 6 months after the installation date of datalogger in the participant's vehicle. Eleven participants were selected with amyloid negative scans and 11 with evidence of amyloid based on centiloid values (Su et al., 2015). Accepted cut-offs of centiloids were based on the mean cortical SUVR with partial volume correction via regional spread function (RSF) [PIB MCSUVR RSF $\geq$ 16.4 and AV45 MCSUVR RSF $\geq$ 20.6] (Su et al., 2015, 2018, 2019). The amyloid negative group had 27.3% the ApoE ε4 allele carriers, and 63.6% were ApoE ε4 allele carriers in the amyloid positive group. The participant's vehicle had to be manufactured in the year 1996 or newer in order to have access to the onboard diagnostic port (OBDII). Study protocols were approved by the Washington University Human Research Protection Office, and written informed consent was obtained from all participants.

Study Procedures

Cognition

At study enrollment, all participants completed cognitive measures, including a global measure of cognition (MMSE) and a task of psychomotor speed and set shifting (Trail Making Parts A and B, respectively) (Reitan, 1956). Trail making was selected because it has been shown to be related to naturalistic driving errors (Papandonatos et al., 2015). Time in seconds to complete the tasks were used in data analyses. Higher scores on Trails A and B reflect worse performance (i.e., slower time).

Technology

Vehicles were equipped with two forms of technology, an event-based video recording system (Drivecam)® was equipped for CDR > 0 drivers and a GPS datalogger was equipped for all drivers. Because these results reflect the combination of drivers from two different parent studies, only the vehicles of the mild AD group were equipped with the camera system. All three groups had the GPS datalogger installed to capture naturalistic driving behavior.

The *DriveCam* video camera is a palm-sized, exception-based video event recorder that was mounted in a bracket secured to the windshield behind the rearview mirror with an adhesive similar to what holds the rearview mirror in place. This system is a validated method for detecting and evaluating driving safety errors in AD (Ott et al., 2017a). Camera views were the forward roadway and the driver in the vehicle interior. Once installed, the camera continuously captured video and temporarily saved the previous several seconds in a video buffer. If the device was not triggered by excessive g-forces, all data was deleted permanently 10 s later. Data were protected from unauthorized access and removal and was only viewable by DriveCam's staff and site research staff. The site research staff, comprised of a neurologist, neuropsychologist, and occupational therapist specializing in driving evaluation, reviewed the events weekly to ensure no egregious events occurred that would prompt recommendation for a road test or driving cessation. DriveCam staff scored all videos according to a standardized procedure developed and validated by their company for commercial drivers (Myers et al., 2012). DriveCam staff were blind to all clinical information regarding dementia severity.

A commercial GPS datalogger (G2 Tracking Device™, Azuga Inc., San Jose, CA, United States) was plugged into the vehicle's OBD-II port and data collected every 30 s. This naturalistic driving methodology, termed the Driving Real World In-Vehicle Evaluation System (DRIVES), is a validated method of driving data collection for older adults (Babulal et al., 2016). The device collected data every time the vehicle was driven and recorded adverse driving events (hard braking, sudden acceleration, and speeding) anytime they occurred during a trip, regardless of the 30 s sampling that occurred for other datalogger measures. Speeding was determined based on the datalogger's GPS, specifically the latitude and longitude and the posted speed limit in the vehicle's location. The device compared the vehicle's speed to the posted speed limit and if the driver was going 6 miles per hour or more above the posted speed limit in that area, an occurrence of speeding was recorded.

Driving Behavior

A total of 3 months of driving were selected for study for all participants from a larger sample of longitudinal driving. Three months were selected because the cognitively impaired group was enrolled in a driving intervention trial where the first 3 months were monitoring only. Inclusion of only the first 3 months of driving avoids any confounds associated with the intervention. To ensure that we were only analyzing the driving behavior of the study driver, videos captured by other drivers were deleted. Since the driver cannot be identified by the GPS logger, drivers had to be driving their vehicle for at least 75% of the time to be included in the current study. Vehicle use was reported by the study partner (M = 94%, range = 75–100). The cognitively normal control group were exclusive drivers of their vehicles (100%). For all groups, any driving events were deleted if they were captured during known times that the study driver was not driving due to illness, travel, etc.

Driving errors captured by video were categorized into the following behaviors and scored according to total demerit points: collisions/near collisions, distractions (food, passengers, cell phone, and other electronic devices), awareness (late response, poor scanning of roadway, and failure to check mirrors), driver conduct (poor judgment, aggressive/reckless), fundamentals (excessive speed for conditions, failure to leave an out, and unsafe lane change), following too close, driver condition (drowsy), traffic violation (rolling stop, failure to stop at stop sign or light, speeding, not on designated roadway, and unsafe/risky behavior). The specific events or problems were graded for safety risk on a 0–10 point demerit scale. A single unsafe driving event could have more than one demerit category, such as judgment error combined with poor awareness of intersection, leading to a combined driving severity rating score for the individual items. Error frequency was used to describe error types. Total demerit points were used to examine correlational relationships between driving events captured with video and GPS logger behavior.

Driving behavior captured by the DRIVES were aggregated from daily trip reports for each vehicle over the course of the participant's entire participation in the study (up to 5 years). Daily trip data included date, starting and ending latitude and longitude, starting and ending time, distance of trip (miles), trip time (minutes), idling time (minutes), and counts of hard braking, sudden acceleration, and speeding. The first 3 months of a participant's driving behavior were extracted and were examined per 100 trips, where a trip was defined from "ignition on" to "ignition off." For example, an excursion from home to the grocery store and back to home without any other stops would be considered two trips. Hard braking, hard acceleration and speeding events per trip were analyzed separately and in combination. As multiple events of a particular type could occur in a single trip, data were analyzed for each event type using a two-step procedure: (a) number of trips per 100 in which such an event occurred and (b) number of events per trip for trips with at least one such event. In addition, other aspects of speeding, such as the duration of speeding episodes, average speed in miles, trip distance and trip time, were analyzed.

Statistical Analysis

Participants with and without preclinical AD (all CDR = 0) were matched on age and gender to participants with symptomatic AD (CDR > 0). Spearman's ρ correlational analyses using Bonferroni correction for multiple comparisons were used to examine relationships between driving behavior captured with video and the GPS datalogger. Generalized linear modeling (GLM) techniques were used to determine effects of amyloid and cognitive status on event frequency per 1,000 miles driven. All analyses were carried out using the GLM function in R 3.5.3[1]. Event frequency was modeled via an over-dispersed Poisson distribution with amyloid group (CN/A−, CN/A+, and AD) as the sole model predictor. Exposure differences were accounted for by adjusting for 1,000 miles driven via an offset variable. Point and interval estimates of group effects were estimated in the logarithmic scale as log (Rate Ratios) and then exponentiated. In addition, measures of cognition (MMSE, Trails A and Trails B time) were analyzed as secondary predictors of event frequency.

RESULTS

Demographic characteristics of these older drivers are presented in **Table 1**. Groups were matched on age (p = 0.95), education (p = 0.19), and gender. As expected, the early AD group had lower MMSE scores than the cognitively normal groups. When examining the early AD group with the video technology, there were four collisions with objects such as curbs, mail boxes, parked cars, but none with other moving vehicles. The most frequently occurring safety events were driving unbelted, late response, too fast for conditions, rolling stop, poor judgment, following too close, speeding, and failing to scan intersections (see **Figure 1**).

[1] https://cran.r-project.org

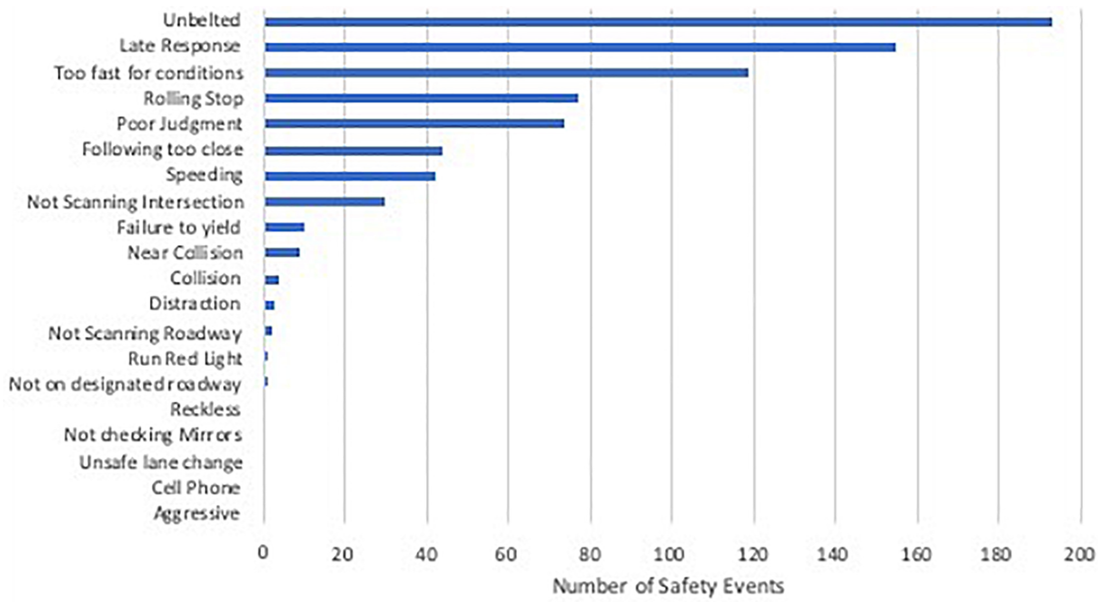


FIGURE 1 | Total number of unsafe driving events video recorded over 3 months of driving in the AD group (n = 11).

The next set of analyses included all three groups to examine differences in driving behavior using the DRIVES. **Table 2** shows that hard braking events were the most commonly occurring behavior on a per trip basis across all three groups, followed by speeding in the CN/A− group and by hard acceleration in the CN/A+ and early AD groups. Summing across event types, driving events occurred in one in five trips taken by CN/A− or early AD drivers vs. one in eight trips taken by CN/A+. When only considering trips in which adverse driving events occurred (see **Table 3**), repeat events in the same trip were only common for speeding, with 0.9 more speeding events per trip on average in the CN/A− and early AD groups compared to the amyloid positive group (4.3 vs. 3.4 speeding events).

To control for driving exposure, driving events were corrected per 1,000 miles driven (see **Table 4** for point estimates and 95% confidence intervals for the driving event rate per 1,000 miles driven). After correcting for driving exposure, speeding was the most common event type in both the CN/A− and early AD groups, while hard braking remained the most common event type in the CN/A+. **Table 5** shows point estimates and 95% confidence intervals for the ratio of driving events per 1,000 miles driven between (a) the cognitively normal groups (CN/A− vs. CN/A+) and (b) early AD vs. preclinical AD (CN/A+). The preclinical AD group had the lowest rate across all three types of driving events (hard braking, hard acceleration, and speeding), but the differences were especially pronounced for speeding behavior.

Spearman correlations were calculated to examine the relationship between error types captured with each technology in the mild AD group. The strongest correlations were between

TABLE 2 | Adverse driving event rate per 100 trips using DRIVES technology.

Event Type	Group		
	CN/A−	CN/A+	AD
Hard breaking	11.07	8.13	8.52
Hard acceleration	5.98	4.45	8.52
Speeding	9.81	2.21	5.52
Overall	21.50	12.61	19.42

TABLE 3 | Adverse driving event count per trip for trips including an event using DRIVES technology.

Event type	Group		
	CN/A−	CN/A+	AD
Hard breaking	1.27	1.26	1.28
Hard acceleration	1.47	1.26	1.41
Speeding	4.27	3.37	4.30
Overall	3.01	1.85	2.40

TABLE 4 | Adverse driving rate per 1,000 miles driven (95% Confidence Interval) using DRIVES technology.

Event type	Group		
	CN/A−	CN/A+	AD
Hard breaking	19.7 (13.9, 27.8)	13.9 (10.1, 19.3)	18.1 (13.2, 25.0)
Hard acceleration	12.3 (3.5, 42.8)	7.6 (3.7, 15.7)	20.0 (8.0, 50.5)
Speeding	58.5 (37.7, 90.6)	10.1 (3.4, 30.4)	39.6 (18.3 86.0)
Overall	90.4 (69.0, 118.4)	31.6 (17.6, 57.0)	77.8 (50.3, 120.3)

TABLE 5 | Ratios of adverse driving rates per 1,000 miles driven (95% Confidence Interval) using DRIVES technology.

Event type	Group	
	CH/Amyloid− vs. CH/Amyloid+	AD vs. CH/Amyloid+
Hard breaking	1.41 (0.88, 2.27)	1.30 (0.83, 2.05)
Hard acceleration	1.62 (0.38, 6.84)	2.64 (0.82, 8.53)
Speeding	5.78 (1.77, 18.90)$^{++}$	3.92 (1.02, 15.10)$^{+}$
Overall	2.86 (1.50, 5.46)$^{+++}$	2.46 (1.18, 5.11)$^{+}$

$^{+}$*Implies* $p < 0.05$, $^{++}$*Implies* $p < 0.01$, $^{+++}$*Implies* $p < 0.001$.

speeding registered by the datalogger and collision/near collisions and fundamentals of driving (i.e., failing to keep an out, too fast for conditions, and failure to yield; rho = 0.81 and 0.87, respectively. Hard breaking and hard acceleration were not consistently related to errors captured by video analysis.

Poisson regression models showed no significant effect of MMSE or Trails B time on driving event frequency (count of braking, speeding, and sudden acceleration in a trip). However, Trails A time had significant non-linear effects on overall event frequency (p = 0.03) that are depicted graphically in **Figure 2**. Further analysis of the non-linear effects (e.g., parabolic curve) taken apart showed that the initial drop in driving events was correlated with lower speeding event rates with increasing Trails A time (p = 0.03), whereas the later increase in events was related to higher hard acceleration event rates with increasing Trails A time (p = 0.09). Hard braking event rates were relatively insensitive to Trails A time (p = 0.64).

DISCUSSION

Two in-vehicle technologies were used in our study to characterize driving errors and behaviors in older adult drivers with preclinical and early symptomatic AD compared to cognitively normal adults without any evidence of AD pathology. In the early AD drivers, g-forced triggered events produced common errors predominantly related to inadequate anticipation of situations, such as late response or driving too fast for the situation. AD drivers also frequently showed errors of judgment and made frequent traffic violations around speeding and responding to road signage (i.e., stop signs and traffic lights). Despite these instances of poor judgment, actual collisions were very rare during the study period. These g-force triggered safety events are consistent with the types of driving events we previously captured with continuous video recording in mild AD drivers compared to cognitively normal older

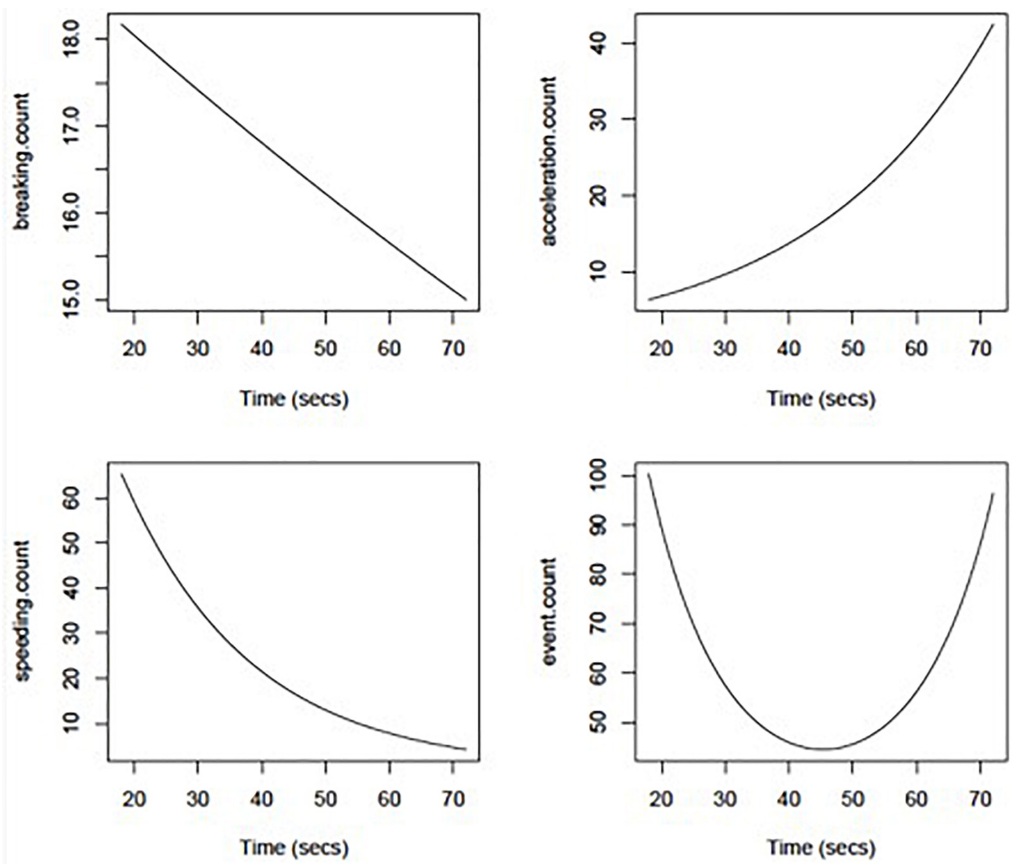


FIGURE 2 | Adverse driving rates per 1,000 miles driven using DRIVES technology as a function of Trail Making Part A duration (in seconds).

adults (Davis et al., 2012). These findings also confirm prior data that used a semiautomated data reduction method to isolate relevant driving errors from continuous video recorded naturalistic driving to detect cognitive impairment-associated driving behaviors in older adults (Davis et al., 2018; Moharrer et al., 2020). This suggests that an event-based approach, rather than more costly and staff time intensive continuous monitoring of behavior via video, may be a sensitive method to study driving risk in AD.

Global positioning system data logger technology was used to address potential differences in driving behaviors among cognitively normal, preclinical AD, and symptomatic AD. From this data, a distinct pattern of driving behavior emerged among drivers with preclinical AD compared to cognitively normal older adults without evidence of AD and early AD. Specifically, drivers with preclinical AD drove more slowly and had the lowest number of aggressive events over the 3-month period. These data are also consistent with prior work showing that older drivers in the preclinical phase of AD restrict their driving compared to healthy peers (Babulal et al., 2019; Roe et al., 2019).

The current results extend these findings by offering insights into how driving may change across the spectrum of normal aging to symptomatic AD by including a group of mild AD drivers for comparison. Results suggest that in the earliest stage of AD there may be a period of self-regulatory behavior during amyloid accumulation but where cognitive functioning remains unaffected. As the disease progresses and cognition begins to decline with disease progression and neurodegeneration, inhibitory control over more aggressive driving behavior may begin to erode. As such, the AD drivers may revert back to "normal" driving habits including excessive speed. This is consistent with prior work showing that AD drivers whose naturalistic driving was video recorded showed poorer tactical self-regulation behavior and made twice as many critical events as healthy older drivers. They were also three times more likely to be unaware of these events (Paire-Ficout et al., 2018). Unfortunately, the early AD group may need to continue to use compensatory strategies (i.e., cautious driving) to prevent more egregious safety errors and accidents. It is possible that early AD drivers could maintain independence longer by increasing their awareness of driving errors and the provision of compensatory strategies. Prior intervention studies suggest that this may be possible. For example, we showed that a behavioral intervention aimed at correcting the specific driving errors reduced the frequency of driving errors in a group of drivers with early AD (Ott et al., 2017a).

Given that the cognitive processes of attention and executive functioning decline in AD and have been shown to relate to driving errors (Anderson et al., 2012; Papandonatos et al., 2015), we examined the relationship between measures of these constructs and driving events captured by the data logger. In this study, simple psychomotor speed was associated with driving behavior. Specifically, more acceleration events, slower speeds, and more overall events were associated with slower psychomotor speed. This suggests a relationship between cognitive impairment and driving behaviors may be captured with the data logger.

This study utilized two different passive monitoring in-vehicle technologies to understand driving behavior. As such, it was of interest to explore the relationship between behaviors captured with g-force triggered video technology vs. the data logger, as the video provides more context to the behaviors captured with the data logger. Data indicated that some, but not all, event types were highly related. Specifically instances of speeding registered by the datalogger were related to instances of collision/near collisions and errors in driving fundamentals. Hard breaking and hard acceleration were not strongly related to safety errors. These relationships could only be examined in a small sample of early AD participants, and more work will be needed to better understand these findings, but preliminarily, these data support the idea that aggressive events captured with the data logger may indeed reflect risky driving behavior.

There are several limitations to this study. First, this is a small sample of older drivers, and results should be viewed as preliminary and only applicable to the aging population. The CN and early AD participants were recruited from two regionally different locations. It is possible that geographic differences in population density, type of driving, and seasonal weather changes may have impacted the results. The cognitively normal older adult drivers (CDR 0) did not have the video technology installed in their vehicles, so it is unclear how these behaviors may occur in a cognitively normal population or the degree to which amyloidosis might influence this relationship based on these data alone. In addition, it is unclear the degree to which events captured with the GPS data logger correlate with actual unsafe behavior or simply more assertive or effective defensive driving in a healthy population. The strong relationship between video captured safety events and GPS events in the cognitively impaired group would suggest that the GPS captured events reflect actual risky behavior, possibly more erratic driving, in the cognitively impaired group. Lastly, results need be replicated in a larger sample of racially and ethnically diverse older drivers as participants in this study were all non-Hispanic white, which limits generalizability to other diverse populations.

Our results offer preliminary findings that suggest that in-vehicle technology can detect behavioral differences between drivers at different points in the spectrum of normal aging to early AD. With the increase in Advanced Driver Assistance Systems (ADAS) as standard features in vehicles, instrumented vehicle technology may offer a unique opportunity to detect early behavioral change in older adults that could signal increased risk for unsafe driving. These types of technology could be used to identify when an individual may need to start considering driving retirement with instances of unsafe behaviors serving as early markers of cognitive decline. Objective measurement of driving changes, in conjunction with report of driving changes using driving questionnaires could lead to further assessment with an occupational therapist, driving specialist, or monitoring by a healthcare provider. Future research should employ multiple modalities for assessing driving behavior over extended periods of time to obtain a more complete characterization of the aging driver.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Rhode Island Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JD and GB contributed to data collection, data interpretation, and manuscript preparation. GP conducted statistical analyses. CR provided data management. BO and CR contributed to the interpretation of results and manuscript preparation. EB provided study coordination and data collection. All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was supported by grant #NPSASA-15-362133 from the Alzheimer's Association, the National Institute on Aging (R01AG056466, R01AG067428, P50AG005681, P01AG003991, and P01AG026276), and the Charles and Joanne Knight Alzheimer's Research Initiative of the Washington University Knight Alzheimer's Disease Research Center.

REFERENCES

Alzheimer's Association (2019). 2019 Alzheimer's disease facts and figures. *Alzheimer's Dement. J.* 15, 321–387.

Anderson, S. W., Aksan, N., Dawson, J. D., Uc, E. Y., Johnson, A. M., and Rizzo, M. (2012). Neuropsychological assessment of driving safety risk in older adults with and without neurologic disease. *J. Clin. Exp. Neuropsychol.* 34, 895–905. doi: 10.1080/13803395.2011.630654

Babulal, G. M., Stout, S. H., Benzinger, T. L. S., Ott, B. R., Carr, D. B., Webb, M., et al. (2019). A naturalistic study of driving behavior in older adults and preclinical alzheimer disease: a pilot study. *J. Appl. Gerontol.* 38, 277–289. doi: 10.1177/0733464817690679

Babulal, G. M., Traub, C. M., Webb, M., Stout, S. H., Addison, A., Carr, D. B., et al. (2016). Creating a driving profile for older adults using GPS devices and naturalistic driving methodology. *F1000Res* 5:2376. doi: 10.12688/f1000research.9608.2

Clark, C. M., Pontecorvo, M. J., Beach, T. G., Bedell, B. J., Coleman, R. E., Doraiswamy, P. M., et al. (2012). Cerebral PET with florbetapir compared with neuropathology at autopsy for detection of neuritic amyloid-beta plaques: a prospective cohort study. *Lancet Neurol.* 11, 669–678. doi: 10.1016/s1474-4422(12)70142-4

Clark, C. M., Schneider, J. A., Bedell, B. J., Beach, T. G., Bilker, W. B., Mintun, M. A., et al. (2011). Use of florbetapir-PET for imaging beta-amyloid pathology. *JAMA* 305, 275–283.

Connors, M. H., Ames, D., Woodward, M., and Brodaty, H. (2018). Predictors of driving cessation in dementia: baseline characteristics and trajectories of disease progression. *Alzheimer Dis. Assoc. Disord.* 32, 57–61. doi: 10.1097/wad.0000000000000212

Davis, J. D., Papandonatos, G. D., Miller, L. A., Hewitt, S. D., Festa, E. K., Heindel, W. C., et al. (2012). Road test and naturalistic driving performance in healthy and cognitively impaired older adults: does environment matter? *J. Am. Geriatr. Soc.* 60, 2056–2062. *doi: 10.1111/j.1532-5415.2012.04206.x*

Davis, J. D., Wang, S., Festa, E. K., Luo, G., Moharrer, M., Bernier, J., et al. (2018). Detection of risky driving behaviors in the naturalistic environment in healthy older adults and mild alzheimer's disease. *Geriatrics* 3:13. doi: 10.3390/geriatrics3020013

Drachman, D. A., and Swearer, J. M. (1993). Driving and Alzheimer's disease: the risk of crashes. *Neurology* 43, 2448–2456. doi: 10.1212/wnl.43.12.2448

Duchek, J. M., Carr, D. B., Hunt, L., Roe, C. M., Xiong, C., Shah, K., et al. (2003). Longitudinal driving performance in early-stage dementia of the Alzheimer type. *J. Am. Geriatr. Soc.* 51, 1342–1347. doi: 10.1046/j.1532-5415.2003.51481.x

Festa, E. K., Ott, B. R., Manning, K. J., Davis, J. D., and Heindel, W. C. (2013). Effect of cognitive status on self-regulatory driving behavior in older adults: an assessment of naturalistic driving using in-car video recordings. *J. Geriatr. Psychiatry Neurol.* 26, 10–18. doi: 10.1177/0891988712473801

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr Res.* 12, 189–198.

Gorrie, C. A., Rodriguez, M., Sachdev, P., Duflou, J., and Waite, P. M. (2007). Mild neuritic changes are increased in the brains of fatally injured older motor vehicle drivers. *Accid. Anal. Prev.* 39, 1114–1120. doi: 10.1016/j.aap.2007.02.008

Jack, C. R. Jr., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., et al. (2018). NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimer's Dement.* 14, 535–562. doi: 10.1016/j.jalz.2018.02.018

Johansson, K., Bogdanovic, N., Kalimo, H., Winblad, B., and Viitanen, M. (1997). Alzheimer's disease and apolipoprotein E epsilon 4 allele in older drivers who died in automobile accidents. *Lancet* 349, 1143–1144. doi: 10.1016/s0140-6736(97)24016-x

Johnson, K. A., Sperling, R. A., Gidicsin, C. M., Carmasin, J. S., Maye, J. E., Coleman, R. E., et al. (2013). Florbetapir (F18-AV-45) PET to assess amyloid burden in Alzheimer's disease dementia, mild cognitive impairment, and normal aging. *Alzheimers Dement.* 9 (5 Suppl), S72–S83. doi: 10.1016/j.jalz.2012.10.007

Kibayashi, K., and Shojo, H. (2002). Incipient Alzheimer's disease as the underlying cause of a motor vehicle crash. *Med. Sci. Law* 42, 233–236. doi: 10.1177/002580240204200307

Marshall, S. C. (2008). The role of reduced fitness to drive due to medical impairments in explaining crashes involving older drivers. *Traffic Inj. Prev.* 9, 291—298. doi: 10.1080/15389580801895244

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., and Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on alzheimer's disease. *Neurology* 34, 939–944. doi: 10.1212/wnl.34.7.939

Moharrer, M., Wang, S., Davis, J. D., Ott, B. R., and Luo, G. (2020). Driving safety of cognitively-impaired drivers based on near collisions in naturalistic driving. *J. Alzheimers Dis. Rep.* 4, 1–7. doi: 10.3233/adr-190159

Molnar, L. J., Eby, D. W., Charlton, J. L., Langford, J., Koppel, S., Marshall, S., et al. (2013). Driving avoidance by older adults: is it always self-regulation? *Accid. Anal. Prev.* 57, 96–104. doi: 10.1016/j.aap.2013.04.010

Morris, J. C. (1993). The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* 43, 2412–2414. doi: 10.1212/wnl.43.11.2412-a

Morris, J. C., Storandt, M., Miller, J. P., McKeel, D. W., Price, J. L., Rubin, E. H., et al. (2001). Mild cognitive impairment represents early-stage Alzheimer disease. *Arch. Neurol.* 58, 397–405.

Myers, L. A., Russi, C. S., Will, M. D., and Hankins, D. G. (2012). Effect of an onboard event recorder and a formal review process on ambulance driving behaviour. *Emerg. Med. J.* 29, 133–135. doi: 10.1136/emj.2010.104034

Ott, B. R., Davis, J. D., and Bixby, K. (2017a). Video feedback intervention to enhance the safety of older drivers with cognitive impairment. *Am. J. Occup. Ther.* 71, 7102260020p1-7102260020p7. doi: 10.5014/ajot.2017.020404

Ott, B. R., Jones, R. N., Noto, R. B., Yoo, D. C., Snyder, P. J., Bernier, J. N., et al. (2017b). Brain amyloid in preclinical Alzheimer's disease is associated with increased driving risk. *Alzheimers Dement.* 6, 136–142. doi: 10.1016/j.dadm.2016.10.008

Ott, B. R., Heindel, W. C., Papandonatos, G. D., Festa, E. K., Davis, J. D., Daiello, L. A., et al. (2008). A longitudinal study of drivers with Alzheimer disease. *Neurology* 70, 1171–1178. doi: 10.1212/01.wnl.0000294469.27156.30

Paire-Ficout, L., Lafont, S., Conte, F., Coquillat, A., Fabrigoule, C., Ankri, J., et al. (2018). Naturalistic driving study investigating self-regulation behavior in early Alzheimer's disease: a pilot study. *J. Alzheimers Dis.* 63, 1499–1508. doi: 10.3233/jad-171031

Papandonatos, G. D., Ott, B. R., Davis, J. D., Barco, P. P., and Carr, D. B. (2015). Clinical utility of the trail-making test as a predictor of driving performance in older adults. *J. Am. Geriatr. Soc.* 63, 2358–2364. doi: 10.1111/jgs.13776

Reitan, R. M. (1956). *Trail Making Test: Manual for Administration, Scoring, and Interpretation*. Indiana, IL: Indiania University Press.

Roe, C. M., Barco, P. P., Head, D. M., Ghoshal, N., Selsor, N., Babulal, G. M., et al. (2017). Amyloid imaging, cerebrospinal fluid biomarkers predict driving performance among cognitively normal individuals. *Alzheimer Dis. Assoc. Disord.* 31, 69–72. doi: 10.1097/wad.0000000000000154

Roe, C. M., Stout, S. H., Rajasekar, G., Ances, B. M., Jones, J. M., Head, D., et al. (2019). A 2.5-year longitudinal assessment of naturalistic driving in preclinical Alzheimer's disease. *J. Alzheimers Dis.* 68, 1625–1633. doi: 10.3233/JAD-181242

Su, Y., Blazey, T. M., Snyder, A. Z., Raichle, M. E., Marcus, D. S., Ances, B. M., et al. (2015). Partial volume correction in quantitative amyloid imaging. *Neuroimage* 107, 55–64. doi: 10.1016/j.neuroimage.2014.11.058

Su, Y., Flores, S., Hornbeck, R. C., Speidel, B., Vlassenko, A. G., Gordon, B. A., et al. (2018). Utilizing the centiloid scale in cross-sectional and longitudinal PiB PET studies. *Neuroimage Clin.* 19, 406–416. doi: 10.1016/j.nicl.2018.04.022

Su, Y., Flores, S., Wang, G., Hornbeck, R. C., Speidel, B., Joseph-Mathurin, N., et al. (2019). Comparison of Pittsburgh compound B and florbetapir in cross-sectional and longitudinal studies. *Alzheimers Dement.* 11, 180–190. doi: 10.1016/j.dadm.2018.12.008

Viitanen, M., Johansson, K., Bogdanovic, N., Berkowicz, A., Druid, H., Eriksson, A., et al. (1998). Alzheimer changes are common in aged drivers killed in single car crashes and at intersections. *Forensic Sci. Int.* 96, 115–127. doi: 10.1016/s0379-0738(98)00114-5

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 03 December 2020
doi: 10.3389/fpsyg.2020.585280

Pedestrians' Understanding of a Fully Autonomous Vehicle's Intent to Stop: A Learning Effect Over Time

Michal Hochman, Yisrael Parmet and Tal Oron-Gilad*

Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, Beersheba, Israel

This study explored pedestrians' understanding of Fully Autonomous Vehicles (FAVs) intention to stop and what influences pedestrians' decision to cross the road over time, i.e., learnability. Twenty participants saw fixed simulated urban road crossing scenes with a single FAV on the road as if they were pedestrians intending to cross. Scenes differed from one another in the FAV's, distance from the crossing place, its physical size, and external Human-Machine Interfaces (e-HMI) message by background color (red/green), message type (status/advice), and presentation modality (text/symbol). Eye-tracking data and decision measurements were collected. Results revealed that pedestrians tend to look at the e-HMI before making their decision. However, they did not necessarily decide according to the e-HMIs' color or message type. Moreover, when they complied with the e-HMI proposition, they tended to hesitate before making the decision. Overall, a learning effect over time was observed in all conditions regardless of e- HMI features and crossing context. Findings suggest that pedestrians' decision making depends on a combination of the e-HMI implementation and the car distance. Moreover, since the learning curve exists in all conditions and has the same proportion, it is critical to design an interaction that would encourage higher probability of compatible decisions from the first phase. However, to extend all these findings, it is necessary to further examine dynamic situations.

Keywords: fully autonomous vehicle, external human-machine interfaces, presentation modality, road crossing, eye movements

OPEN ACCESS

Edited by:
Christian Collet,
Université Claude Bernard Lyon 1, France

Reviewed by:
Michal Przemyslaw Muszynski,
Université de Genève, Switzerland
Enkelejda Kasneci,
University of Tübingen, Germany

****Correspondence:***
Michal Hochman
hochmanm@post.bgu.ac.il

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Psychology

Received: *20 July 2020*
Accepted: *12 November 2020*
Published: *03 December 2020*

Citation:
Hochman M, Parmet Y and Oron-Gilad T (2020) Pedestrians' Understanding of a Fully Autonomous Vehicle's Intent to Stop: A Learning Effect Over Time. Front. Psychol. 11:585280. doi: 10.3389/fpsyg.2020.585280

INTRODUCTION

Crossing the street in the Fully Autonomous vehicle (FAV) era will differ from road crossing today since, among other things, the crossing decision will not be influenced by informal pedestrian – driver human-human communication (like eye contact, facial expressions, gestures, or body movements) that is necessary to understand driver intention (Rasouli et al., 2018). Thus, in the FAV era, with the absence of a human driver, the main challenge would be to establish pedestrians' understanding of FAV intentions so that they can make safe crossing decisions.

Simulation studies reported that an external human-machine interfaces (e-HMI) mounted on the vehicle enhances the interaction with pedestrians by reducing the uncertainty regarding FAV intent, improving pedestrians' initial trust and understanding (Deb et al., 2018; Ackermann et al., 2019; Ackermans et al., 2020). It was claimed that pedestrians have high trust and confidence in the e-HMI, even before getting to know it, and they tend to comply with its instructions

(Holländer and Butz, 2019). Moreover, even after a malfunction, trust and confidence recovered quickly (Holländer and Butz, 2019). Inconsistent with this claim, a Wizard of Oz (WoZ) study suggested that people prefer to decide for themselves when to cross, as they do today, based on the FAV's distance and speed from their crossing point (Clamann et al., 2017). Another video-based study followed by questionnaires reported similar trends (Mahadevan et al., 2018).

Few studies dealt with the form of the visual e-HMI messages. One distinction is between advice messages that suggest to the pedestrian whether to cross the road or not (e.g., "please cross," "walk," "stop") and status messages that display the FAV status, like "Driving," "Stopping," etc (Deb et al., 2018; Ackermann et al., 2019). Another distinction was between text and symbol messages (Deb et al., 2018; Ackermann et al., 2019). A study that looked at pedestrians' comprehension of the e-HMI messages through questionnaires revealed that participants assessed advice messages as more comfortable than status messages, independent of text or symbol-based presentation (Ackermann et al., 2019). On the contrary, Deb et al. (2018) found that a textual "Braking" status message was preferred over textual advice "Walk" message.

Studies also varied in the way they measured pedestrians' understanding. One way is to measure the time it took the pedestrian to decide whether to cross the road in a VR simulation (Clamann et al., 2017; Dey et al., 2018). Decision time was faster when e-HMI display included text or symbol compared to no e-HMI (Dey et al., 2018). Another way is through subjective questionnaires and ratings (Deb et al., 2018; Ackermann et al., 2019). A third way is through accuracy rate, that is, whether the pedestrian's decision was in agreement with what was being displayed on the e-HMI [compatible responses, noted as the e-HMI proposition in Ackermann et al. (2019)]. This can also be measured through the error probability (i.e., the probability of incompatible responses), that is, decisions that were not in agreement with the e-HMI display.

When examining an e-HMI, it is essential to explore learnability. Learnability was found to significantly affect users adopting new technology and on user satisfaction from a product (Noel et al., 2005). Also, it was found that learnability directly influences safety when considering drivers (Noel et al., 2005). When investigating the learnability of a pedestrian's interaction with a FAV, in a WoZ field experiment, researchers found a learning curve over time but in a rather limited form as the authors based the learning on rating questionnaires over time (Faas et al., 2020). Researchers investigated learnability with a single item: the participant agreement with the statement, "It is easy to learn that the light signal on the vehicle indicates yielding" (strongly disagree – strongly agree) while comparing steady, flashing, and sweeping light signals.

The current study aims to investigate factors that influence pedestrians' understanding of a FAV's intention by looking at their decisions and scanning patterns when aiming to cross the road, in fixed simulated scenes from the perspective of the pedestrian, in general, and over time. The factors examined are related to the characteristics of the e-HMI, color, message type (advice or status message) and modality (text or symbol), and the crossing context; FAV size and distance from the crossing place. Also, using eye-tracking to measure pedestrians' visual attention distribution while deciding to cross is common in pedestrian behavior studies (e.g., Tapiro et al., 2016). Explicitly, it can indicate whether pedestrians looked at the e-HMI and for how long before the decision to cross or not. Field research investigated pedestrians' gaze patterns, but only with a manual car that did not include e-HMI (Dey et al., 2019). Another research reported a negative correlation between pedestrians' subjective understanding of the FAV intention and their gaze fixation duration (Liu et al., 2020). Furthermore, to our knowledge, the interaction between the crossing decision making (to cross or not cross) and pedestrians' gaze behavior on the FAV's e-HMI is yet to be investigated in general and learnability over time. Also, with regard to the measurement of response time and error probability.

The following hypotheses are suggested: h1- the e-HMI's proposition would lead pedestrians to make more compatible decisions, particularly when it conflicts with the crossing conditions (e.g., short distance*)*. This hypothesis is based on previous contradicting findings regarding what affects pedestrians to cross in the FAV world like distance (Clamann et al., 2017) or e-HMI proposition (Holländer and Butz, 2019). *h2-* Advice message would reduce error probability compared to status message (Ackermann et al., 2019). *h3-* is regarding learnability, we expect error probability and response time to reduce over time regardless of the crossing context and e- HMI display characteristics due to learnability.

MATERIALS AND METHODS

Participants

Twenty students aged 21–34 (M = 26, SD = 3, 11 females) participated in the experiment. One participant's data were excluded due to technical problems. As compensation, seven participants received course credit and 13 a payment of $10. All participants had normal contrast sensitivity and visual acuity of at least 6/6. Participants were free to withdraw from the study at any time.

Apparatus

Experimental Environment

The study was conducted at the Eye Tracking laboratory using a desktop test station computer with a 22″ screen. Participants were situated approximately 70 cm from the screen. The Gaze point eye tracker was located below the screen (see **Figure 1**).

Fixed Scene Generation

One hundred and eight fixed scenes were generated using the VT-MAK VR tools[1] with a typical local city's 3D terrain model. The crossed road was a one direction two-lane urban road. To add realism to the scene, the city's typography included buildings, light posts, vegetation, etc (**Figure 2**). The images were taken from the pedestrian's perspective as if standing on the curb and looking to the left before crossing the street. Each image included

[1] https://www.mak.com/

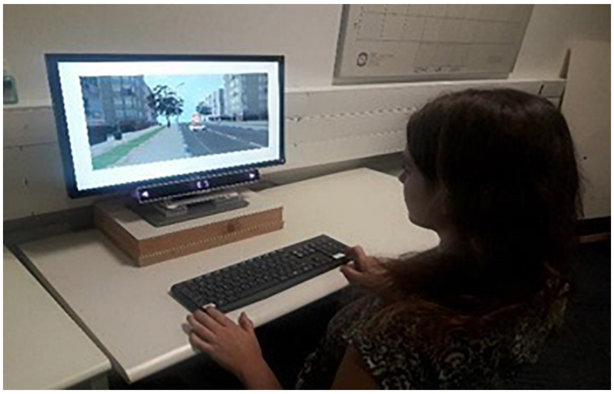


FIGURE 1 | The experimental testbed, consisting of a 22″ screen, the Gaze point eye-tracking system and the keyboard to collect participants' responses.

a combination of a single FAV (small or big) on the closer lane, either far (20 m) or close (9 m) to the pedestrian's crossing point in the simulation. The e-HMI size in the far distance was 0.9 cm × 0.9 cm and in the close distance 20 cm × 20 cm. The e-HMI was located on the roof of the car (this location was found to be very useful in previous research (Bazilinskyy et al., 2020). It included a sign that could convey either a written message (text) or a symbolized message (see **Figures 2**, **3**). Also, the message content could be a status message ("Slowing" or "Driving") or advice message ("Cross" or "Don't Cross"). Also, the e-HMI background color was green or red. I previous research, it was found that color convention helped pedestrians understand the FAV intention; that is, a green e-HMI indicated it was s safe to cross, and the red implies that it was unsafe to cross (Rouchitsas and Alm, 2019; Bazilinskyy et al., 2020). Besides, baseline images without the e-HMI were created, with a variation of car size and crossing distance (for the content of the entire images, see **Supplementary Appendix 1**).

Eye-Tracking System

The Gaze point eye-tracking system was used to measure pupil diameter and gaze direction with an accuracy visual angle of 0.5–1 degree (**Figure 1**). The system uses an eye camera and an infra-red eye illuminator to sample a close image pupil at a sample rate of 60 Hz.

Road Crossing Task

Each participant took part in three consecutive sessions (**Figure 4**). In each session, participants were asked to observe 36 consecutive crossing scenes and decide for each one, as quickly as possible, if it was safe to cross the road or not. The decision was made by selecting the "Safe to cross" or "Not safe to cross" designated keyboard buttons.

Dependent Variables

Estimated Error Probability

An error was defined as the incompatibility of the participant's selection (whether to cross or not) with the sign meaning (as *a priori* defined). If the selection had not the same value (safe/unsafe), it was counted as an error. In the model, we predicted the estimated error probability. Within the images that had no e-HMI, the incompatible response was unknown, and therefore, the error probability was undefined.

Response Time

Time from the moment the image was displayed until the participant pressed a decision button.

Eye-Tracking Measures

Total fixation duration and the total number of fixations on the e-HMI. A fixation was defined as a period of at least 100 ms that the eyes remain relatively still. The gaze data is based on the position variance technique (Jacob, 1995), that is, a sequence of gaze data estimates spatially located within a local region are determined to belong to the current fixation, while subsequent data outside of this local region is identified as the beginning of a new fixation. The fixations counted were only the ones within the area of interest (AOI), which was defined as the e-HMI sign (**Figure 5**).

Learnability

Learnability was defended as the improvement in performance over time, from the trial to trial, that is, the reduction in the error probability, response time, and the number of fixations.

Subjective Measurements

A written explanation of the sign meaning and rating its comprehension level (on 10-point rating scale), followed by an open interview on how each participant made their crossing decisions.

Experimental Design

A within-subject design. The following independent variables were defined: e-HMI (included/none), message type (status message/advise message), modality (text/symbol), car size (big/small), color (red/green) and car distance (close/far), altogether a 2^6 factorial design.

Procedure

Participants were invited individually to the lab for approximately 30 min. Following instructions and signing a consent form, they performed visual acuity and contrast sensitivity tests (Ginsburg, 1984). Next, the eye calibration was done. After calibration, participants performed a short practice of the road crossing task with five baseline images (no e-HMI). The experiment was divided into three consecutive sessions. After each session, there was a 30 s break. The sessions and the images within them were given in random order. Sessions included images with all combinations of car size, distance from the crossing place, and e-HMI content options. Each session contained four baseline images. Throughout the experiment, each image variation appeared three times with slight variations of the surrounding urban crossing road environment (e.g., building facade). Following the three sessions, participants were asked to explain each sign's meaning and rate their comprehension level. Then an open interview was conducted. Then an open interview was conducted. The experimental flow is described in **Figure 4**.

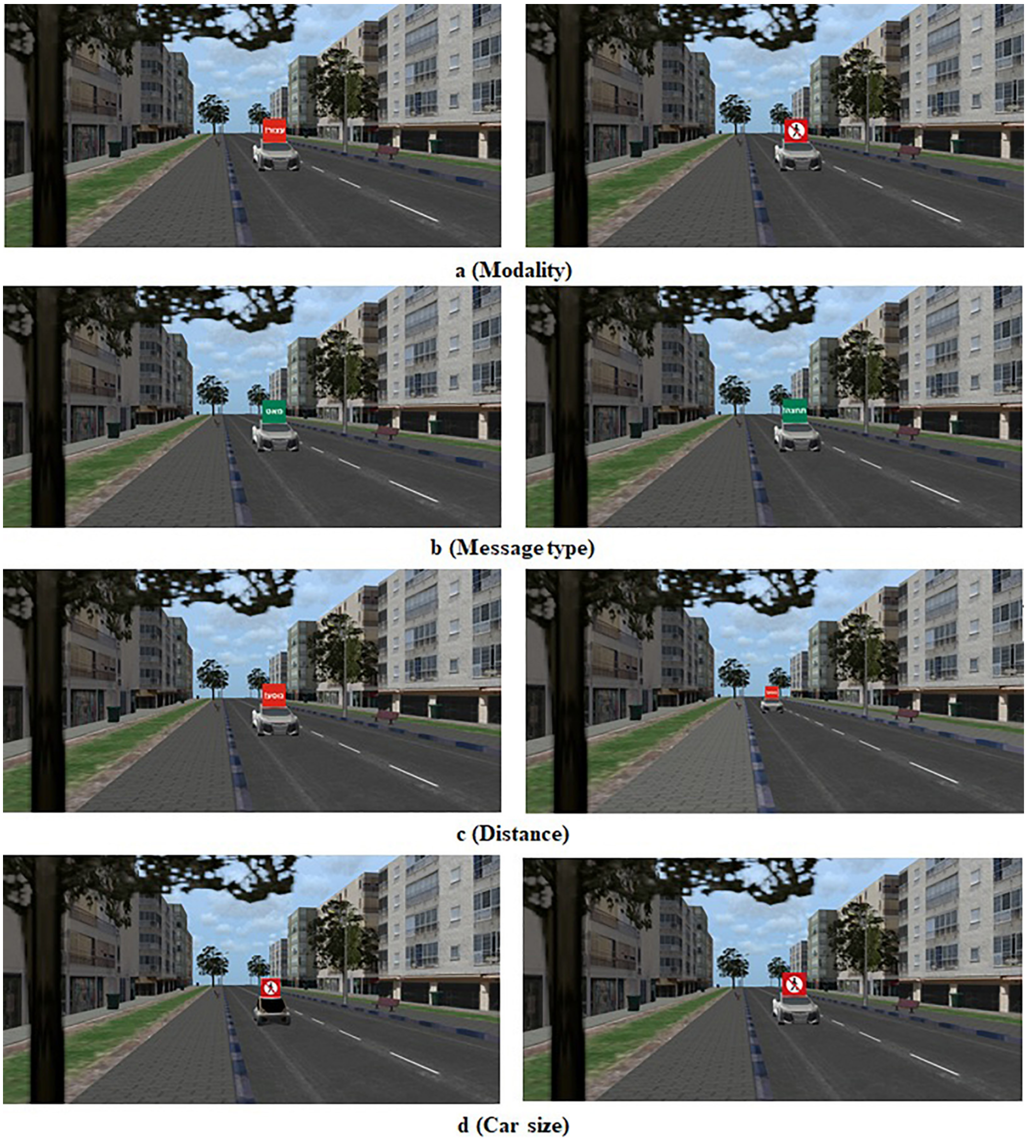


FIGURE 2 | Sample crossing scenes, as seen from the perspective of the pedestrian. Each row **(a–d)** demonstrates an examined factor. **(a)** Modality: left- Text, right- Symbol. **(b)** Message type: left- Status, right- Advice. **(c)** Distance: left- close, right- far. **(d)** Car size: left – Small (Kancil),right- Large (Audi).

Eye-Tracking and Area of Interest (AOI) Definition

Eye movements and fixations data were collected and synced with the experimental timeline for each crossing scene through a designated software. Once the experiment ended, the software was used to determine whether the fixations were within the defined AOI and only that data (within the AOI) was summed per image. In the baseline images (no e-HMI), the entire image (car and environment) was defined as the AOI. For the rest, the area around the e-HMI was defined as the AOI. Its exact size was defined as the minimum size that can be expected around 0.5–1 degree in a high-end eye-tracker when the computer distance from the participant was 68.6 cm (Bojko, 2013). Hence, the AOI is defined as the multiplication of each side in the e-HMI (sign) frame length by 1.43 (see **Figure 5**).

Data Analysis

A Wilcoxon test was performed to examine whether there was a difference in the dependent variables between the compatible responses and the incompatible ones. Next, a Generalized Linear Mixed Model (GLMM) was used to analyze the effect of the independent variables (message type, modality, e-HMI background color, car size, and distance) on the estimated error probability (incompatible responses, a binary logistic regression within the GLMM) of all responses over time, to examine learnability from trial to trial. Then, the effects of the independent variables were further examined on response

FIGURE 3 | The e-HMI Messages; each column displays the message type (status or advice) in both written (in Hebrew) and symbolic messages. The background color is compatible with the message meaning according to the color convention.

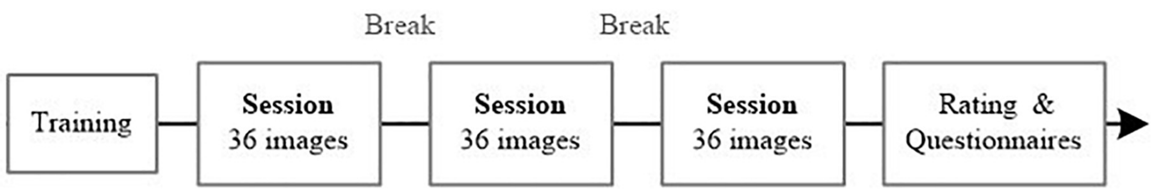


FIGURE 4 | The experimental flow.

time (ln transformed response time, a normal regression within the GLMM) and the number of fixations on the AOI over time both for all responses and for the compatible response [(ln transformed number +1), a normal regression within the GLMM]. Beyond the fixed effect, participants and image numbers were included as random effects to account for individual differences among participants and variation among images. Utilizing a stepwise process, only the main effects and the significant interactions were included in the final model. All three models used the same predicting effects- message type, modality, e-HMI background color, car size, and car distance. The final model included only significant effects or interaction related parameters.

FIGURE 5 | Defining the AOI around the e-HMI.

RESULTS

Crossing Decisions, Response Time, and Eye-Tracking Data

Overall, 75% (1401 out of 1867) of the decisions were compatible, and 25% (466) were incompatible. Wilcoxon tests revealed that the number of fixations for compatible responses was significantly smaller (Mean = 3.46, SD = 2.67) compared to incompatible ones (Mean = 3.85, SD = 3.02, $p < 0.001$). In addition, response time for the compatible responses was significantly shorter (Mean = 1.27 s, SD = 0.97) compared to incompatible responses (Mean = 1.50, SD = 1.25, $p < 0.001$). **Table 1** shows that when the FAV is close and the e-HMI background is red, response times and the number of fixations were about twice as high in the incompatible responses compared to the compatible ones. Delving into the details, only 14 responses of all trials were incompatible (compared to 452 that were compatible in the same conditions) when the e-HMI background was red, and the car was close, and a single participant made 8 of them. This participant had dispersed response times (0.49–12.04 s), including two considerably longer ones (9.36 and 12.04 s) that occurred at the beginning of the experiment. Longer

TABLE 1 | Crossing decisions number of fixations and response time.

Measurements	Compatible crossing decision				Incompatible crossing decision			
	Red		Green		Red		Green	
	Close	Far	Close	Far	Close	Far	Close	Far
Number of fixations								
Mean	3.25	3.52	4.39	3.13	6.14	3.57	3.78	4.06
Median	3	3	3	3	4	3	3	3
Confidence interval	0.006	0.011	0.015	0.007	0.09	0.02	0.01	0.02
Response Time [sec]								
Mean	1.14	1.44	1.48	1.15	3.00	1.43	1.38	1.72
Median	0.91	1.10	1.08	0.94	1.74	1.16	1.27	1.27
Confidence interval	0.00	0.01	0.01	0.00	0.06	0.006	0.01	0.01

FIGURE 6 | Sample images with high error rates (the right image received 50% error rates and the left 74%). The commonality amongst them was the green background e-HMI and close distance (for both symbol and text modality).

response time may imply that when pedestrians take a risk and decide to cross in close distance and red e-HMI, they tend to hesitate before crossing. It may suggest that they understood the risks of crossing and decided to cross despite them.

Images That Received High Incompatible Responses

Twenty images out of the 96 images yielded an error rate of 45% or higher per image. All of these images had a green background. The FAV distance was close in eighteen of them, which implies that according to the FAV's e-HMI, pedestrians could have crossed, but they decided not to (sample images are shown in **Figure 6**). One specific symbol message (the car slowing status symbol, see **Figure 6**) on the right) received the highest error rate. This symbol was also the lowest-ranked in the comprehensive subjective ratings (average score of 3.6 out of 10).

Estimated Error Probability in General and Over Time

Distance and Color

A significant interaction was found between the color the distance in the estimated error probability [χ^2 (df = 1) = 185.3, $p < 0.001$]; see **Table 2** and **Figure 7** Top. In the close distance, there was a significant difference in the estimated error probability between the e-HMI colors, compared to the far distance. *Post hoc* (Tukey's-HSD) analyses revealed that in the close distance, the red background e-HMI had a significantly higher probability of compatible responses compared to the green background e-HMI

TABLE 2 | The effect of e-HMI related factors (message type, modality, color) and crossing context factors (car size, distance, and order) on the Estimated error probability (GLMM).

	Estimated error probability	
Factors	**χ^2 (df = 1)**	***p*-value**
Order	11.07	0.00***
Message type	3.74	0.053
Modality	8.56	0.003**
Color	13.11	0.00***
Car size	0.27	0.60
Distance	32.31	0.00***
Message type * Modality	17.59	0.00***
Car size * Distance	17.27	0.00***
Message type * Color	4.9	0.03*
Modality * Color	9.52	0.002**
Car size * Color	7.88	0.005**
Distance * Color	185.3	0.00***

*The χ^2 ratio is a measure of the overall significance of the model. *p < 0.05, **p < 0.01, ***p < 0.001.*

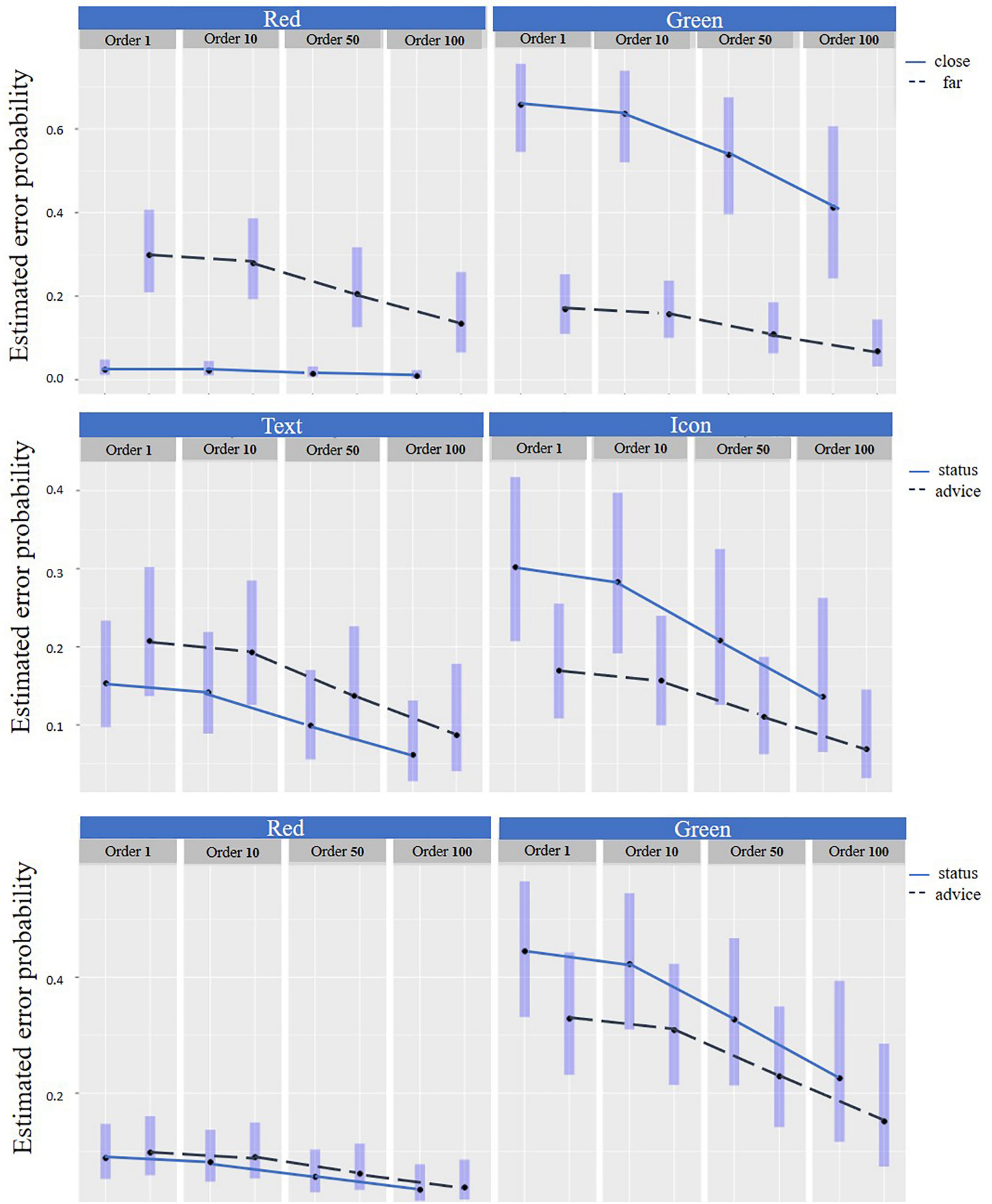


FIGURE 7 | Estimated error probability. Top: By distance and color over time. Middle: By message type and modality over time and Bottom: By message type and color over time. Note: in the graphs, for visualization only, the estimated error probability was sampled in 4 chronologic locations (order) – after the first impression (image #1), at the beginning (after image #10), middle (following image #50) and at the end of the experiment (image #96), and the estimated error probability average is displayed for each sample.

($z = -13.63$, $p < 0.001$). An opposite trend was found in the far distance; the estimated error probability in the green e-HMI was much lower from the red in the far one ($z = 4.1$, $p < 0.001$). Also, overall, results revealed a strong interaction between the fixed image order (each image had a random chronological location in each trial) and the estimated error probability. It was found that the estimated error probability reduced over time for each color – distance combination [χ^2 (df = 1) = 11.07, $p < 0.001$]; that is, there was a learning effect over time. See **Figure 7** top.

Modality, Message Type, and Color

It was found that there was an interaction between the message type and the color [χ^2 (df = 1) = 4.90, $p < 0.05$], see **Table 2**. *Post hoc* (Tukey's-HSD) analyses revealed that there was no significant difference in the estimated error probability in the red background e-HMI for the different message types ($z = -0.51$, $p = 0.6$) (**Figure 7** Bottom). This finding means that when there was a red background e-HMI, pedestrians tended to decide not to cross in both messages type. However, in the green background e-HMI, pedestrians had higher errors when they received status messages compared to advice messages ($z = 2.9$, $p < 0.05$) (**Figure 7** Bottom). Also, there was an interaction between the message type and the modality [χ^2 (df = 1) = 17.59, $p < 0.001$] (**Figure 7** Middle). In the status message, the estimated error probability for text messages was lower than for symbol messages ($z = -4.5$, $p < 0.001$). A learning effect over time that is being reflected by the reduction of the estimated error probability seems to have a similar pattern for each message type-modality combination (**Figure 7** Middle) and each message type-color combination (**Figure 7** Bottom).

Response Time in General and Over Time

Distance and Color

There was an interaction between the distance and color for response time of the compatible responses [$F(1,1389) = 34.0$, $p < 0.001$] see **Table 3**. *Post hoc* (Tukey's-HSD) analyses revealed that in the close distance, response times for compatible responses were shorter for the red background e-HMI color (Mean = 1.14 s, SD = 0.71) compared to the green (Mean = 1.48 s, SD = 1.18, $p < 0.001$), as shown **Figure 8**. In the far distance, response time was shorter when the e-HMI background color was green (Mean = 1.15 s, SD = 0.69) compared to red (Mean = 1.44, SD = 1.29, $p < 0.05$). Overall, response time was shorter over time for each combination of distance-e-HMI background color, for all responses [$F(1,1389) = 38.2$, $p < 0.01$] and for compatible responses [$F(1,1389) = 37.01$, $p < 0.01$], as seen in **Table 3** and **Figure 8**. Thus, there was a learning effect over time. Moreover, the learning effect shown through the reduction of response time seems to have a similar pattern for all four distance-color combinations.

TABLE 3 | The effect of color and crossing context factors (car size, distance, and order) on Response time (ln transformed) and number of fixations for all responses and the compatible responses (GLMM).

	Response time		Number of fixations	
	All responses	Compatible responses	All responses	Compatible responses
Factors	$F(1,1855)$	$F(1,1389)$	$F(1,1855)$	$F(1,1389)$
Order	38.20**	37.01***	12.00*	11.18*
Car size	–	–	6.55*	6.40*
Distance	1.12	0.17	16.60***	15.73***
Color	2.74	4.32*	5.58*	5.54*
Distance * Color	39.06***	34.0***	20.62***	20.42***

The F ratio is a measure of the overall significance of the model.$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.*

Number of Fixations in General and Overtime

Distance and Color

In general, there was an interaction between the color, distance, and the number of fixations [$F(1,1389) = 20.42$, $p < 0.001$] for the compatible responses (**Table 3** and **Figure 9**). *Post hoc* analysis revealed that in the close distance, there were less fixations on the red e-HMI background (Mean = 3.25 SD = 2.27) compared to the green one (Mean = 4.39, SD = 3.43, $p < 0.001$). Findings reveal that for both colors, the number of fixations was reduced over time [$F(1,1389) = 11.18$, $p < 0.05$], which indicates upon learnability. However, in the close distance, the number of fixations was reduced more notably compared to the far distance. In other words, the learnability overtime was more significant in the close distance compared to the far distance (**Figure 9**).

Rate of Fixations per Millisecond

One can rightfully argue that the number of fixations will increase if response time increases, which is why it is also necessary to look at the rate of fixations. This is a similar analysis to the one in 3.5.1 of the number of fixations but now with response time as a covariate, leading to an examination of the rate of fixations per millisecond. If the response time as a covariate in the model is statistically significant, it implies that the fixation rate changes over time. Depending on the estimated mean of this covariate, one can identify the rate of change in the number of fixations over time. If the rate estimate is less than one, it means that as response time increases, the increase in the number of fixations decreases (indicating that fixations are becoming longer). Oppositely, if the rate estimate is larger than one, it indicates that the number of fixations increases as the time progresses (indicating a more erratic movement of the eyes).

The final statistical model yielded the following significant effects: Learnability over time remained significant [$F(1,1378) = 7.24$, $p < 0.007$], main effects for car size [$F(1,1374) = 4.54$, $p < 0.033$] distance [$F(1,1375) = 24.09$, $p < 0.001$], and an interaction for color and modality [$F(1,1375) = 8.98$, $p < 0.003$]. Most importantly for this analysis, response time as a covariate was statistically significant [$F(1,1392) = 1272.03$, $p < 0.00001$], indicating that indeed the rate of fixations changes over time. The estimated rate was 0.663 (SE = 0.018) thus, less than 1, indicating that as response time increases, the number of fixations increases too, but at a lower rate. Hence, most likely fixations are becoming longer in time.

DISCUSSION

This study aimed to explore which parameters affect pedestrians' understanding of the FAV's intentions as expressed in crossing

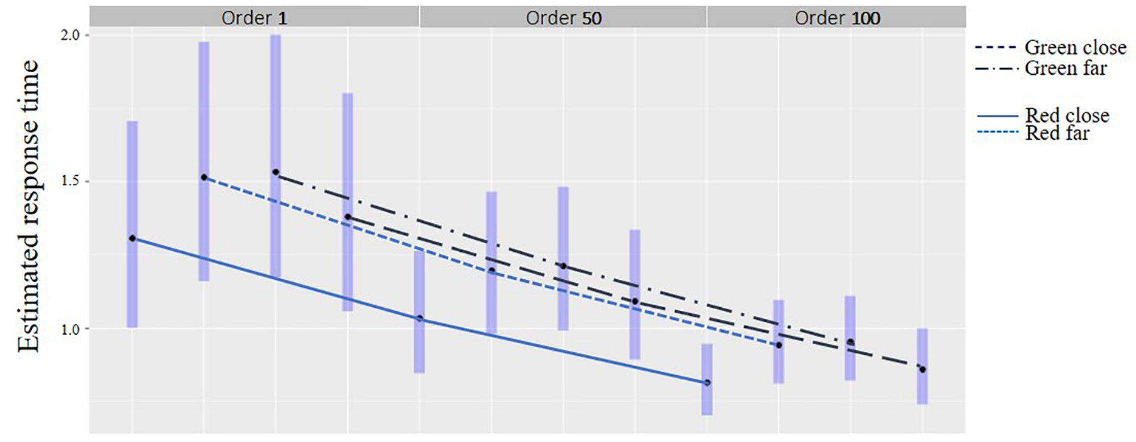


FIGURE 8 | Response time for compatible responses by distance and color over time.

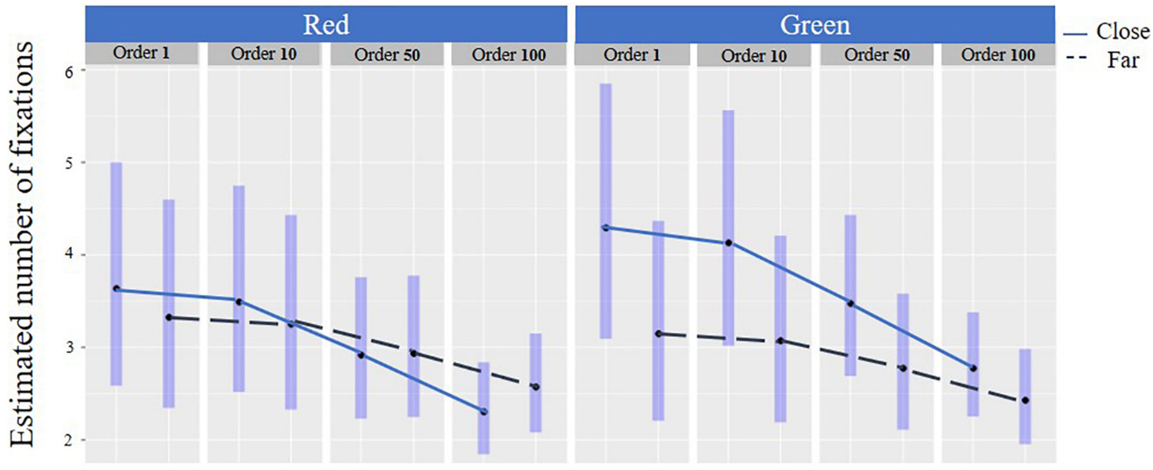


FIGURE 9 | Number of fixations for compatible responses: By distance and color over time.

decisions of participants on fixed crossing scenes, as well as the change of crossing decisions over time (learnability). Results revealed that pedestrians fixated on the FAV's e-HMI, in line with previous research (Dey et al., 2019; Eisma et al., 2020). But, unlike what has been suggested in a previous study (Holländer and Butz, 2019), pedestrians do not always base their decision on the e-HMI proposition as demonstrated through the e-HMI background color, or message type - instruction or status. It was found that in 25% of the time, pedestrians made crossing decisions that were incompatible with what the e-HMI proposed. From observing the images that got the most incompatible responses, one can attain that in those, pedestrians made their decisions based on the FAV distance from the crossing place and decided not to cross when the FAV was close. Yet, the e-HMI background was green and proposed to cross. Also, it was found that when the e-HMI background was red and the distance was far, pedestrians sometimes decided to take the risk and cross (**Figure 7** Top). This finding is in line with previous research that explored the effect of distance on pedestrians' crossing decision (Clamann et al., 2017).

Nevertheless, when pedestrians made the compatible crossing decision when the FAV was close and the e-HMI was green, they lingered and did not decide to cross immediately. These findings were pronounced by longer response times and a higher number of fixations compared to green background e-HMI in the far distance (**Figure 8** and **Table 3**). Also, when the FAV was far, and a red background e-HMI appeared, pedestrians also hesitated and took some time to decide (**Table 3** and **Figure 9**). These results imply that, most likely, pedestrians base their decisions on a combination of distance and the e-HMI proposition. These findings can be explained by color conventions and distance. When the color convention fits the pedestrian's expectations and risk due to the car's distance, fewer fixations were needed. However, in cases where the e-HMI color convention conflicted with pedestrians' expectations, it was necessary to further gaze on the e-HMI to understand the message and take more time to decide (**Figure 8** and **Table 3**). These findings confirm *h1* that the e-HMI can help make the compatible decision when there are conflicts but not always, as shown in the 25% of the incompatible responses.

Learnability

Overall, there was a learning effect over time for the various fixed effects. This was reflected in the reduction in error probability over time (**Figure 7** and **Table 2**), as well as in the shortening of response times and the reduction in the number of fixations over time in all conditions (**Figures 8**, **9** and **Table 3**). The learning curve seems to have a similar pattern for all combinations of conditions crossing conditions (e.g., distance-color combinations). These findings are aligned with previous research findings regarding the learning effect over time (Faas et al., 2020) and strengthen them. Thus we can confirm *h3* that there is a learning effect regardless of the crossing context and e-HMI display content.

Message Type, Modality, and Color

Results revealed that the green background e-HMI for advice message tended to be more intuitive since it had a lower estimated error probability than the green background for e-HMI status message (see **Table 2**). Also, in the advice message, there was no difference between the two modalities (see **Figure 7** Bottom). These results confirm *h2* and can be explained by the fact that pedestrians today are more familiar with advice messages, in both modalities, and not familiar with status messages in general and with regard to FAVs. Further, it was easier to express a status message through text than through symbols, but this may change in the future when symbols become more standardized and common.

This study sheds more light on the contradicting findings of previous studies and emphasizes that pedestrians are not yet in a stage where they trust FAV e-HMI entirely in contrast to some findings (Holländer and Butz, 2019). However, they do not ignore it (in contrast to Clamann et al., 2017; Mahadevan et al., 2018). This study revealed, from analyzing the number of fixations and response time, that pedestrians tend to decide for themselves whether to cross the road based on a combination of the FAV distance from the crossing place and the e-HMI background color and instructions.

Our study highlights the importance of the e-HMI and how it may affect pedestrians' decision to cross. However, several limitations must be noted. A major limitation is in the crossing conditions of only one FAV and one pedestrian at a specific time, unlike the real world. Another limitation refers to the form of presentation, that is, the fixed scenes. Although this form allows us to examine pedestrian behavior parameters (such as understanding) more deeply, it ignores other parameters that are associated with the dynamicity of the road crossing task. Lastly, the study population included a convenience sample of students. Future studies should examine our findings in dynamic scenarios and with more complex and varied crossing conditions such as multiple FAVs on the road, different car types, etc. Further, pedestrians' decisions may be influenced by the presence of other pedestrians, which we did not examine. Last but not least, as shown in pedestrian studies (e.g., Tapiro et al., 2016, 2020), findings must be further evaluated across cultures and with regard to children and older adults. Finally, while this study addressed learnability, we still do not know enough about how and if the eHMI proposition will lead pedestrians to behave in compliance with its recommendation even in conflict situations, to establish this, we need to examine the learnability curve further using varied learnability inflators, such as system errors, misses and false alarms, varying trust level, etc.

CONCLUSION

Over time, learning was apparent in response times and gaze for crossing context and e-HMI characteristics combinations for the compatible and all responses. Therefore, it is essential to provide e-HMI designs that will minimize error probability and provide fast response times, and need for only a minimal number of fixations on the e-HMI from the first phase. The existence of learning is encouraging, as it implies that crossing performance can be improved over time. Further, color conventions play a significant role in pedestrians crossing decisions today, and they will probably influence decisions in the FAV world as well, at least in the upcoming decade. This finding is important and emphasizes that it is essential to adhere to existing color convention, in line with Bazilinskyy et al. (2020) and not use neutral colors for all FAV messages, as some researchers have suggested (Dey et al., 2020). Yet, pedestrians also considered the FAV's distance from the crossing place when deciding to cross, especially in conflict situations. This skill of estimating the risk of the crossing from the distance of the vehicle is necessary for today's pedestrians. Still, it may diminish in the future FAV world if and when pedestrians will over trust the e-HMI and base their decisions solely on its recommendations and status. This research used fixed scenes, which allows examining in-depth, how pedestrians related to the crossing scene over time in the FAV world and how the e-HMI influenced their decision. However, to extend these findings, it is necessary to conduct further studies with dynamic various crossing complexities, examine further learnability inflators, and include diverse, multicultural populations, such as the elderly and children.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Department Internal Review Board, Department of Industrial Engineering and Management, Ben-Gurion University of the Negev. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.585280/full#supplementary-material

REFERENCES

Ackermann, C., Beggiato, M., Schubert, S., and Krems, J. F. (2019). An experimental study to investigate design and assessment criteria: What is important for communication between pedestrians and automated vehicles? *Appl. Ergonom.* 75, 272–282. doi: 10.1016/j.apergo.2018.11.002

Ackermans, S., Dey, D., Ruijten, P., Cuijpers, R. H., and Pfleging, B. (2020). "The Effects of Explicit Intention Communication, Conspicuous Sensors, and Pedestrian Attitude in Interactions with Automated Vehicles," in *Prooceeding of the CHI 2020 - Conference on Human Factors in Computing Systems* Hawaii. 1–14.

Bazilinskyy, P., Kooijman, L., Dodou, D., and De Winter, J. C. F. (2020). How should external Human-Machine Interfaces behave? Examining the effects of colour, position, message, activation distance, vehicle yielding, and visual distraction among 1,434 participants. *Manuscript submitted for publication*. doi: 10.13140/RG.2.2.11345.56164

Bojko, A. (2013). *Eye Tracking the User Experience: A Practical Guide to Research*. New York, NY: Rosenfeld Media.

Clamann, M., Aubert, M., and Cummings, M. L. (2017). "Evaluation of vehicle-to-pedestrian communication displays for autonomous vehicles (No. 17-02119)," in *Proceeding of the 96th Annual Transportation Research Board*, Washington DC.

Deb, S., Strawderman, L. J., and Carruth, D. W. (2018). Investigating pedestrian suggestions for external features on fully autonomous vehicles: a virtual reality experiment. *Trans. Res. F: Traffic Psychol. Behav. 59*, 135–149. doi: 10.1016/j.trf.2018.08.016

Dey, D., Habibovic, A., Pfleging, B., Martens, M., and Terken, J. (2020). "Color and Animation Preferences for a Light Band eHMI in Interactions Between Automated Vehicles and Pedestrians," In *Proceeding of the* CHI Conference on Human Factors in Computing Systems, Hawaii, 1–13. doi: 10.1145/3313831.3376325

Dey, D., Martens, M., Wang, C., Ros, F., and Terken, J. (2018). "Interface concepts for intent communication from autonomous vehicles to vulnerable road users," In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, Toronto, ON, pp. 82-86. doi: 10.1145/3239092.3265946

Dey, D., Walker, F., Martens, M., and Terken, J. (2019). "Gaze patterns in pedestrian interaction with vehicles: Towards effective design of external human-machine interfaces for automated vehicles,". *Proceeding of the 11th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Toronto, ON: AutomotiveUI), 369–378. doi: 10.1145/3342197.3344523

Eisma, Y. B., Bergen, S. VanBrake, S. M., Hensen, M. T. T., Tempelaar, W. J., and DeWinter, J. C. F. (2020). External Human – Machine Interfaces: The Effect of Display Location on Crossing Intentions and Eye Movements. *Information* 11:13. doi: 10.3390/info11010013

Faas, S. M., Kao, A. C., Baumann, M., & Ag, M. (2020). "A Longitudinal Video Study on Communicating Status and Intent for Self-Driving Vehicle – Pedestrian Interaction,"In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Media Technology and Interaction Design, MID), 1–14.

Ginsburg, A. P. (1984). A new contrast sensitivity vision test chart. *Am. J. Optom. Physiol. Opt.* 61, 403–407. doi: 10.1097/00006324-198406000-00011

Holländer, K., and Butz, A. (2019). "Overtrust in External Cues of Automated Vehicles: An Experimental Investigation," In *Proceedings of the* 11th International Conference (Automotive). doi: 10.1145/3342197.3344528

Jacob, R. J. (1995). "Eye tracking in advanced interface design," in *Virtual Environments and Advanced Interface Design* (New York, NY: Oxford University Press), 258–288.

Liu, H., Hirayama, T., Morales, L. Y., and Murase, H. (2020). What is the gaze behavior of pedestrians in interactions with an automated vehicle when they do not understand its intentions? *arXiv* [Preprint]. arXiv:2001.01340. Available online at: https://arxiv.org/abs/2001.01340

Mahadevan, K., Somanath, S., and Sharlin, E. (2018). "Enabling Pedestrian Communication with Autonomous Vehicles," In *Proceedings of the CHI 2018 Workshop Interacting with Autonomous Vehicles: Learning from Other Domains*, (CHI EA 18: Extended Abstracts of the 2018).

Noel, E., Nonnecke, B., and Trick, L. (2005). "A comprehensive learnability evaluation method for in-car navigation devices,"In *Proceedings of the SAE 2005 World Congress and Exhibition*, Detroit, MI. doi: 10.4271/2005-01-1605

Rasouli, A., Kotseruba, I., and Tsotsos, J. K. (2018). Understanding Pedestrian Behavior in Complex Traffic Scenes. *IEEE Trans. Intell. Veh. 3*, 61–70. doi: 10.1109/tiv.2017.2788193

Rouchitsas, A., and Alm, H. (2019). External Human–Machine Interfaces for Autonomous Vehicle-to-Pedestrian Communication: A Review of Empirical Work. *Front. Psychol.* 10:2757. doi: 10.3389/fpsyg.2019.02757

Tapiro, H., Borowsky, A., Oron-Gilad, T., and Parmet, Y. (2016). "Where do older pedestrians glance before deciding to cross a simulated two-lane road? A pedestrian simulator paradigm," In *Proceedings of the Human Factors and Ergonomics Society*, (Human Factors and Ergonomics Society), 11–15. doi: 10.1177/1541931213601003

Tapiro, H., Oron-Gilad, T., and Parmet, Y. (2020). Pedestrian distraction: The effects of road environment complexity and age on pedestrian's visual attention and crossing behavior. *J. Safe. Res.* 72, 101–109. doi: 10.1016/j.jsr.2019.12.003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 18 February 2021
doi: 10.3389/fpsyg.2021.596038

Classification of Drivers' Workload Using Physiological Signals in Conditional Automation

Quentin Meteier[1], Marine Capallera[1], Simon Ruffieux[1], Leonardo Angelini[1], Omar Abou Khaled[1], Elena Mugellini[1], Marino Widmer[2] and Andreas Sonderegger[3]*

[1] HumanTech Institute, University of Applied Sciences of Western Switzerland, Haute École Spécialisée de Suisse Occidentale, Fribourg, Switzerland, [2] Department of Informatics, University of Fribourg, Fribourg, Switzerland, [3] Bern University of Applied Sciences, Business School, Institute for New Work, Bern, Switzerland

The use of automation in cars is increasing. In future vehicles, drivers will no longer be in charge of the main driving task and may be allowed to perform a secondary task. However, they might be requested to regain control of the car if a hazardous situation occurs (i.e., conditionally automated driving). Performing a secondary task might increase drivers' mental workload and consequently decrease the takeover performance if the workload level exceeds a certain threshold. Knowledge about the driver's mental state might hence be useful for increasing safety in conditionally automated vehicles. Measuring drivers' workload continuously is essential to support the driver and hence limit the number of accidents in takeover situations. This goal can be achieved using machine learning techniques to evaluate and classify the drivers' workload in real-time. To evaluate the usefulness of physiological data as an indicator for workload in conditionally automated driving, three physiological signals from 90 subjects were collected during 25 min of automated driving in a fixed-base simulator. Half of the participants performed a verbal cognitive task to induce mental workload while the other half only had to monitor the environment of the car. Three classifiers, sensor fusion and levels of data segmentation were compared. Results show that the best model was able to successfully classify the condition of the driver with an accuracy of 95%. In some cases, the model benefited from sensors' fusion. Increasing the segmentation level (e.g., size of the time window to compute physiological indicators) increased the performance of the model for windows smaller than 4 min, but decreased for windows larger than 4 min. In conclusion, the study showed that a high level of drivers' mental workload can be accurately detected while driving in conditional automation based on 4-min recordings of respiration and skin conductance.

Keywords: automated driving, classification, driver, workload, physiology, secondary task, machine learning

OPEN ACCESS

Edited by:
David Shinar, Ben-Gurion University of the Negev, Israel

Reviewed by:
Bruce Mehler, Massachusetts Institute of Technology, United States
Karel Brookhuis, University of Groningen, Netherlands

Correspondence:
Quentin Meteier
quentin.meteier@hes-so.ch

Specialty section:
This article was submitted to Performance Science, a section of the journal Frontiers in Psychology

Received: *18 August 2020*
Accepted: *20 January 2021*
Published: *18 February 2021*

Citation:
Meteier Q, Capallera M, Ruffieux S, Angelini L, Abou Khaled O, Mugellini E, Widmer M and Sonderegger A (2021) Classification of Drivers' Workload Using Physiological Signals in Conditional Automation. Front. Psychol. 12:596038. doi: 10.3389/fpsyg.2021.596038

1. INTRODUCTION

According to the National Highway Traffic Safety Administration (NHTSA), 2,935 fatal crashes occurred on U.S. roadways due to driver's distraction in 2017. This represents 9% of all fatal crashes (NHTSA, 2017). Performing a secondary task while driving is one cause that increases the risk to have an accident, among other factors such as fatigue, mood or demanding driving conditions.

The latter lead to hazardous drivers states as named by Darzi et al. (2018). To solve that issue, car manufacturers aim at reducing the rate of accidents by proposing an increasing level of automation in cars to support the driver. According to the Society of Automotive Engineers (SAE) classification (SAE, 2018), the next generation of vehicles that will emerge on our roads will be conditionally automated cars, corresponding to Level 3 of the SAE taxonomy. At this automation level, the driver will no longer be in charge of the main driving task, neither monitoring the environment. However, the car alerts the driver that he or she has to take over control of the car when the automation is reaching its limit. The commonly accepted approach is sending a takeover request (TOR) to the driver (Kim et al., 2019). Various ways of alerting the driver are being tested (Petermeijer et al., 2017), such as visual, auditory and haptic alerts, or a combination of those. Thus, the driver must be ready to take over control at any moment during the ride. The role of the driver in such a situation will switch quickly from passenger behind the wheel to driver. Besides, on the basis of decisions taken by the authorities concerned, drivers could be allowed to engage in a Non-Driving-Related Task (NDRT) during periods of conditionally automated driving. The driver might be out-of-the-loop if he or she is engaged in a NDRT. It was recently defined by Merat as being "not in physical control of the vehicle, and not monitoring the driving situation, or in physical control of the vehicle but not monitoring the driving situation" (Merat et al., 2019). The engagement of drivers in a NDRT would distract them from the supervision of the environment for which they are responsible. They could be distracted visually, orally, cognitively, or bio-mechanically (Pettitt et al., 2005). These are not exclusive and drivers' could be distracted in different ways at the same time. The distraction induced by performing a NDRT using another sensory channel might also increase the mental workload (MWL; Mehler et al., 2009).

Previous studies showed that performing NDRTs that involve various modalities affect the gaze behavior and takeover performance of drivers (Nakajima and Tanaka, 2017; Wandtner et al., 2018). To address this issue, Parasuraman et al. (2000) suggested that "well-designed information automation can change human operator mental workload to a level that is appropriate for the system tasks to be performed." If we want the drivers to safely engage in NDRTs, it is crucial to find a way to measure continuously their state and use this information to dynamically support the driver. Various types of measures that depict the operator's state could be used such as performance, subjective or physiological measures. Under real situations, it might not be the best option to rely on subject ratings for adapting the level of automation. Driving data were suggested in previous studies to show drivers' distraction and elevations of MWL induced by a NDRT in manual driving (Engström et al., 2005). However, this source of data cannot be used in conditionally automated driving since the car is performing the main driving task most of the time, except during takeover situations. Previous research has shown that increases of MWL and cognitive distraction can easily be detected with cameras using eye-tracking, face-based or EEG features (Li and Busso, 2013; Hogervorst et al., 2014). For practical issues, a EEG headset would not be comfortable to wear for drivers, therefore, we are not considering this signal in this study. When the driver's gaze is toward the windshield and facing the camera, these features could be useful for predicting driver's workload continuously. However, the driver's gaze often changes direction in the car. Thus, it may be difficult to use only these data sources in real driving conditions to measure driver cognitive load.

In this context, physiological signals seem to be the best option to evaluate continuously and non-intrusively changes in MWL of drivers while performing a secondary task in conditionally automated driving. Recent advances in technology allow for a recording of physiological signals with embedded sensors that drivers could wear in real-world environments such as wristbands, smartwatches, or smart clothes (Sonderegger, 2013; Angelini et al., 2014). Features computed from raw physiological signals could be used to classify the driver's state using recent machine learning techniques. This information could be used to adapt either the automation level or the interaction level between the driver and the car. Such a system would help reducing fatalities due to bad takeover behavior and performance and therefore increase safety on roads. Besides, this system could also increase the user experience in automated cars because drivers would be able to engage in their favorite activity during the ride. This is because, based on such information, an automated system could adapt the type of warning (e.g., display of a loud and startling sound in case of low activation of the driver vs. smooth visual hint in case of situations of high activation) or even decide not to send out a TOR because it calculated that there is not enough time to safely take over control under given conditions (e.g., travel speed, distance to the object, state of the driver).

2. RELATED WORK

2.1. Mental Workload

Being one of the most widely invoked concepts in human factors and ergonomics, MWL represents a topic of increasing importance in research and practice (Young et al., 2015). MWL can be explained in terms of the balance between the demands of a situation (task and environmental context) and the resources an individual has available to overcome the situation (Wickens, 2008). While task demands are generally referred to as stress, strain describes the impact of task demands on the human (Schlegel, 1993). MWL is generally defined as a multidimensional construct which is determined by task characteristics, operator characteristics (e.g., attentional resources, skills), and the environmental context (Young et al., 2015). One of the main reasons for the increasing interest in MWL lies within its link to human performance and hence the possibility to identify suboptimal workload conditions that might lead to stress, errors and incidents in the driving context (Brookhuis and De Waard, 2001). It is generally agreed upon that MWL can be considered a basic precursor of stress, errors, and accidents—it has been difficult however to establish an exact relationship between the concepts, which mainly might be due to difficulties in the accurate measurement of MWL (Young et al., 2015).

Three main approaches have been put forward for assessing MWL, including measures of task performance (primary and/or secondary task), subjective ratings based on questionnaires and physiological measures (Gawron, 2019). The first approach is based on techniques measuring task performance on a primary and a secondary task. While generally an acceptable level of performance in the primary task can be maintained in high workload conditions, performance on the secondary task is highly correlated with MWL since the secondary task is associated with the spare capacity unused for completion of the primary task (Young et al., 2015). The primary-secondary task paradigm has been shown to be a good indicator of MWL in experimental research. Its implementation however is linked with some rather severe drawbacks (e.g., artificial setup of test environment with a high need for standardization and control of the task scenarios; Fisk et al., 1986). In some driving studies, the performance on the primary task was measured using driving data (longitudinal and lateral parameters) and the secondary task performance was used as an indicator of MWL (Engström et al., 2005; Mehler et al., 2009). The second approach is to assess MWL as subjective state based on subjective ratings. This implies the assumption that humans are capable of evaluating and expressing the level of MWL they experience in a specific task after task completion. Some widely used questionnaires for measuring MWL subjectively are the NASA-TLX (Hart and Staveland, 1988) or the Rating Scale Mental Effort (Zijlstra and Doorn, 1985). Task load questionnaires are easy to apply and interpret but come along with some methodological issues which are due to the subjectivity of the measure and the retrospective bias of post-task assessments (Bulmer et al., 2004). The third approach to measure MWL is the assessment of physiological indicators. In this regard, two groups of physiological indicators can be differentiated, indicators of the autonomic nervous system and indicators of the central nervous system. Cardiovascular indicators (e.g., heart rate and heart rate variability) as well as electrodermal activity (e.g., tonic and phasic skin conductivity) are often referred to as useful indicators of MWL in research (De Waard, 1997). However, considerable drawbacks for the assessment of MWL via physiological parameters are the troublesome procedure of applying electrodes, the generally rather high signal to noise ratio as well as the interfering influence of physical activity (Huigen et al., 2002). As mentioned before, it is difficult to measure MWL using task performance or subjective ratings in real-world conditions. In this study, these measures are used to control the success of MWL manipulation. Besides, the use of physiological signals as a potential source of data for measuring MWL in conditionally automated driving is explored. In the present study, we concentrate on measurements of the autonomic nervous system for classifying drivers' workload. This is because we consider EEG or near-infrared spectroscopy are being less suitable under real-world conditions since drivers might be averse to wearing a headset. Besides, drivers' gaze can constantly switch between the windshield, the dashboard and potentially a tablet or a smartphone held in the hands during conditionally automated driving, which makes it challenging to continuously capture this feature. However, we are convinced that these measures could also represent interesting indicators and should be considered in future research.

2.2. Definition of Physiological Indicators

2.2.1. Electrodermal Activity (EDA)

The first selected physiological signal is the EDA, which is defined as the changes in the electrical conductivity of the skin, caused by the fluctuations of sweat in glands regulated by the autonomic nervous system (Cacioppo et al., 2007). The latter can be declined in two main components. One feature which can be derived from EDA data is the tonic level of EDA which refers to the slow-acting components of electrical activity such as the mean level of EDA or slow climbing and decreases over time. The most common measure of this component is the skin conductance level. Changes in this measure reflect general changes in arousal (Cacioppo et al., 2007). The second component is the phasic component of EDA, which refers to fast-changing properties of the signal. It is measured with the Skin Conductance Responses (SCRs). Previous research suggested that both components are important and may rely on different neural mechanisms (Cacioppo et al., 2007). Phasic SCRs can be distinguished into two categories called non-specific SCRs (NS-SCRs) and event-related SCRs (ER-SCRs). The first one gathers responses occurring in the absence of identifiable eliciting stimuli, while the second one characterizes subjects' electrodermal reaction to stimuli. One commonly used indicator is the frequency of NS-SCRs, which is generally between one and five per minute at rest, and more than 20 per minute in periods of high arousal. To characterize ER-SCRs, indicators such as latency, amplitude, rise time and half recovery time are usually used (Boucsein, 2012). The same indicators can be calculated identically for NS-SCRs, except for latency which requires a time-stamped triggered event to be calculated.

2.2.2. Electrocardiogram (ECG)

The second selected physiological signal is the ECG. Various indicators can be computed based on ECG data such as the heart period (the time interval between successive heart cycles) and the heart rate variability (HRV; Camm et al., 1996). The heart period is also known as the inter-beat interval (IBI). Another widely used metric to evaluate the cardiac activity is the heart rate (HR) which corresponds to the number of heartbeats per unit of time, usually per minute. HRV is a general term that refers to time changes in IBI. These measures are used as indices of autonomic nervous system regulatory activities and have been related to individual differences in attention and cognition in various groups of populations (Cacioppo et al., 2007). Previous studies showed that mental effort is related to changes in cardiovascular state (Aasman et al., 1987; Bernston et al., 1993) and more specifically in HRV (Mulder, 1992). The HRV can be quantified by two different categories of methods. The first method is the time-domain method. This category contains both statistical and geometric measures, depicting the variability of time between heartbeats (Camm et al., 1996). Malik and Terrace (1996) recommend indicators to use in this regard that are the standard deviation of IBI (SDNN, for estimating the overall HRV), the HRV triangular index (an estimate of the overall HRV), the standard deviation of IBI calculated over short periods (SDANN, an estimate of long-term changes of HRV) and the square root of the mean squared differences of successive IBI (RMSSD, an estimate of short-term components of HRV). While some specialists in the field advise calculating common statistical time-domain HRV measures such as SDNN or RMSSD using at least a 5-min ECG recording (Camm et al., 1996; Malik and Terrace, 1996), other investigations have utilized ultra-short term measures (below 5 min) (Shaffer and Ginsberg, 2017). Studies showed that 10 s for HR, 30 s for RMSSD and 60 s for other metrics such as pNN50 could be enough to get a reliable measure of cardiac activity (Salahuddin et al., 2007; Baek et al., 2015). Other metrics such as SDANN and the HRV triangular index require long-time monitoring (at least 20 min, preferably 24 h; Malik and Terrace, 1996). The second method to evaluate changes in HRV is the frequency-domain method. Power spectral density method provides information on how power (e.g., variance) distributes as a function of frequency. Three main components can be distinguished in periods of two to 5 min of recording, including the Very Low Frequency (VLF; below 0.04 Hz), the Low Frequency (LF; between 0.04 and 0.15 Hz) and the High Frequency (HF; between 0.15 and 0.4 Hz; Malik and Terrace, 1996). More recent researches suggest that 20–90 s could be enough to evaluate the components of HF and LF (Salahuddin et al., 2007; Baek et al., 2015). However, the use of VLF should be avoided to interpret recordings shorter than 5 min. The ratio LF/HF is also an indicator used to emphasize the behavior of the two main branches of the autonomic nervous system.

2.2.3. Respiration

The third physiological signal recorded in this study is the respiration of drivers. The respiratory system is complex and sensitive to other psychological variables (Cacioppo et al., 2007). Respiration forces the chest to expand and this movement of chest expansion can be measured by piezoelectric sensors. The respiratory system is linked with other muscles of the body as well as with the nervous system. Under ideal conditions, the respiratory activity is regular and harmonious but it can be perturbed when experiencing stressful situations. Previous research showed that respiration influences both EDA and heart activity (Cacioppo et al., 2007). Several measures can be extracted based on the information provided by breathing transducers such as the breathing rate (BR), which corresponds to the number of breathing cycles per minute. Inspiratory and expiratory volumes and durations, the ratio of both, and the complexity of the signal (through spectral analysis) are also measurements that can be derived from the raw breathing signal.

2.2.4. Respiratory Sinus Arrhythmia (RSA)

Heart rate changes as a function of the respiratory cycle. This phenomenon is called respiratory sinus arrhythmia. RSA has become of great interest in recent years since the tight coupling of both signals can be used as an index of the vagal control of the heart (Cacioppo et al., 2007). Many factors influence RSA such as posture, age, or activity. The main measure is the magnitude of RSA but both frequency and time domain methods can be used as well since they showed similar results.

2.3. Influence of MWL on Physiological Measures

Previous studies already investigated the influence of increased MWL induced by cognitive tasks on the physiological state of subjects. The goal is to summarize previous findings in order to get a better appreciation of the expected results in this study. Studies that manipulated MWL by administering a secondary NDRT to drivers were reviewed, as well as studies that manipulated MWL with a cognitive task that subjects had to perform on a computer under experimental laboratory conditions. Previous research showed that the EDA level increases with increasing task difficulty. It has been shown for subjects performing oral or auditory tasks while driving in the real field (Collet et al., 2009) or in a simulated environment (Mehler et al., 2009, 2012). The same effect was found for a visual task performed while driving in both real field and simulated environments (Engström et al., 2005) or on a computer (Ikehara and Crosby, 2005). However, no effect of task difficulty was found on EDA for an auditory task performed in a driving simulator (Engström et al., 2005). A possible explanation was provided by Mehler et al. (2012) in their follow-on study for this non-consistent effect of incremental difficulty of task on EDA in the study of Engström et al. (2005). Some participants might have disengaged from the task when performed at high levels of difficulty, resulting in a lower physiological activation. This shows the high importance of controlling the performance of the participants with regard to the secondary task. For measures describing the cardiac activity, HR and IBI were also shown to be sensitive to increased task difficulty. IBI decreases (e.g., HR increases) with increasing difficulty of the visual task performed in both simulated and real environment (Engström et al., 2005). The same effect was found for auditory tasks (with verbal prompt or not) performed under real driving conditions (Engström et al., 2005; Mehler et al., 2009). Collet et al. (2009) found similar results since HR of participants increased when performing various oral and auditory tasks while driving. An effect of increased task demand induced by the environment in simulated driving was also found on HR and frequency-based HRV measures (Brookhuis et al., 2004; Brookhuis and de Waard, 2010). In addition, the respiratory activity of subjects is also sensitive to the performance of the auditory prompt-verbal response "n-back" task while driving (Mehler et al., 2009). This study showed a plateau effect between the 1-back and 2-back conditions for BR and EDA, suggesting that it might be difficult to distinguish two different levels of high cognitive workload using physiological measures. Subsequently, Mehler et al. (2012) found that the seeming plateau effect in the earlier study was an artifact of the methodology employed and that when the order of task difficulty is randomized, significant differences in EDA level between the 1-back and 2-back were observed, confirming that the mean EDA level increases with task demand. In summary, previous studies led in various experimental settings already showed that changes in driver's workload can be measured using physiological signals. Some significant results were found in both real field and simulated environments (Engström et al., 2005), although the simulation was probably less realistic than it is now. Some indicators such as the mean EDA level or IBI can be used to measure changes in workload. Increasing task difficulty lead to increasing MWL, which goes with reduced IBI (e.g., higher HR), increased mean EDA level and increased BR. Some of these measures showed to vary when engaging in tasks involving different sensory channels such as auditory, oral or visual tasks. The fact that a speech-based cognitive task increases EDA, HR, and BR (Collet et al., 2009; Mehler et al., 2009) is particularly relevant for our study. We expect that these indicators will play a significant role in the classification of drivers' condition.

2.4. Classification of Workload

Our contribution is to classify drivers' workload using physiological measures during conditionally automated driving. In this regard, the classification results and procedure from previous studies, including the type of chosen physiological signals, the features generation and selection techniques, the selected classifiers, the validation techniques, or the number of classes to predict are reported. In this section, we reviewed studies for which the experimental task was the accomplishment of a NDRT during manual driving periods (real or simulated driving) or the performance of a single cognitive task in a laboratory. For each study, the method employed by authors to perform the classification is described.

Ferreira et al. (2014) asked two groups of adults (young vs. old) to perform two different cognitive tasks on a computer, testing their perceptual speed and visio-spatial cognitive processing capabilities. Each group performed three blocks of these two tasks, with two different difficulty levels of the tasks. One hundred and twenty-eight features were extracted from the EEG, ECG, EDA, heat flux, and respiration raw signals. Features were computed from 10 and 60-s segments using sliding windows with a step of one second. An inter-subject classification achieved results from 64 to 86% accuracy to distinguish two difficulty levels, depending on the task, the age group and the time window used for classification. The best scores were achieved mostly with data collected from young participants but with a high variation.

Haapalainen et al. (2010) administered six elementary tasks to 20 young subjects on a computer. Tasks were asking for visual perception and cognitive speed. A Naive Bayes classifier had to choose between two levels of cognitive load (low vs. high) using features derived from non-overlapping segments of psychophysiological measures during the tasks. Features such as statistical indicators of pupil diameter, GSR, heat flux, mean absolute deviation of ECG, EEG power values, two mental state outputs, heart rate and time-based HRV features were calculated. A leave-one-out approach was used for validation. Finally, the authors averaged the classification results across all participants, using the best feature from each sensor. An average of 76 and 71.4% of accuracy was achieved with respectively the heat flux and mean absolute deviation from ECG. An accuracy of 81.1% was achieved by combining both features. Besides, the classification with EDA as an input feature showed the lowest performance.

In another study led by Hogervorst et al. (2014), 14 participants had to perform the visual n-back task on a computer at different levels of difficulty (rest, 0, 1, and 2-back). Each participant did 8 epochs of 2 min of that task in the 3 difficulty levels. Features such as frequency-based indicators from EEG, mean EDA level, time-based HRV indicators, breathing frequency, and eye-related indicators were used. The best classification accuracy reached was a little over 90% for distinguishing high (2-back) and low (0-back) workload on the basis of 2 min segments with all indicators. The breathing frequency was the most useful physiological measure for classifying workload level. Using only physiological features, the best accuracy achieved was around 75% to distinguish 2-min segments of 0-back and 2-back task using the support vector machine classifier. This score decreased slightly under 70% when using 30-s segments for the classification.

Son et al. (2013) collected driving, physiological and eye movement data of 30 participants performing the auditory n-back task while driving. Task difficulty was varied on three levels (0, 1, and 2-back task) for a duration of 2 min each. HR and skin conductance level were used as physiological features. Ten-second windows across all 2-min windows were used to compute the features. A support vector machine classifier with a nested cross-validation technique was used to classify periods of normal driving and dual-task periods (NDRT and driving). The heart rate showed the best accuracy as a single feature to classify workload with 80% accuracy. In addition, all models with the two physiological inputs (HR and skin conductance level) obtained at least 82.6% accuracy.

A recent study led by Darzi et al. (2018) aimed at identifying the causes of hazardous driver states, using a combination of driver characteristics, vehicle kinematics, and physiological measures. 21 drivers were asked to perform four 45-min sessions of simulated driving. Each driving session contained eight scenarios with changing weather, traffic density and NDRT. The classification of cell phone usage periods only with physiological data is the most relevant result for our study because it is close to what we are trying to achieve in this experiment, except that it is in the context of conditionally automated driving. During cell phone use, participants indicated to have a higher MWL with regard to NASA-TLX (NASA Task Load Index, Hart and Staveland, 1988) results. Seventeen features were computed from four physiological signals (ECG, EDA, respiration, and temperature) from each 4-min scenario. It included time-based and frequency-based HRV measures, skin temperature, indicators of tonic and phasic EDA and respiration rate and variability. To automatically classify drivers' condition, the support vector machines, logistic regression and decision trees were selected as classifiers. For the physiological features, the baseline value was subtracted to driving value and then normalized using the minimum and maximum during a session. Only significant features to the stepwise forward feature selection (threshold of 0.05) were selected and the leave-one-out validation method was employed. Only with physiological features, classifiers were able to detect that participants used the cell phone while driving with a 81.8% accuracy. The most useful

TABLE 1 | Summary of state of the art.

References	Physiological features			Best classifier	Best performance (accuracy)	Best features
	ECG	EDA	RESP			
Ferreira et al. (2014)	Yes	Yes	Yes	QDA	86% (60 s)	EEG, respiration rate
Haapalainen et al. (2010)	Yes	Yes	No	NB	83.70%	Heat flux, HRV
Hogervorst et al. (2014)	Yes	Yes	Yes	SVM	75% (2-back vs. 0 back, 120 s)	EEG, respiration, RMSSD
Son et al. (2013)	Yes	Yes	No	SVM	82.9%	HR, skin conductance
Darzi et al. (2018)	Yes	Yes	Yes	LR	82.3% (cell phone or not)	ECG gradient, respiration rate
Solovey et al. (2014)	Yes	Yes	No	MLP	75.7%; HR only : 74%	
	Yes	Yes	No	LR	90% (30 s); HR : 80–85%	
Le et al. (2018)	No	No	No	DT	89.91%	

QDA, Quadratic Discriminant Analysis; NB, Naïve Bayes; SVM, Support Vector Machine; LR, Logistic Regression; MLP, Multilayer Perceptron; DT, Decision Tree.

physiological features for classification were the mean breathing rate and the absolute value of the gradient of the ECG signal.

Another main study in this domain had been led by Solovey et al. (2014). They conducted two field studies with 20 and 99 drivers to classify workload of drivers only with physiological data using respectively a subject-dependant and a subject-independent classification approach. Once again, the n-back task was used to manipulate drivers' workload (auditory prompt and verbal answer with digits). Various size of time windows (10–30 s) and overlapping factors (0–75%) were tested. Statistical measures of HR, skin conducatance level and vehicle velocity were used for classification. For the subject-dependant classification, accuracies around 75% were achieved with all classifiers (except for the k-Nearest Neighbor one) using all features. The model accuracy did not decrease much using only HR features for classification. For the subject-independent classification, accuracies around 90% were achieved only with physiological measures. The additional driving features did not increase accuracy, suggesting that physiological measures alone have a great potential for classifying drivers' workload in automated driving (where driving features are not available). Logistic Regression, Multilayer Perceptron and Naïve Bayes classifiers were the most efficient ones. Increasing the time widows for computing features increased the accuracy, with the best accuracy achieved with a 30-s time window. However, the overlapping factor did not affect the accuracy of the system.

Finally, Le et al. (2018) recently considered using near-infrared spectroscopy to similarly classify drivers workload being cognitively distracted by a NDRT. Again, the n-back task was chosen for manipulating workload and 6 features were computed from the sensor data. Five-fold cross-validation and a principal-component analysis were applied to data before the final classification. Five different classifiers were tested to classify three workload levels (driving only vs. driving + 1-back vs. driving + 2-back), including decision-tree, discriminant analysis model, logistic regression, support vector machine and nearest neighbor classifiers. Scores above 88% accuracy were achieved for subject-dependant classification and between 84 and 90% for subject-independent classification. The results obtained in this study are very promising the results are obtained for classifying three levels of workload compared to other studies that classified only two levels. However, the sensor was placed on the forehead of drivers, who may not be willing to wear such a device under real driving conditions.

To summarize the findings from previous studies, the main results of each article are presented in **Table 1**. Overall, decreasing the time window for computing physiological measures showed to decrease accuracy. Apart from psychological features such as EEG, some of the best physiological features to classify MWL were the breathing rate, HR or the mean absolute deviation of IBIs. Another main result to consider is that the models developed in several studies always benefited from sensor fusion. This leads to a compromise to classify the driver's condition. Previous research reviewed here raises a fundamental issue if we want to implement such systems in vehicles. The stake is to find the best trade-off between the number of physiological signals, features and time window to build a reliable and robust model (e.g., high accuracy with low variance). If too many signals are selected, it is difficult for the driver to wear many sensors under real driving conditions. Also, if a time window of a few minutes is needed to get an acceptable accuracy, this takes us away from real-time MWL assessment and therefore makes an implementation of such a system less credible.

3. CURRENT STUDY

In this paper, we propose a solution that classifies the driver's MWL (high vs. low) based on physiological data during conditionally automated driving. In particular, the following contributions are made:

- Creation of a dataset containing three physiological signals (ECG, EDA, and respiration) of 90 subjects in the specific context of conditionally automated driving in a simulator.
- Manipulation of drivers' MWL through a verbal cognitive task with a rigorous experimental approach. The selected task is similar to a task that drivers might engage in under real driving conditions (e.g., talking on the phone or to another passenger).
- Validation of the success of workload manipulation by means of the widely used questionnaire NASA Task Load Index (NASA-TLX, Hart and Staveland, 1988).
- Training of three different classifiers to predict drivers' condition, using a k-fold cross-validation approach.

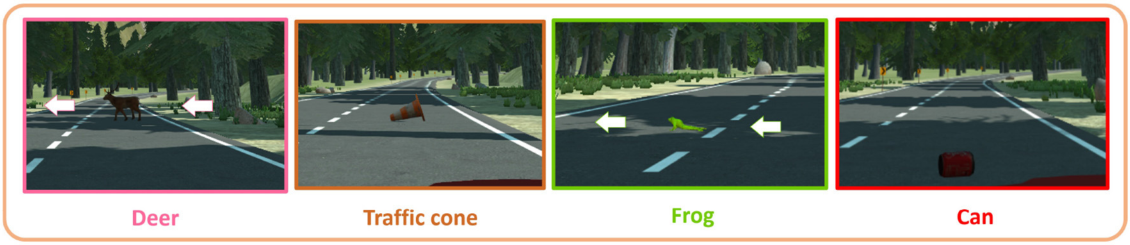


FIGURE 1 | Takeover situations.

FIGURE 2 | The driving simulator.

- Evaluation of the effect of selected physiological signals and segmentation level (e.g., size of time windows used to compute features) on performance of classifiers.

Our approach differs from previous studies because it investigates the change of MWL of drivers in the specific context of conditionally automated driving. With the future rise of automated driving, it is important to validate that the results of previous findings are consistent with the increase of drivers' MWL at higher levels of automated driving. Besides, the measures used to classify drivers' MWL differ from some previous studies that used eye-tracking, EEG or driving features. In this study, only physiological signals that can be collected in real-world conditions using smart embedded sensors are used. Therefore, findings from this study are relevant to the potential use of physiological signals for detecting changes in MWL of drivers in future conditionally automated cars.

4. MATERIALS AND METHODS

4.1. Experimental Method

4.1.1. Participants and Experimental Design

90 young participants (24.15 ± 5.95 years old) within a tight age range were recruited for this study. 40 of them identified themselves as male, 49 as female and 1 as other. Participants were mainly students. All participants were required to hold a driving license and be of good general health. Students received course credit for their participation. All the research and measurements followed the tenets of the Helsinki agreement and written informed consent was obtained from all participants.

The experimental design was a 2 × 6 mixed-design with the task difficulty as a between-subject variable (secondary task vs. no secondary task) and the takeover situation as a within-subject variable (deer vs. traffic cone vs. frog vs. can vs. false alarm 1 and 2). The cognitive NDRT that half of the participants had to perform was a verbal cognitive task named oral backward counting (Siegenthaler et al., 2014; Krueger et al., 2019). It consisted of counting backwards for 20 min from 3,645 by step of 2. This artificial task was chosen because it is a continuous task similar to a discussion on the phone or between passengers in the car. With such task, a higher level of MWL was continuously induced over a long period of time. This gave the possibility to investigate the effect of segmentation on physiological signals. Also, the engagement in such difficult task could be measured. The six takeover situations included four obstacles that led to taking over control: a deer and a frog crossing the road, as well as a traffic cone and a can standing on the track (**Figure 1**). Participants also received two false alarms. They could choose to take over control if they estimated that the situation was dangerous for them and the car. The takeover situations were implemented in the scenario to make it more realistic and engaging for participants. However, the effect of the takeover situation on the physiological state of subjects is not presented in this work.

4.1.2. Material and Instruments

The experiment was conducted on a fixed-base simulator, as shown in **Figure 2**. It is composed of two adjacent car seats with seat belts and a Logitech G27 steering wheel with the gas, brake and clutch pedals. The clutch was not used in this study since the car used in the simulation had an automatic gearbox. The orientation and the longitudinal position of the seats toward the steering wheel were adaptable like in a real car. All this structure was installed in front of a large screen where the driving simulation software was back-projected with a projector (model Epsilon EH-TW3200). Two speakers were set up behind the seats to immerse the driver in the simulated driving environment. A cabin-like room with low ambient lighting contained all of this installation. The driving simulation used GENIVI software, developed with Unity by a consortium of car manufacturers. The scenario used for the experiment was a replication of the Yosemite National Park (USA) and included conditionally automated driving features. It was modified to add the obstacles triggered by the experimenter, leading to six takeover requests throughout the driving session. During the experiment, the Biopac MP36 hardware recorded the physiological signals of drivers, at a sample rate of 1,000 Hz. Appropriate parameters were used for each channel. A digital low pass filter with a frequency of 66.5 Hz and a Q factor of 0.5 reduced the noise of the three signals. For the EDA and RESP signals, the filter had a respective gain of 2,000 and 1,000. The SS57LA and SS2LB lead sets (Biopac) with disposable Ag/AgCl pre-gelled electrodes (EL507 and EL503, Biopac) respectively collected the EDA and ECG of participants. The SS5LB respiratory effort transducer (Biopac) recorded the respiration via chest expansion and contraction. The voice of participants assigned to the manipulation group was recorded by an audio recorder placed behind the dashboard.

FIGURE 3 | Upper part: The dashboard displaying the icon indicating the autopilot mode, the speed of the car and the number of engine's revolutions per minute. Bottom-left: gray icon—Autopilot OFF, Bottom-middle: green icon—Autopilot ON, Bottom-right: Red icon—Takeover Request (TOR).

4.1.3. Measures

Physiological signals of participants were recorded throughout the whole experiment, including ECG, EDA, and RESP. Based on these data, physiological indicators were calculated. The creation of features from these indicators is presented later in the article (section 4.2.1). The subjective workload was assessed using the widely used questionnaire NASA-TLX (Hart and Staveland, 1988). It is a 6-item questionnaire where participants report their subjective level of workload during a task. After their experience in the driving simulator, participants were asked to rate their workload during the main driving session. The scale was modified due to visualization problems on the questionnaire. Hence, each item rated on a 11-point scale, from 0 to 10 (0 = Low, 10 = High). The mean score of the six items was computed to create a global score of MWL rating from participants. To ensure that participants were engaged enough in the NDRT throughout the driving session, we also measured NDRT performance of the participants. The frequency of orally spoken number (i.e., the number of orally spoken numbers per minute) was calculated from recordings obtained with the voice recorder. From times to times, participants stopped counting since the task was monotonous. For that reason, we also counted the number of times the experimenter asked the participant to resume counting.

4.1.4. Procedure

After initial instructions about the experiment, participants answered a questionnaire containing socio-demographic questions (i.e., age, gender, driving experience, accidents, etc.). To record the physiological signals, the experimenter attached electrodes and the respiration belt to the participants' body. Three electrodes were attached to record the ECG, two above both ankles and one at the right wrist. For the EDA, two electrodes were attached to the index and middle finger of the right hand of participants. Then, the experimenter asked them to take a seat in the simulator. The experiment took place in three distinct periods. Oral instructions were given by the experimenter before each period to ensure that participants understood the experimental procedure. As described in the section 4, the three periods took place in the same scenic environment. During the first period, participants had to monitor the environment of the car while it was driving in conditional automation for 5 min. They were told that no takeover could be requested during this period. Indicators computed during this phase corresponded to the physiological baseline of participants.

The second period served as a practice session for the participants. During 5 min, they could familiarize with the takeover process as well as with the driving functions of the simulator (e.g., sensitivity of the steering wheel, gas and brake pedal etc.). Before starting, the experimenter reminded the subjects that they were driving a level 3 vehicle. The meaning of icons showing the state of the autopilot on the dashboard was explained to the driver (cf. **Figure 3**). For each TOR, the simulation displayed a red icon on the dashboard and played an audio chime in the speakers. The experimenter also explained how the participants could take over control of the car, either by steering the wheel, braking or pressing the upper-right button placed on steering wheel. For the practice session, drivers were told that three false alarms would be triggered to become familiar with the process. After the three false alarms were triggered, the experimenter made sure that participants understood the process. Then, they had the chance to drive manually until the end of the 5 min. This study does not include the analysis of data during the practice drive.

The third period consisted of the main driving session that lasted 20 min. The experimenter reminded the participants to take over control of the car only in a situation that they considered dangerous for themselves and the vehicle. They had to react accordingly to six TORs. Each one was randomly triggered between 1 min and a half and 4 min after the previous TOR. The randomization of time between takeover was implemented to avoid an expectation effect. Once participants gained control over the critical situation and considered it as safe again, they were instructed to reengage the autopilot. To do that, they had to position the car in the center of the right lane and press a button on the steering wheel. In addition, half of the participants had to perform the speech-based cognitive secondary task while the car was driving. At the end of the session, participants were asked to stop the car and leave the simulator. The experimenter removed electrodes and the participants could fill in the last part of the questionnaire. Then, they were thanked and discharged.

4.1.5. Pilot Study

Eight people took part in the pilot study. The purpose was to check that all the data were correctly recorded (physiological signals and driving data) and that the driving scenario was running flawlessly. Shadows on the lane due to the reflection of sunlight on trees were removed from the driving environment. Indeed, obstacles were not triggered at the same location for all participants to minimize the impact of the visibility of drivers during the takeover situations. Also, the original design contained a third experimental condition. This condition was to count backwards by step of 13 to induce a higher cognitive load. However, we realized that it was too demanding to perform this task for 20 min.

4.1.6. Statistical Analysis

The analyses were performed using IBM SPSS Statistics 25. By examining the audio files recorded during the execution of the NDRT, we found that the calculation task was performed correctly and accurately. Four participants made some errors during the NDRT. Three of them made a mistake in the transition from 3,001 to 2,999, starting again from 3,999. The fourth participant obtained an even score at the end. However, they were not removed from the analysis because they kept counting, which was the most important for the inducement of MWL. For the subjective ratings of the NASA-TLX, nine participants were removed due to due to issues with the online questionnaire. To test for the difference of MWL between the control group and the treatment group, analyses of variances (ANOVAs) were calculated for each questionnaire item and the global score of MWL. Cohen's effect size is reported when the ANOVA showed a significant result.

4.2. Classification Method

This section describes the methodology used to classify drivers' condition (secondary task vs. no secondary task) based on the recorded physiological signals. A first goal was to investigate the effect of sensor fusion on classification accuracy. The classification was performed for each signal independently (ECG, EDA, RESP), each possible pair of signals and all signals combined. A second goal was to observe the effect of segmentation level. In other words, the main driving session was segmented into windows of different size that were used to compute features. Six segmentation levels were tested : 1, 2, 5, 10, 20, and 40. With a segmentation level of 1, the features were computed from one 20-min window, whereas a segmentation level of 40 consisted of 40 30-s windows for computing features. The higher the segmentation level was, the more training examples the algorithm had for training. This process aimed at investigating the shortest time required to record physiological parameters to classify accurately the MWL of drivers. Overall, this work will help to find the best trade-off between the number and type of physiological signals needed, the optimal time-span for recording physiological data and the performance of a model to classify the level of MWL workload, with the ultimate goal of implementing such model in future automated vehicles under real-world conditions.

4.2.1. Data Preprocessing

The preprocessing of raw physiological data was automated using the Neurokit library in Python (Makowski et al., 2021). Neurokit is a module that provides high-level integrative functions to process and exploit bio-signals. Signals from the baseline and driving phases were processed separately. The result of the processing step resulted in the computation of physiological indicators. To summarize the indicators computed in this study, a definition of each indicator calculated from physiological raw signals is proposed in **Table 2**.

The EDA signal was processed using methods of convex optimization (Greco et al., 2016), which defines EDA as the sum of three terms: the phasic component, the tonic component, and an additive white Gaussian noise term incorporating model prediction errors as well as measurement errors and artifacts. To be able to process the EDA signal with the convex optimization method, it had been down-sampled to 50 Hz to reduce computation time. The signal had also been filtered with a Finite Impulse Response low-pass filter of fourth order with a cut-off frequency of 5 Hz and smoothed using the convolution of a filter kernel with the input signal (Smith, 1999). That smoothing process used the moving average principle, with a window size of three-quarters the sampling rate. The output was the EDA raw signal, the filtered signal, the tonic component, the phasic component, the SCR onsets, peak indexes and amplitudes. Based on the related work, we chose to use the filtered signal, the tonic component and indicators that characterize NS-SCRs because we evaluate changes in drivers' state over a long period. Hence, EDA indicators including the minimum, maximum, standard deviation and mean values of filtered and tonic EDA signals were computed in this study, in addition to the frequency and the mean amplitude of NS-SCRs.

The ECG signal was filtered with a Finite Impulse Response band-pass filter of fourth-order with cut-off frequencies of 3 and 45 Hz. A QRS-detector algorithm was used to locate R-peaks from the ECG signal (Hamilton, 2002). The output was the ECG raw signal, the filtered signal and the R-peaks indexes. From that, HR and HRV indicators were computed. HRV indicators included time domain, frequency domain and non-linear domain indicators.

The respiration signal was filtered with a Butterworth band-pass filter of second-order with cut-off frequencies of 0.1 and 0.35 Hz and smoothed with the same process than EDA and a rectangular window size (also known as Dirichlet window) of 3 s (Smith, 1999). The output was the respiration raw signal, the filtered signal, the respiratory cycles onsets, and respiratory phases (inspirations and expirations). From that, indicators of rate and variability of respiration were computed.

Also, from both respiration and ECG signal, RSA features were computed using the peak-to-trough (P2T) and Porges–Bohrer methods. The P2T algorithm computes all RSA estimates in a given period. For each breath, an estimate of RSA is

TABLE 2 | Summary of physiological indicators computed from raw physiological signals.

Signal	Indicator	Domain	Description
EDA	Mean raw level		The mean value of filtered EDA signal
	Min raw value		The minimum value of filtered EDA signal
	Max raw value		The maximum value of filtered EDA signal
	Std raw value		The standard deviation of filtered EDA signal
	Mean tonic level		The mean value of tonic EDA signal
	Max tonic value		The minimum value of tonic EDA signal
	Min tonic value		The maximum value of tonic EDA signal
	Std tonic value		The standard deviation of tonic EDA signal
	Amp. NS-SCRs		The mean amplitude of NS-SCRs (computed from phasic EDA signal)
	Freq. NS-SCRs		The number of NS-SCRs per minute (computed from phasic EDA signal)
ECG/RESP	Mean Rate	Time	The mean number of cardiac cycles per minute
	Mean		The mean time of IBIs/BBs
	Median		The median of the absolute values of the successive differences between adjacent IBIs/BBs
	MAD		The mean absolute deviation of IBIs/BBs
	SD		The standard deviation of IBIs/BBs
	SDSD		The standard deviation of the successive differences between adjacent IBIs/BBs
	CV		The Coefficient of Variation, i.e., the ratio of SD divided by Mean
	mCV		Median-based Coefficient of Variation, i.e., the ratio of MAD divided by Median
	RMSSD		The square root of the mean of the sum of successive differences between adjacent IBIs/BBs
	CVSD		The coefficient of variation of successive differences; the RMSSD divided by Mean
	LF	Frequency	The spectral power density pertaining to low frequency band (0.04 to 0.15 Hz)
	HF		The spectral power density pertaining to high frequency band (0.15 to 0.4 Hz)
	LF/HF		The ratio of low frequency power to high frequency power
	SD1	Non-linear	Measure of the spread of IBIs/BBs on the Poincaré plot perpendicular to the line of identity
	SD2		Measure of the spread of RR intervals on the Poincaré plot along the line of identity
	SD2/SD1		Ratio between long and short term fluctuations of IBIs (SD2 divided by SD1)
ECG	pNN50	Time	The proportion of successive IBIs greater than 50 ms, out of the total number of IBIs
	pNN20		The proportion of successive IBIs greater than 20 ms, out of the total number of IBIs
	TINN		The baseline width of IBIs distribution obtained by triangular interpolation
	HTI		The HRV triangular index (total number of IBIs divided by the height of IBIs histogram)
	VHF	Frequency	Variability, or signal power, in very high frequency (0.4–0.5 Hz)
	LFn		The normalized low frequency, obtained by dividing the low frequency power by the total power
	HFn		The normalized high frequency, obtained by dividing the low frequency power by the total power
	LnHF		The log transformed HF
	CSI	Non-linear	The Cardiac Sympathetic Index (longitudinal variability of Poincaré plot divided by transverse variability)
	CVI		The Cardiac Vagal Index (logarithm of the product of longitudinal and transverse variability)
	CSI_modified		The modified CSI (the square of the longitudinal variability divided by transverse variability)
RESP	Mean amplitude	Time	The mean respiratory amplitude
	ApEn	Non-linear	The approximate entropy of RRV
	DFA2		A long-term fluctuation value. Only be computed if mora than 640 breath cycles in the signal
RSA	Mean		Mean of RSA estimates
	Mean Log		The logarithm of the mean of RSA estimates
	SD		The standard deviation of all RSA estimates
	NoRSA		The number of breath cycles from which RSA could not be calculated
	RSA_PB		The Porges–Bohrer estimate of RSA, optimal when the signal to noise ratio is low, in ln(ms^2)

Identical indicators computed from both ECG and respiration (RESP) signal are grouped together. IBIs refers to interbeat intervals (ECG) and BBs refers to breath-to-breath cycles (RESP).

calculated by subtracting the shortest heart period during inspiration from the longest heart period during a breath cycle (Lewis et al., 2012). RSA features included the mean, the standard deviation and the logarithm of the P2T estimates (in milliseconds), in addition to a measure computed with the Porges–Bohrer method, as explained in Lewis et al. (2012).

4.2.2. Feature Generation and Normalization

The same feature engineering process was applied to data of all participants, regardless of their experimental condition. Each indicator presented above was calculated for each segment of the driving phase. It was taken as a feature for the classification. For each indicator, an additional feature was computed, corresponding to the difference of that indicator between the driving segment and the baseline. That feature engineering process aimed at taking into account the physiological state of participants at rest and evaluate the individual changes on each indicator during the driving session (Darzi et al., 2018). In this way, the model performance should be higher for a between-participant validation procedure, since the ultimate goal is to build a model that would perform well with any subject inside the car. In total, the three raw physiological signals served to compute 122 features corresponding to 61 physiological indicators (10 from EDA, 27 from ECG, 19 from RESP, five from RSA). For classifiers sensitive to the range of features, data were normalized using the maximum and minimum of each feature during a driving segment. The minimum value was subtracted to the feature value and then divided by the difference between the maximum and minimum values.

4.2.3. Feature Selection

Statistical analysis techniques are usually employed to check for the effect of a between-subjects factor on dependant variables. Therefore, we chose to do an ANOVA on each one of the 122 features independently. Only the physiological features that reached the significance level (p-value lower than 0.05) were used for classification. The number of features was not the same depending on the segmentation level and the physiological signals used for the classification task.

4.2.4. Selected Algorithms

At this step of the procedure, the dataset consisted of some selected features that were used as input of classifiers for the training and validation procedure. Three algorithms were selected based on results from previous research in the field (Son et al., 2013; Solovey et al., 2014; Darzi et al., 2018) and for their ease of implementation. They have been implemented in Python using the scikit learn machine-learning framework (Pedregosa et al., 2011). The effect of selected physiological signals and segmentation level was tested with each classifier. Their classification principle is detailed below:

Random Forest Classifier (RF): A random forest is a meta estimator that fits some decision tree classifiers on various sub-samples of the dataset and use averaging to improve the predictive accuracy and control over-fitting (Breiman, 2001).

C-Support Vector Classifier (SVC): The support vector classifier uses boundaries (linear or more complex) to separate data in the input feature space. The separation boundary is defined by a kernel. In this experiment, we tested four different kernels: the linear, the sigmoid and the polynomial ones, as well as the radial basis function (Hsu et al., 2010).

Multi-Layer Perceptron Classifier (MLP): A multi-layer perceptron consists of a set of nodes distributed in a number of layers. It contains at least three layers of nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a non-linear activation function. The multi-layer perceptron utilizes backpropagation as a supervised learning technique for training (Hastie et al., 2009). Here, we use the multi-layer perceptron as a classifier, meaning that the output layer contains only two nodes that output the probability that the driver was performing a secondary task or not, based on input features. The classifier contained one hidden layer and we only tested to change the number of neurons in that hidden layer.

4.2.5. Optimization and Validation

To maximize the performance of classifiers, an optimization of hyperparameters of the three selected classifiers was done. The hyperparameter search aims to find the set of hyperparameters that minimizes the loss and maximizes the classification accuracy (Claesen and De Moor, 2015). The grid search technique was chosen to search for the best set of hyperparameters. It consists of predefining a range of values to test for each hyperparameter. The classifier tests all possible combinations of parameters for training and validation procedures. A first iteration of that grid search technique (GridSearch_1) was performed with a wide range of values. The goal was to eliminate values of hyperparameters for which the model does not perform well and hence reduce this range for the final optimization during the validation procedure. It was done on the entire dataset which was split into a training set (75% of samples) and a validation set (25% of samples). The hyperparameters that have been tested during this first optimization procedure can be found in **Table 3**. The definition and the chosen range of values for each parameter are presented. This first procedure was done for each level of segmentation and the feature selection process was also applied. The second hyperparameter optimization process was done during the final validation procedure. It used a reduced range of values defined after the first optimization. The k-fold cross-validation method was select to validate the performance of classifiers and prevent classifiers from overfitting the data (Hastie et al., 2009). In this procedure, the dataset was split into 10-folds. Classifiers were trained using data from 9 subsets and then validated on the remaining subset. The validation was repeated 10 times, with each subset acting as the validation subset once. The second step of optimization with a refined range of parameters (GridSearch_2) was performed within the final validation pipeline. The 10-fold validation procedure was performed once for each set of hyperparameters. Graphs and tables report results for the set of hyperparameters that gave the best mean accuracy for the classification overall 10 subsets.

TABLE 3 | Tweaked hyperparameters during the first iteration of the grid search procedure (GridSearch_1), with chosen ranges and step values for each parameter.

Classifier	Parameter name	Parameter definition	Range
RF	n_estimators	Number of trees in the forest.	[10, 507, 1,005, 2,000]
	max_features	Number of features to consider when looking for the best split.	Sqrt
	max_depth	Maximum depth of the tree. If None, then nodes are expanded until all leaves are pure or until all leaves contain less than 2 samples.	[None, 10, 57, 105, 152, 200]
SVC	kernel	Specifies the kernel type to be used in the algorithm	[linear, RBF]
	C	Regularization parameter.	[2e-3, 2e-1, 2e1, 2e7, 2e9, 2e11]
	gamma	Kernel coefficient for RBF kernel.	[2e-13, 2e-9] by step of 10
MLP	solver	Solver used for weight optimization.	[lbfgs, adam]
	max_iterations	Maximum number of iterations. Solver iterates until convergence or number of iterations.	[500, 1,500]
	alpha	L2 penalty (regularization term) parameter.	[1e-4, 1] by step of 10
	hidden_layer_sizes	The number of neurons in the hidden layer.	[32, 64, 128, 256, 512]
	random state	Determines random number generation for weights and bias initialization.	[0, 42]

RBF refers to Radial Basis Function.

5. RESULTS

5.1. Statistical Validation of MWL Inducement

5.1.1. Engagement on Task

The indicator used to check for the engagement on task was the frequency of orally spoken numbers. The participants counted backward, on average, to the number 2,740 ($M = 2740.03$, $SD = 311.28$), making an average of 452 ($M = 452.49$, $SD = 155.64$) calculations throughout the driving session. It is equivalent to 22.6 numbers orally spoken per minute, e.g., approximately one number every 3 s. During the experiment, the experimenters asked participants to resume counting on average twice ($M = 2.00$, $SD = 1.77$).

5.1.2. Subjective Ratings (NASA-TLX)

To control the success of the MWL manipulation, subjective ratings of workload collected from the NASA-TLX questionnaire were used. Results indicate higher level of reported MWL for the group that performed the secondary task ($M = 4.64$, $SD = 0.90$) compared to the control group ($M = 3.90$, $SD = 1.42$; $F_{(1, 79)} = 7.77$, $p < 0.05$, $d = 0.63$), regarding the global score of the NASA-TLX. The difference was also significant between both groups for mental demand ($F_{(1, 79)} = 59.85$, $p < 0.001$, $d = 1.73$), performance ($F_{(1, 79)} = 9.07$, $p < 0.05$, $d = 0.67$) and frustration ($F_{(1, 79)} = 6.83$, $p < 0.05$, $d = 0.58$). Means and standard deviations for all components of the questionnaire are shown in **Figure 4**.

5.2. Classification of Drivers' Workload

5.2.1. Reduction of Hyperparameter Range

The first iteration of the grid search (GridSearch_1) gave insights about the influence of hyperparameter values on the performance of the model. The RF classifier obtained the poorest results with 2,000 estimators and a maximum depth of 200. For the final pipeline, we reduced the range for these two parameters. The SVC classifier performed best across all segmentation levels with the linear kernel and C-values of 2e-1, 2e1, and 2e7. Therefore, we chose to only use the linear kernel and refine the final range of C-values. For some segmentation levels, the MLP classifier did not converge to achieve the best score after 1,500 iterations. Therefore, it was set to 2,000 for the final pipeline. The lbfgs solver (which stands for Limited-memory Broyden–Fletcher–Goldfarb–Shanno) was selected for the final optimization process because it gave better results more often than the adam solver. The smallest alpha value (1e-4) did not show satisfying results so it was excluded from the final range of values. The number of neurons in the hidden layer did not have much influence (except for 512 neurons). Therefore, a similar range of values was chosen. Finally, the random state was set at 42 for challenging the model with random initialization of weights and biases during the final procedure. The chosen range and step values for each hyperparameter tested during the final optimization procedure (GridSearch_2) are summarized in **Table 4**.

5.2.2. Influence of the Number of Selected Physiological Signals

Figure 5 shows the means and standard deviations of the classification accuracy using the 10-fold validation procedure. Results are reported for each classifier depending on the type and the number of chosen physiological signals, with a segmentation level of 1. For each combination of selected signals, **Table 5** shows the best mean accuracy (and standard deviation) and the classifier which performed best to classify drivers' condition over the 10-folds. Using only EDA as the input signal, the model showed the lowest performance, achieving between 69 and 73% accuracy regardless of the classifier. The model with ECG alone achieved 82–89% accuracy. With only one physiological signal as the input of classifiers, the respiration achieved the best results with an accuracy close to 90% on average over the 10-folds.

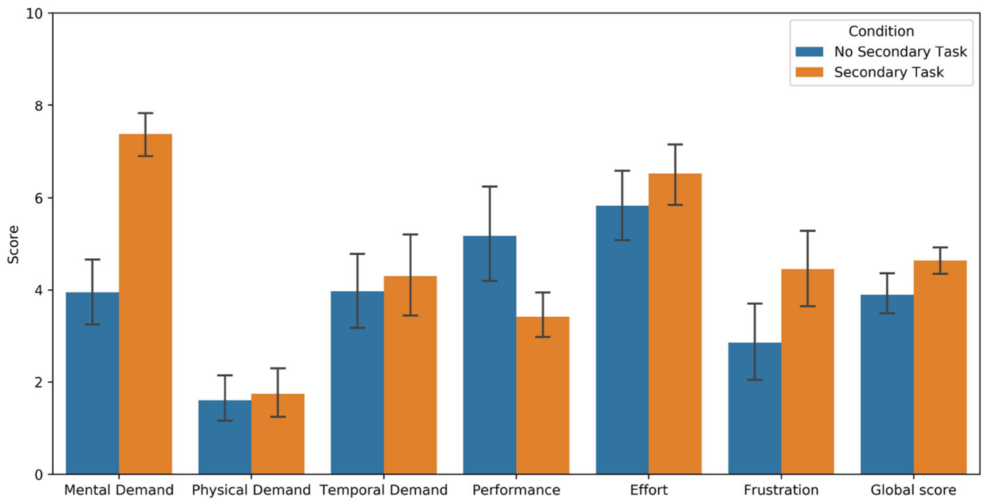


FIGURE 4 | Subjective ratings of MWL collected from the NASA-TLX questionnaire. Low value means high performance for the *Performance* item.

TABLE 4 | Final range of values tested for each hyperparameter (GridSearch$_2$).

Classifier	Parameter name	Range
RF	n_estimators	[10, 257, 505, 752, 1,000]
	max_features	sqrt
	max_depth	[None, 10, 40, 70, 100]
SVC	kernel	linear
	C	[2e-3, 2e7] by step of 10
MLP	solver	lbfgs
	max_iterations	2,000
	alpha	[1e-3, 1] by step of 10
	hidden_layer_sizes	[32, 64, 128, 256]
	random state	42

With two signals as the input of classifiers, the combination of EDA and ECG features showed the lowest accuracy, between 82 and 89% accuracy. The combination of EDA and respiration as input signals gave 87–89% accuracy. The best combination of two signals was respiration and ECG, achieving 92–94% accuracy depending on the selected classifier. Finally, the combination of the three signals resulted in accuracy between 91 and 92% for classifying drivers' condition.

5.2.3. Influence of the Segmentation Level

For each classifier and each segmentation level from 1 to 40, **Figure 6** shows the means and standard deviations achieved by the model after the classification task. The results are reported only for selected signals that achieved accuracy over 85% with a segmentation level of 1 with at least two classifiers. It includes the respiration alone, both pairs of EDA with respiration and ECG with respiration, and the fusion of the three signals. Best results for each level of segmentation are summarized in **Table 6**. The classifier and the combination of signals that gave the best results are also reported.

6. DISCUSSION

6.1. Manipulation of Workload

Regarding the results from the experimental manipulation, measures of task performance showed that participants were sufficiently involved in the NDRT they were asked to perform. Indeed, they counted orally with a rate of one number every 3 s on average. Subjective ratings of MWL showed that participants in the NDRT condition reported a significantly higher level of MWL than participants in the control group. Mental demand was the component that showed the largest effect size. Results from task performance and subjective ratings indicate that the manipulation of MWL of participants was successful. We can hence consider that performing such speech-based NDRT for 20 min in conditionally automated driving is increasing the MWL of drivers. From that, the effect of a higher level of MWL on the collected physiological data of drivers can be analyzed. A procedure using machine learning techniques for classifying drivers' MWL was used and an interpretation of results is proposed below.

6.2. Interpretation of Results Depending on the Selected Signals

The results are first interpreted for the effect of selected physiological signals on classification performance. Features were computed with a segmentation level of 1, meaning that physiological indicators were calculated from the entire driving period (20 min). With only one physiological signal selected as an input of the model, results showed that the model was performing poorest when the only EDA signal was selected. Using ECG alone, the model performed best, achieving an accuracy of 89% with the RF classifier. However,

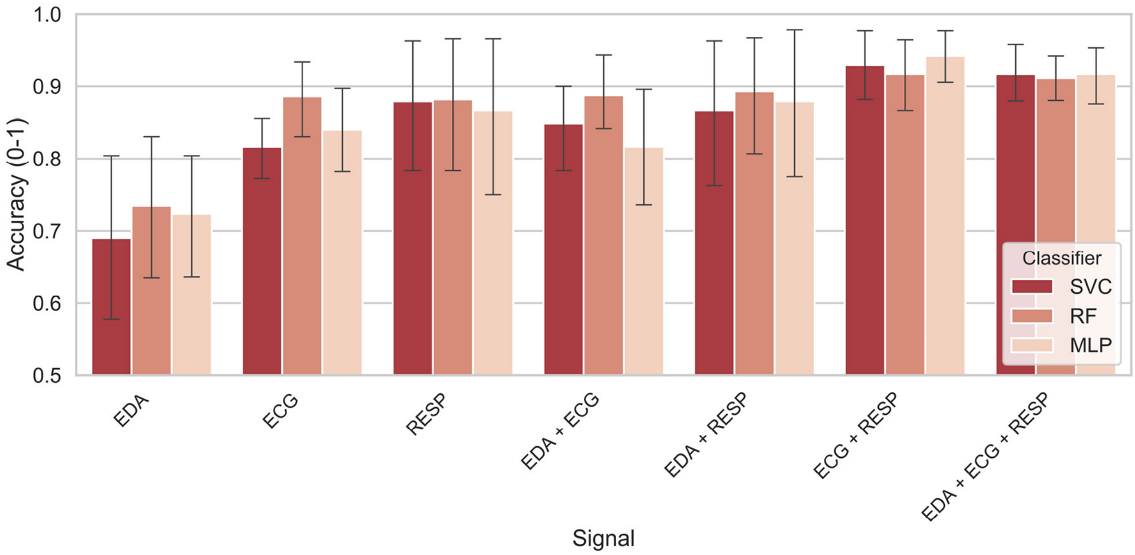


FIGURE 5 | Classification accuracy as a function of selected physiological signals and classifier.

TABLE 5 | Best score for each combination of selected signals (with a segmentation level of 1).

Selected signal	Best classifier	Best accuracy [Mean (SD)]
EDA	RF	0.73 (0.15)
ECG	RF	0.89 (0.09)
RESP	RF	0.88 (0.15)
EDA + ECG	RF	0.89 (0.09)
EDA + RESP	RF	0.89 (0.13)
ECG + RESP	**MLP**	**0.94 (0.06)**
EDA + ECG + RESP	SVC/MLP	0.92 (0.09)

Bold values show the best score achieved by the model.

from the three physiological signals alone, the model showed more consistent results across classifiers with the respiration signal selected alone as input. Indeed, each of the three classifiers achieved 87–88% of accuracy, but with higher variance compared to ECG. We can consider that both features computed independently from the respiration and ECG signals are useful to distinguish the driver conditions (verbal secondary task vs. no secondary task).

If we now look at the effect of sensor fusion on classification results, the fusion of EDA and ECG did not give better results than the ones achieved with ECG alone. In the same way, the fusion of EDA and respiration signals was not better than respiration alone. In previous studies, EDA indicators such as mean skin conductance level were shown to be sensitive to an increase of MWL (Engström et al., 2005; Mehler et al., 2009). Similar indicators were computed in this work such as mean tonic and raw level of EDA. Additional indicators relating the long-term changes of driver's state such as the frequency of NS-SCRs were supposed to improve the accuracy of the system. Results suggest that EDA features were the least useful ones to classify drivers' condition, as found by Haapalainen et al. (2010) and Son et al. (2013). Nevertheless, the model can achieve 73% with EDA features, which confirms that drivers' skin conductance is affected by the performance of a secondary task involving a verbal function (Engström et al., 2005; Collet et al., 2009; Mehler et al., 2009).

However, the model benefited from the fusion of two sensors without EDA. Indeed, the fusion of respiration and ECG signals showed to increase the accuracy of the system compared to the respiration or ECG alone, achieving accuracy levels of over 90% using all classifiers. This is consistent with statements made above, confirming that the ECG and respiration features are useful for classification. Also, the variance of scores obtained over the 10-folds was lower. This suggests that the performance of classifiers varied less from one-fold to the other during the classification task, making the model more robust. Besides, the fusion of the three signals as inputs of the model performed similarly (or slightly worse) than the one of respiration and ECG. The variance and accuracy achieved were also similar regardless of the classifier. Overall, the fusion of ECG and respiration showed to achieve the best performance to specify the drivers' condition, with an accuracy of 94% and a standard deviation of 0.06 across the 10-folds with the MLP classifier (**Table 5**). It is probably due to the additional respiratory sinus arrhythmia features that were computed during the processing of ECG and respiration signals. These features were taken into account in the classification procedure and might have helped the model to capture more information about the change of phase between ECG and respiration signals during the execution of the task.

If we compare the results to reviewed studies reported in **Table 1**, the accuracy achieved in this study is better, using only physiological features for the classification. Still, results must be compared carefully since the experimental settings

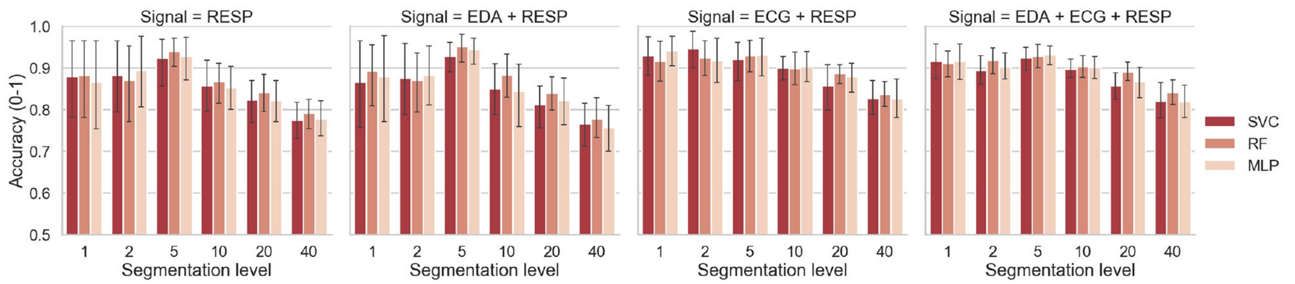


FIGURE 6 | Classification accuracy as a function of the segmentation level and classifier.

TABLE 6 | Best score for each segmentation level, with corresponding signals and classifier.

Segmentation level	Best selection of signals	Best classifier	Best accuracy [Mean (SD)]
1	ECG + RESP	MLP	0.94 (0.06)
2	ECG + RESP	SVC	0.95 (0.07)
5	**EDA + RESP**	**RF**	**0.95 (0.05)**
10	ECG+RESP/EDA+ECG+RESP	SVC	0.90 (0.05)
20	ECG+RESP/EDA+ECG+RESP	RF	0.89 (0.04)
40	ECG+RESP/EDA+ECG+RESP	RF	0.84 (0.05)

Bold values show the best score achieved by the model.

varied from one study to another: the driving environment, the task to complete or the classification procedure. The tasks performed by participants in previous studies were either visual or auditory. The only study in which the task was similar to the task administered in our experiment is the one led by Solovey et al. (2014). Participants had to perform the auditory n-back task and answer verbally to targets. Overall, if we compare our results with those of the latter study based on accuracy measurement, a better accuracy was achieved in this work, probably because the features were calculated over a 20-min time window.

6.3. Interpretation of Results for the Effect of Segmentation Level

In this study, the effect of segmentation on the performance of the model was also investigated. For each driver, the physiological signals collected during the driving session were split into several parts (from 1 to 40) and physiological indicators were computed for each segment. Regardless of the classifier and the chosen physiological signals, increasing the segmentation level from 1 to 5 showed to increase the accuracy of the model. Especially for respiration alone and respiration with EDA signals, the model gained around 10% of accuracy, as shown on **Figure 6**. For these signals that gave a lower accuracy with a segmentation level of 1, the model benefited from sensor fusion. The features computed on 4-min time windows (segmentation of 5) were more accurate to depict the condition of drivers. For segmentation levels of 5–40, increasing the segmentation level showed to decrease the accuracy, regardless of the selected signals. Even if the model had more training example for the classification task, it was more difficult to predict the driver's condition when the features were computed on time windows shorter than 4 min. However, even with 30-s time windows, the model was still able to achieve 84% accuracy with both ECG and respiration and the three signals (**Table 6**). For small time windows, Solovey et al. (2014) also found that enlarging the time window used for computing features (e.g., decreasing the segmentation level) increases the accuracy of the model. Again, if we compare our results with those of the latter study based on accuracy measurement, a lower accuracy was obtained in our study over 30-s time windows (84 vs. 90%). However, they used sliding windows to compute features so their model probably had more training data than our model to maximize its performance. Finally, another interesting result is that increasing the segmentation level showed to reduce the variance for some pairs of signals (error bars on **Figure 6**). This would suggest that the model could be more robust if a reduced time window is used for evaluating driver's state.

6.4. Selection of Best Trade-Off Between Performance and Number of Physiological Signals

For an implementation of a model able to classify drivers' MWL in future automated cars, the goal is to select the best trade-off between the length of the time window used to compute features from the physiological signals and the performance of the system. Based on results obtained in this study, we would select a time window of 4 min to compute features since that segmentation level gave the best accuracy with low variance. For the selection of signals, three options would be possible based on results obtained in this work. The first option would be to choose only the respiration alone as input signal. It would facilitate the implementation of such model under real-world conditions since only one sensor would be necessary to detect driver's MWL with a high accuracy (over 90%). The respiration could be measured either using non-contact respiratory monitoring methods (Min et al., 2010; Al-Khalidi et al., 2011) or contact-based methods using a piezoelectric sensor mounted in the seat belt. The second option would be to select the three signals, because it showed the lowest variance over the 10-folds, meaning that the prediction in real-time of driver's condition would be more reliable from one time to the next. If EDA and ECG would be selected as inputs signals in addition of RESP, both signals could be collected from an intelligent watch or from sensors integrated in smart garments (Sonderegger, 2013; Schneegass et al., 2015). The third option would be to use EDA and respiration signals, that showed the highest performance (high accuracy and low variance) in the present study. In practice, recent advances in technology allow for a continuous recording of the EDA and ECG signals. EDA can be collected from wearable devices such as watches, but they might not give measures as sensitive as the ones obtained with the gold-standard sensors used in this experiment. This can be explained by the lower sensitivity of the wrist tissue and the lower density of eccrine glands in this area compared to the volar surface of the hands (palms) or the foot (sole) (Taylor and Machado-Moreira, 2013). Besides, watches are currently using a plethysmograph sensor and do not provide fine-grained HRV features. However, we must consider that advances in wearable devices and smart clothes might give the possibility to collect robustly and continuously ECG, EDA and respiration signals in a near future. Technologies such as built-in sensors in the seat, radars or intelligent clothes with electrodes such as socks or chest strap are conceivable. Since, we cannot predict the pace of development of new technologies in the field of smart sensors and garments, we make here a proposition based on the empirical results of this work, taking into account the three physiological signals. Therefore, based on results obtained in this experimental study it can be argued that the combination of EDA and respiration signals with a time window of 4 min should be selected for an optimal prediction of orally induced MWL in conditionally automated driving. In previous studies, physiological indicators showed a great potential to detect an increase of driver's MWL due to the performance of a secondary task while driving manually. This study showed that it is also possible to use such indicators for distinguishing two different levels of driver's MWL at a higher level of automation. Previous findings on MWL evaluation can hence be considered in the specific context of automated driving (Level 3 or more according to the SAE taxonomy; SAE, 2018). Therefore, physiological sensors could be worn by drivers so that the car could evaluate their state continuously in conditionally automated driving. This evaluation of driver's state could be used by the car along with the evaluation of the driving situation to provide an optimal support to the driver through in-car interfaces. However, the successful implementation and acceptance of such algorithm depends on people's willingness to wear such sensors in the car. However, although highly interesting and challenging for the future development of the car industry, this is a different topic and not the subject of this paper.

6.5. Limitations and Further Research

There are several limitations that need to be discussed. A first limitation is that the verbal task might have influenced the respiratory pattern of subjects and therefore influenced our physiological indicators (Cacioppo et al., 2007). Therefore, based on the present findings, we can only state that a higher level of MWL induced by a continuous verbal task can be accurately detected in the context of conditionally automated driving. Future research needs to be conducted to investigate to what extent similar results could be obtained in situations of high MWL induced by a task that does not require the participants to speak. In addition, only a subset of all available features were used for the classification task. However, some features that were excluded could have been useful for the classification because of their correlation with other features. Therefore, different strategies for feature selection should be explored. Also, similar experiments should be conducted to collect physiological data from drivers performing cognitive tasks that involve other modalities. The visual and/or auditory n-back task (without verbal answer) could be used to manipulate the MWL of drivers, as done in previous studies (Mehler et al., 2009; Son et al., 2013; Hogervorst et al., 2014; Solovey et al., 2014). Therefore, the model developed as part of this study needs to be evaluated using other NDRTs.

Another stake for the emergence of driver's state systems under real conditions is to be able to evaluate MWL in real-time. Since, Solovey et al. (2014) obtained an accuracy of around 90% using sliding windows of 30 s, it would be interesting to test the effect of sliding windows on our data to generate more training examples and see if it increases the performance of the model. Further studies should focus on evaluating MWL on shorter epochs of cognitive task. The duration of the task performed by drivers in this study was rather long (20 min). Even if the segmentation of data was tested, it might have facilitated the model to achieve good results. Another experiment should be led with participants performing cognitive NDRTs on shorter periods. In this way, it would be closer to reality because drivers might not perform verbal task during 20 min. Based on the new collected data, the same model will be tested to see if it still performs well to predict the MWL level of drivers on shorter periods. If the model's performance decreases too much, the model will need to be refined. To do that, we could consider using model architectures that are efficient with temporal data such as recurrent neural networks. The perceptron used in the MLP classifier would be replaced by gated recurrent units or long-short term memory cells (Hochreiter and Schmidhuber, 1997).

7. CONCLUSION

The main contribution was to use machine learning techniques to specify drivers' condition (verbal task or no task). Three different classifiers along with sensor fusion and six levels of data segmentation were compared. Results show that the model was able to successfully classify the state of the driver with an accuracy of 95% using physiological features from two signals, computed from 4-min windows. The model benefited from sensors' fusion when the respiration and ECG were both selected as input signals. We also showed that increasing the segmentation level from 1 to 5 increased the performance of the classifiers, but increasing the segmentation level from 5 to 40 decreased the performance. For the concrete implementation of such a model under real driving conditions, a fusion of EDA and respiration signals with a time window of 4 min should be considered to compute physiological features in order to classify drivers' MWL in conditionally automated driving.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the dataset is saved in a repository and used only in the context of the AdVitam project. Requests to access the datasets should be directed to Quentin Meteier, quentin.meteier@hes-so.ch.

ETHICS STATEMENT

This study involving human participants was reviewed and approved by Ethics Committee of the Department of Psychology of the University of Fribourg. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

EM, AS, LA, MW and OA developed the research question. QM and AS were responsible for study design and data collection. QM, SR and MC developed machine learning algorithms and conducted data analysis. QM and AS wrote the manuscript. QM and AS designed the experimental procedure and collected the data. QM and SR implemented the classification procedure. All authors contributed to the writing and revision processes.

FUNDING

This work has been funded by Hasler Foundation (Switzerland).

ACKNOWLEDGMENTS

This work has been supported by Hasler Foundation in the framework of the AdVitam project. The authors would like to thank all the persons who contributed to this paper. Special thanks are due to Christian Océane Minot, Laura Rege Colet, Stanislav Riss and Christian Schoepfer, for their help in collecting the data.

REFERENCES

Aasman, J., Mulder, G., and Mulder, L. J. M. (1987). Operator effort and the measurement of heart-rate variability. *Hum. Factors* 29, 161–170. doi: 10.1177/001872088702900204

Al-Khalidi, F. Q., Saatchi, R., Burke, D., Elphick, H., and Tan, S. (2011). Respiration rate monitoring methods: a review. *Pediatr. Pulmonol.* 46, 523–529. doi: 10.1002/ppul.21416

Angelini, L., Khaled, O. A., Caon, M., Mugellini, E., and Lalanne, D. (2014). "Hugginess: encouraging interpersonal touch through smart clothes," in *ISWC '14 Adjunct* (Seattle, WA). doi: 10.1145/2641248.2641356

Baek, H., Cho, C.-H., Cho, J., and Woo, J. (2015). Reliability of ultra-short-term analysis as a surrogate of standard 5-min analysis of heart rate variability. *Telemed. J. e-Health* 5, 404–414. doi: 10.1089/tmj.2014.0104

Bernston, G. G., Cacioppo, J. T., and Quigley, K. S. (1993). Respiratory sinus arrhythmia: autonomic origins, physiological mechanisms, and psychophysiological implications. *Psychophysiology* 30, 183–196. doi: 10.1111/j.1469-8986.1993.tb01731.x

Boucsein, W. (2012). *Electrodermal Activity*. Springer US. doi: 10.1007/978-1-4614-1126-0

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Brookhuis, K., Waard, D., and Samyn, N. (2004). Effects of mdma (ecstasy), and multiple drugs use on (simulated) driving performance and traffic safety. *Psychopharmacology* 173, 440–445. doi: 10.1007/s00213-003-1714-5

Brookhuis, K. A., and De Waard, D. (2001). "Assessment of drivers' workload: performance, subjective and physiological indices," in *Stress, Workload and Fatigue*, P. A. Hancock and P. A. Desmond (Lawrence Erlbaum Associates), 321–333. doi: 10.1201/b12791-2.5

Brookhuis, K. A., and de Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. *Accident Anal. Prevent.* 42, 898–903. doi: 10.1016/j.aap.2009.06.001

Bulmer, M., De Vaus, D. A., and Fielding, N. (2004). *Questionnaires*. London; Thousand Oaks, CA: Sage Publications.

Cacioppo, J. T., Tassinary, L. G., and Berntson, G. G. (Eds.) (2007). *Handbook of Psychophysiology, 3rd Edn*. Cambridge University Press. doi: 10.1017/CBO9780511546396

Camm, A. J., Malik, M., Bigger, J. T., Breithardt, G., Cerutti, S., Cohen, R. J., et al. (1996). Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task force of the European society of cardiology and the north American society of pacing and electrophysiology. *Circulation* 93, 1043–1065.

Claesen, M., and De Moor, B. (2015). Hyperparameter search in machine learning. *arXiv preprint arXiv:1502.02127*.

Collet, C., Clarion, A., Morel, M., Chapon, A., and Petit, C. (2009). Physiological and behavioural changes associated to the management of secondary tasks while driving. *Appl. Ergon.* 40, 1041–1046. doi: 10.1016/j.apergo.2009.01.007

Darzi, A., Gaweesh, S. M., Ahmed, M. M., and Novak, D. (2018). Identifying the causes of drivers' hazardous states using driver characteristics, vehicle kinematics, and physiological measurements. *Front. Neurosci.* 12:568. doi: 10.3389/fnins.2018.00568

De Waard, D. (1997). *The measurement of drivers' mental workload* (Ph.D. thesis). Traffic Research Centre, University of Groningen, Haren, Netherlands.

Engström, J., Johansson, E., and Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transport. Res. F Traffic Psychol. Behav.* 8, 97–120. doi: 10.1016/j.trf.2005.04.012

Ferreira, E., Ferreira, D., Kim, S., Siirtola, P., Roning, J., Forlizzi, J. F., et al. (2014). "Assessing real-time cognitive load based on psycho-physiological measures for younger and older adults," in *2014 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)* (Orlando, FL: IEEE), 39–48. doi: 10.1109/CCMB.2014.7020692

Fisk, A. D., Derrick, W. L., and Schneider, W. (1986). A methodological assessment and evaluation of dual-task paradigms. *Curr. Psychol. Res. Rev.* 5, 315–327. doi: 10.1007/BF02686599

Gawron, V. J. (2019). *Human Performance, Workload, and Situational Awareness Measures Handbook*. CRC Press. doi: 10.1201/9780429019562

Greco, A., Valenza, G., Lanata, A., Scilingo, E. P., and Citi, L. (2016). cvxEDA: a convex optimization approach to electrodermal activity processing. *IEEE Trans. Biomed. Eng.* 63, 797–804. doi: 10.1109/TBME.2015.2474131

Haapalainen, E., Kim, S., Forlizzi, J. F., and Dey, A. K. (2010). "Psycho-physiological measures for assessing cognitive load," in *Proceedings of the 12th ACM International Conference on Ubiquitous Computing - Ubicomp '10* (Copenhagen: ACM Press), 301. doi: 10.1145/1864349.1864395

Hamilton, P. (2002). "Open source ECG analysis," in *Computers in Cardiology* (Memphis, TN: IEEE), 101–104.

Hart, S. G., and Staveland, L. E. (1988). "Development of NASA-TLX (task load index): results of empirical and theoretical research," in *Advances in Psychology, Volume 52 of Human Mental Workload*, eds P. A. Hancock and N. Meshkati (Amsterdam: North Holland), 139–183. doi: 10.1016/S0166-4115(08)62386-9

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd Edn*. New York, NY: Springer-Verlag.

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput*. 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735

Hogervorst, M. A., Brouwer, A.-M., and van Erp, J. B. F. (2014). Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. *Front. Neurosci*. 8:322. doi: 10.3389/fnins.2014.00322

Hsu, C.-W., Chang, C.-C., and Lin, C.-J. (2010). A practical guide to support vector classification. *Technical report*.

Huigen, E., Peper, A., and Grimbergen, C. A. (2002). Investigation into the origin of the noise of surface electrodes. *Med. Biol. Eng. Comput*. 40, 332–338. doi: 10.1007/BF02344216

Ikehara, C., and Crosby, M. (2005). "Assessing cognitive load with physiological sensors," in *Proceedings of the 38th Annual Hawaii International Conference on System Sciences* (Big Island, HI), 295a.

Kim, H., Kim, W., Kim, J., Lee, S.-J., and Yoon, D. (2019). "A study on the effects of providing situation awareness information for the control authority transition of automated vehicle," in *2019 International Conference on Information and Communication Technology Convergence (ICTC)* (Jeju Island), 1394–1396. doi: 10.1109/ICTC46691.2019.8939867

Krueger, E., Schneider, A., Sawyer, B. D., Chavaillaz, A., Sonderegger, A., Groner, R., et al. (2019). Microsaccades distinguish looking from seeing. *J. Eye Movement Res*. 12, 1–14. doi: 10.16910/jemr.12.6.2

Le, A. S., Aoki, H., Murase, F., and Ishida, K. (2018). A novel method for classifying driver mental workload under naturalistic conditions with information from near-infrared spectroscopy. *Front. Hum. Neurosci*. 12:431. doi: 10.3389/fnhum.2018.00431

Lewis, G. F., Furman, S. A., McCool, M. F., and Porges, S. W. (2012). Statistical strategies to quantify respiratory sinus arrhythmia: are commonly used metrics equivalent? *Biol. Psychol*. 89, 349–364. doi: 10.1016/j.biopsycho.2011.11.009

Li, N., and Busso, C. (2013). "Analysis of facial features of drivers under cognitive and visual distractions," in *2013 IEEE International Conference on Multimedia and Expo (ICME)* (San Jose, CA: IEEE), 1–6. doi: 10.1109/ICME.2013.6607575

Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., et al. (2021). Neurokit2: a python toolbox for neurophysiological signal processing. *Behav. Res. Methods*. doi: 10.3758/s13428-020-01516-y

Malik, M., and Terrace, C. (1996). Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. *Eur. Heart J*. 17, 354–381. doi: 10.1093/oxfordjournals.eurheartj.a014868

Mehler, B., Reimer, B., and Coughlin, J. (2012). Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task. *Hum. Factors* 54, 396–412. doi: 10.1177/0018720812442086

Mehler, B., Reimer, B., Coughlin, J., and Dusek, J. (2009). The impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transport. Res. Rec*. 2138, 6–12. doi: 10.3141/2138-02

Merat, N., Seppelt, B., Louw, T., Engstrom, J., Lee, J. D., Johansson, E., et al. (2019). The "Out-of-the-Loop" concept in automated driving: proposed definition, measures and implications. *Cogn. Technol. Work* 21, 87–98. doi: 10.1007/s10111-018-0525-8

Min, S. D., Kim, J. K., Shin, H. S., Yun, Y. H., Lee, C. K., and Lee, M. (2010). Noncontact respiration rate measurement system using an ultrasonic proximity sensor. *IEEE Sensors J*. 10, 1732–1739. doi: 10.1109/JSEN.2010.2044239

Mulder, L. J. M. (1992). Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biol. Psychol*. 34, 205–236. doi: 10.1016/0301-0511(92)90016-N

Nakajima, Y., and Tanaka, K. (2017). "Effects of active and passive secondary tasks in a take-over situation during automated driving," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (Banff, AB), 1161–1166. doi: 10.1109/SMC.2017.8122769

NHTSA (2017). *Distracted Driving in Fatal Crashes (Traffic Safety Facts Research Note*. National Center for Statistics and Analysis.

Parasuraman, R., Sheridan, T., and Wickens, C. (2000). A model for types and levels of human interaction with automation. *IEEE Trans. Syst. Man Cybernet. A Syst. Hum*. 30, 286–297. doi: 10.1109/3468.844354

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res*. 12, 2825–2830. doi: 10.5555/1953048.2078195

Petermeijer, S., Bazilinskyy, P., Bengler, K., and de Winter, J. (2017). Take-over again: investigating multimodal and directional TORs to get the driver back into the loop. *Appl. Ergon*. 62, 204–215. doi: 10.1016/j.apergo.2017.02.023

Pettitt, M., Burnett, G., and Stevens, A. (2005). "Defining driver distraction," in *Intelligent Transportation Society of America - 12th World Congress on Intelligent Transport Systems* (San Francisco, CA), 5.

SAE. (2018). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*.

Salahuddin, L., Cho, J., Jeong, M. G., and Kim, D. (2007). "Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Lyon), 4656–4659. doi: 10.1109/IEMBS.2007.4353378

Schlegel, R. E. (1993). Chapter 17/21, Driver mental workload. *Automot. Ergon*. 359–382.

Schneegass, S., Hassib, M., Zhou, B., Cheng, J., Seoane, F., Amft, O., et al. (2015). "Simpleskin: towards multipurpose smart garments," in *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers, UbiComp/ISWC'15 Adjunct* (New York, NY: Association for Computing Machinery), 241–244. doi: 10.1145/2800835.2800935

Shaffer, F., and Ginsberg, J. (2017). An overview of heart rate variability metrics and norms. *Front. Public Health* 5:258. doi: 10.3389/fpubh.2017.00258

Siegenthaler, E., Costela, F. M., McCamy, M. B., Di Stasi, L. L., Otero-Millan, J., Sonderegger, A., et al. (2014). Task difficulty in mental arithmetic affects microsaccadic rates and magnitudes. *Eur. J. Neurosci*. 39, 287–294. doi: 10.1111/ejn.12395

Smith, S. W. (1999). *The Scientist and Engineer's Guide to Digital Signal Processing*. California Technical Pub., San Diego, CA.

Solovey, E. T., Zec, M., Garcia Perez, E. A., Reimer, B., and Mehler, B. (2014). "Classifying driver workload using physiological and driving performance data: two field studies," in *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI '14* (Toronto, ON: ACM Press), 4057–4066. doi: 10.1145/2556288.2557068

Son, J., Oh, H., and Park, M. (2013). Identification of driver cognitive workload using support vector machines with driving performance, physiology and eye movement in a driving simulator. *Int. J. Precis. Eng. Manufactur*. 14, 1321–1327. doi: 10.1007/s12541-013-0179-7

Sonderegger, A. (2013). "Smart garments-the issue of usability and aesthetics," in *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication, UbiComp '13 Adjunct* (New York, NY: Association for Computing Machinery), 385–392. doi: 10.1145/2494091.2495969

Taylor, N., and Machado-Moreira, C. (2013). Regional variations in transepidermal water loss, eccrine sweat gland density, sweat secretion rates and electrolyte composition in resting and exercising humans. *Extreme Physiol. Med*. 2:4. doi: 10.1186/2046-7648-2-4

Wandtner, B., Schomig, N., and Schmidt, G. (2018). Effects of non-driving related task modalities on takeover performance in highly automated driving. *Hum. Factors* 60, 870–881. doi: 10.1177/0018720818768199

Wickens, C. D. (2008). Multiple resources and mental workload. *Hum. Factors* 50, 449–455. doi: 10.1518/001872008X288394

Young, M. S., Brookhuis, K. A., Wickens, C. D., and Hancock, P. A. (2015). State of science: mental workload in ergonomics. *Ergonomics* 58, 1–17. doi: 10.1080/00140139.2014.956151

Zijlstra, F., and Doorn, L. (1985). *The Construction of a Scale to Measure Perceived Effort*. Department of Philosophy and Social Sciences.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 19 February 2021
doi: 10.3389/fpsyg.2021.596278

The Effects of an In-vehicle Collision Warning System on Older Drivers' On-road Head Movements at Intersections

Rachel Shichrur[1]*, *Navah Z. Ratzon*[2], *Arava Shoham*[2,3] *and Avinoam Borowsky*[4]

[1] Occupational Therapy Department, Faculty of Health Sciences, Ariel University, Ariel, Israel, [2] Occupational Therapy Department, School of Health Professions, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv-Yafo, Israel, [3] Occupational Therapy Clinics, Clalit Health Services, Dimona, Israel, [4] The Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, Be'er-Sheva, Israel

OPEN ACCESS

Edited by:
Haneen Farah,
Delft University of Technology, Netherlands

Reviewed by:
Sophie Lemonnier,
EA7312 Laboratoire de Psychologie Ergonomique et Sociale pour l'Expérience Utilisateurs (PERSEUS), France
Carryl L. Baldwin,
Wichita State University, United States

***Correspondence:**
Rachel Shichrur
rachelsh@ariel.ac.il

Specialty section:
This article was submitted to Performance Science, a section of the journal Frontiers in Psychology

Received: 18 August 2020
Accepted: 25 January 2021
Published: 19 February 2021

Citation:
Shichrur R, Ratzon NZ, Shoham A and Borowsky A (2021) The Effects of an In-vehicle Collision Warning System on Older Drivers' On-road Head Movements at Intersections. Front. Psychol. 12:596278. doi: 10.3389/fpsyg.2021.596278

With age might come a decline in crucial driving skills. The effect of a collision warning system (CWS) on older drivers' head movements behavior at intersections was examined.

Methods: Twenty-six old-adults, between 55 and 64 years of age, and 16 Older drivers between 65 and 83 years of age, participated in the study. A CWS (Mobileye Inc.) and a front-back in-vehicle camera (IVC) were installed in each of the participants' own vehicles for 6 months. The CWS was utilized to identify unsafe events during naturalistic driving situations, and the IVC was used to capture head direction at intersections. The experimental design was conducted in three phases (baseline, intervention, and carryover), 2 months each. Unsafe events were recorded by the CWS during all phases of the study. In the second phase, the CWS feedback was activated to examine its effect on drivers' head movement' behavior at intersections.

Results: Older drivers (65+) drove significantly more hours in total during the intervention phase ($M = 79.1$ h, $SE = 10$) than the baseline phase ($M = 39.1$ h, $SE = 5.3$) and the carryover phase ($M = 37.7$ h, $SE = 5.4$). The study revealed no significant differences between the head movements of older and old-adult drivers at intersections. For intersection on the left direction, a significant improvement in drivers' head movements' behavior was found at T-junctions, turns and four-way intersections from phase 1 to phase 3 ($p < 0.01$), however, two intersection types presented a decrease along the study phases. The head movements' behavior at roundabouts and merges was better at phase 1 compared to phase 3 ($p < 0.01$). There was no significant reduction of the mean number of CWS unsafe events across the study phases.

Conclusions: The immediate feedback provided by the CWS was effective in terms of participants' head movements at certain intersections but was harmful in others. However, older drivers drove many more hours during the active feedback phase, implying that they trusted the system. Therefore, in the light of this complex picture, using the technological feedback with older drivers should be followed with an additional mediation or follow-up to ensure safety.

Keywords: technology—assistive/supportive, older drivers, in-vehicle camera, feedback, head movements, naturalistic driving

INTRODUCTION

Older Drivers' Safety

The percentage of older individuals (typically defined as ≥65 years; Vespa et al., 2018; National Highway Traffic Safety Administration, 2020). in society has been steadily increasing worldwide and is expected to reach 90 million in 2050 in only the United States. This population constitutes about a quarter of all licensed drivers (Pomidor, 2015). In 2018, 6,907 drivers above the age of 65 were killed on US roads, constituting 19% of all road fatalities in the US (National Highway Traffic Safety Administration, 2020). In modern life, older adults, similar to all age groups, are dependent on driving as the primary mode of transportation, allowing them to maintain autonomy (Classen, 2010). However, the prevalence of medical impairments, decline in vision, cognition, motor abilities, and somatosensory functions rises with age, which may have substantial effects on driving skills, including the ability to perform proper scanning (Karthaus and Falkenstein, 2016; Samuel et al., 2016).

Researchers suggest that the underlying frailty, medical conditions, and medication-use contribute significantly to crash disparities between older and younger drivers and increased risks of injury and fatality in older drivers (McGwin et al., 2000; Langford and Koppel, 2006). The data of Bédard et al. (2002) showed that older drivers are more vulnerable to the traumatic effects of crashes. The odds of a fatal injury for older drivers (65–79) were 2.3 times of that of drivers aged 40–49 and the odds of a fatal injury for drivers older than 80 was even five times more of that of the younger drivers. Consistently, several studies confirmed that drivers aged ≥65 pose danger to themselves and to other road users as compared with drivers at younger age groups (Dellinger et al., 2004; Awadzi et al., 2008). Findings from a driving simulator study (Park et al., 2017) indicated that older adults (65.6 $\pm$ 5 years) have some limitations, primarily relating to left turns against oncoming traffic and while overtaking a lead vehicle. In a study by Bao and Boyle (2009), for example, older drivers (65–80) had a significantly smaller proportion of visual sampling to the left and right-hand side of the intersection during intersection negotiations when compared to younger (18–25) and middle-aged (35–55) drivers.

These circumstances raise the need to balance between encouraging independent living and protecting the rights of the safe older drivers, vs. the practitioners' duty to identify unsafe driving and protect other road users.

Vision and Safe Driving

Studies agree that for a driver, vision is crucial for collecting driving-relevant information from the driving environment (Van Houten and Retting, 2001; Green, 2002). Visual attention, a critical skill for avoiding crashes while driving, is used to direct information processing resources (using eye and head movements) to spot potentially important visual events. Older drivers (65+) tend to identify hazards less often when hazards are located in the periphery of the visual scene (Bromberg et al., 2012).

Moreover, safe driving relies on the drivers' ability to make quick head turns and eye movements, scan other spatial locations such as mirrors, lead cars, pedestrians, and road traffic signs, and shift their attention to the road. The timing of performing glances before moving forward at an intersection is critical. It takes 1.8–2.9 s to identify approaching vehicles before leaving the stop line (Hostetter et al., 1986 as cited in Fisher et al., 2016). Most drivers make glances to the left and the right; however, only a few make a secondary glance (Fisher et al., 2016). A secondary glance is defined as "a glance toward an area from which a threat might emerge at a time after the foot moves from the brake, or after the start of acceleration into the intersection when there is not a stop" and is considered "the last best chance to abort the movement into the intersection" (Fisher et al., 2016, p. 94).

In this study, head movements were used as a proxy of gaze position. It has already been shown that for horizontal visual angles larger than 30° an observer must move his head toward the target area and the gaze position follow this (Land and Tatler, 2009). In this study the focus was identifying glance position at intersection, which generally requires scanning at large visual angles. In addition, Metz and Krueger (2010) recommend using head movement analysis instead of eye movement at intersections due to severe data loss of eye movements that often occurs when scanning requires wide visual angles that require head movements. The authors added that in their study head movement was a good and reliable alternative to eye movements. Capturing head movements during a drive is becoming more common especially due to the prevalence of low cost in-vehicle cameras that allow it.

Visual Search for Threats

Visual search is a prominent process that drivers must apply in order to identify road hazards. Visual search evidence in the driving domain shows that scanning patterns are typically different between older and younger-experienced drivers. Bao and Boyle (2009), for example, showed that older drivers do not utilize their full scanning range when compared to middle-aged drivers, and tend to check fewer areas before executing a maneuver through intersections, specifically during left and right turns. Romoser and Fisher (2009), concluded that regardless of driver's cognition, speed-of-processing, or useful field of view (UFOV) status if drivers do not turn their heads to scan for cross-traffic when turning at intersections, they will fail to detect unanticipated vehicles that may conflict with their turn. Inattention can cause drivers to exhibit unsafe behaviors during the driving task, such as greater lane position variability, reduced headway distance, and reduced time-to-collision. Additionally, inattention may also reduce a driver's capability to respond to hazardous situations, as indicated by delayed reactions (Strayer et al., 2003; Lees and Lee, 2007).

Interventions to Improve Older Drivers' Road Scanning Skills

Despite older drivers deteriorated scanning behavior, some studies have shown that visual search for threats is a skill that can be trained and improved. Romoser and Fisher (2009), for example, used a driving simulator to train older drivers (70+). They found that active training with direct feedback (using a driving simulator), compared to passive training (e.g., telling drivers where to look without actually driving), is a more effective strategy for increasing the likelihood that older drivers will look for threats during a turn. Their active training increased the likelihood of hazard detection during a right or left turn by nearly 100% in both post-training simulator tests and field drives (Romoser and Fisher, 2009). Similarly, Pollatsek et al. (2012) were able to consistently train older drivers (70+) to both scan the roadway environment and to learn to allocate their attention more effectively.

Nevertheless, this evidence shows only an initial step toward finding ways to preserve and even improve older drivers' visual search skills in order to make them safer drivers and compensate for age-related deterioration of driving skill. An important aspect that was overlooked in these studies is the possibility of exploiting the advantages of in-vehicle technology and specifically using collision-warning systems (CWS). This exploitation would be used with the purpose of (1) facilitating more efficient road scanning and (2) providing an overall better safe-driving performance while combining humans with technology vs. humans alone.

Although there currently is encouraging evidence regarding the potential of an intervention aimed to improve older drivers' driving performance at least partially, the way toward making older adults safer drivers is still long. This delay is especially worrying if one considers their over-representation in fatal traffic crashes, poor scanning performance, and overall reduced driving performance compared to younger-experienced drivers. Considering that older drivers look less often to the left and right at intersections than younger adults do, the question then becomes whether there is a way to improve older adults' scanning at intersections by using current in-vehicle technology (Caserta and Abrams, 2007; Bao and Boyle, 2009).

In-vehicle Data Recorders (IVDR)

In recent years, due to various technological improvements, IVDR technology offers CWS that can help drivers to pay attention to road hazards and objects to avoid collisions (Maltz and Shinar, 2004; Wang et al., 2016; Hubele and Kennedy, 2018). This technology may be especially valuable when other tasks compete for driver's attention. The task of the warning system is to attract the drivers' attention back on to the road, especially when the road demands increase.

Modern in-vehicle safety technologies offer Advanced Driver Assistance Systems (ADAS) that helps drivers drive safer and pay attention to road hazards. ADAS are becoming more ubiquitous in newer cars and may significantly reduce crashes related to impaired visual search, distraction, or lack of attention. Hickman et al. (2015) have collected retrospective crash data from 14 motor carriers including a total of 151,624 truck-years on different types of roads. The authors have demonstrated that lane departure warning (LDW) significantly reduced a LDW-related crashes by 1.92 times and roll stability control (RSC) significantly reduced a RSC-related crashes by 1.56 times. Integrating between real-time on-road driving data, with systems that monitor information regarding the drivers' scanning behavior (such as IVC), ADAS are designated to help drivers identify unnoticed road hazards as well as use the feedback received by the system to facilitate their road scanning behavior training when distracted (Carr and Grover, 2020). Analyzing the different patterns of spatial attention and driving behavior will assist in modifying inferior behaviors in an effort to improve road safety.

Recruiting such technologies to assist the older driver in driving safer, as well as in taking advantage of the feedback received by the system in order to facilitate their scanning behavior, has been set as a goal for current researchers (Carr and Grover, 2020). This research is expected to provide further insight regarding older drivers' spatial attention and head movements behavior in correlation with unsafe driving-related events on the road. Therefore, the present study proposes an intervention procedure that combines CWS and IVC as an integrated tool to enhance older drivers' safety and awareness of safety while driving. The effects of an CWS's feedback on older drivers' unsafe CWS events and head movements will be analyzed.

Hypotheses

This study has three hypotheses:

1. The feedback-based intervention provided by CWS in phase 2 will be found effective in improving the head movements behavior of study population at intersections and in reducing their involvement in hazardous driving-related events (provided by the CWS).
2. Old-adults group (55–64) will have better head movements at intersections than the older drivers' group (+65).
3. Positive correlations will be found between low quality of the head movements at intersections as measured by IVC and hazardous driving-related events as obtained from CWS in the study phases.

METHODS

The study population included 42 drivers: 26 old-adults (55–64 years old, $M = 59.4$, $SD = 3.1$), and 16 older drivers (65–83 years old, $M = 70.9$, $SD = 5.32$). The study population included 25 men and 17 women. All participants were independent drivers with a valid driver's license and with normal or corrected to normal vision. Recruitment was done via e-mails, and snowball sampling method, targeting old-aged volunteers who pursue an independent lifestyle. Volunteers were compensated for the time they spent participating in the study for 6 months.

To exclude potential effects of depression and cognitive impairments, participants had to score five or fewer points on the Geriatric Depression Scale (Yesavage et al., 1982) and to score above 24 in the Mini-Mental State Examination (Folstein et al., 1975). Other exclusion criteria included a medical history of neurological, orthopedic, and/or psychiatric conditions with permanent impairments or using drugs that, according to the guidelines of the pharmaceutical company, may interfere with driving. In each participant's private vehicle, a CWS and IVC were installed to identify unsafe events during on-road driving and to capture head movements' behavior data at intersections.

Primary Outcome Measure Tools

Collision-Warning Systems (CWS)

ADAS technology offers collision-warning systems (CWS) that provide a forward collision warning (FCW) and a lane departure warning (LDW), helping drivers drive safely. The CWS is a vision-based tool for the vehicle, which continually measures the time/distance to the vehicle in front and lane marks. The ultra-speed data processor includes algorithms that follow lane marks and road curves, detect cars and pedestrians, and follow the headway distance, working in the rain and at night. It is programmed to detect only five types of objects, specifically: trucks, cars, motorcycle, bicycles, and people. The system follows four types of safety outcomes.

(a) *Urban Forward Collision Warning (uFCW)*—A risk warning for urban collision is activated when driving is <30 km/h. The system calculates the time it will take to stop the car without touching the other car, and a risk event is recorded when the distance between the cars is 2.7 s.

(b) *Forward Collision Warning (FCW)*—A risk warning for rear-end collision is activated when driving 30 km/h and above. The system calculates the time it will take to stop the car without touching the other car, and a risk event is recorded when the distance between the cars is 2.7 s.

(c) *Unsafe Headway Warning (HW)*—The warning is activated from 30 km/h and above. The system calculates the time it will take to stop the car without colliding with the object in front of the vehicle, and a risk event is recorded when the distance between the car and one of the aforementioned objects is 1 s.

(d) *Sudden Lane Deviations Warning (LDW)*—A lane departure without signaling warning is activated from 55 km/h and above. The CWS focuses on driving safety and analysis of the technical skill of the driver and provides a driving profile. The riskiness grade used in this study obtained by the CWS is a calculation of the mean number of all types of unsafe events per hour from the four types of safety outcomes (uFCW, FCW, HW, and LDW).

In-vehicle Front-Back Camera (IVC)

Each vehicle was equipped with a dual-lens video camera that captured both the driving scene from the driver's perspective (i.e., driving context) as well as the driver's face. This front-back camera configuration allowed capturing the participant's head direction dependent on the driving context (e.g., driving on a straight road, approaching an intersection). The camera had a high definition (HD) video quality and a 64 GB SD card to allow recording 9 h of driving each time. Once the card was full, it was removed from the vehicle, analyzed on a computer, and replaced with a blank card. Recording the driver's face provided the information regarding head movements.

Data Preparation

As mentioned above, the front-facing camera provided an immense amount of real-world driving data of the traffic environment from a driver's perspective. Thus, at each phase of the study, it was decided to focus on the first 120 intersections that each participant encountered and examine the head movement behavior of each participant at each intersection. The intersections were classified into five types according to their geometric structure (merging road, roundabout, turn, T junction, and four-way intersection). The direction of travel of the driver (right/left/straight and also the presence of secondary glances when needed) and the presence of other road users (vehicles and pedestrians) were registered. Examples of different types of intersections can be seen in **Figures 1A–C**.

Procedure for Coding Research Data

Each coder was given a list of proper road scanning behavior expected at each intersection that was predefined by the research team [a complete list of all head movements demands for each intersection is provided in Appendix A (**Supplementary Material**)]. Each coder was asked to indicate whether the driver scanned the intersection properly (given a score of "2"), whether the road scanning behavior was only partially correct (given a score of "1"), or whether the scanning behavior at that specific intersection was improper (given a score of "0") for each intersection per participant. In cases where a coder was not sure how to classify the driver's head movement behavior at a specific intersection, he or she consulted with the research team, who made a classification decision based on a discussion.

Primary, a pilot data-coding procedure was conducted where four researchers from the research team analyzed two drivers' videos. Each researcher viewed the camera videos independently, identified 30 intersections, classified them, defined the head movements requirements at each intersection, and reviewed the actual head-based behavior of each driver. A between rater reliability test of this pilot data coding procedure showed reliability of only 76.6%. In order to improve the inter-rater reliability percentage, three additional meetings were required in which the researchers consulted on uniformity in interpreting the data obtained from the videos. This process was repeated until the raters were able to achieve inter-rater reliability of over 90%. To promote the research, three additional research assistants, trained by the research team, were recruited to view camera videos and code the participants' horizontal head rotations. Each research assistant viewed the same 30 intersections of the same driver, and classified his head

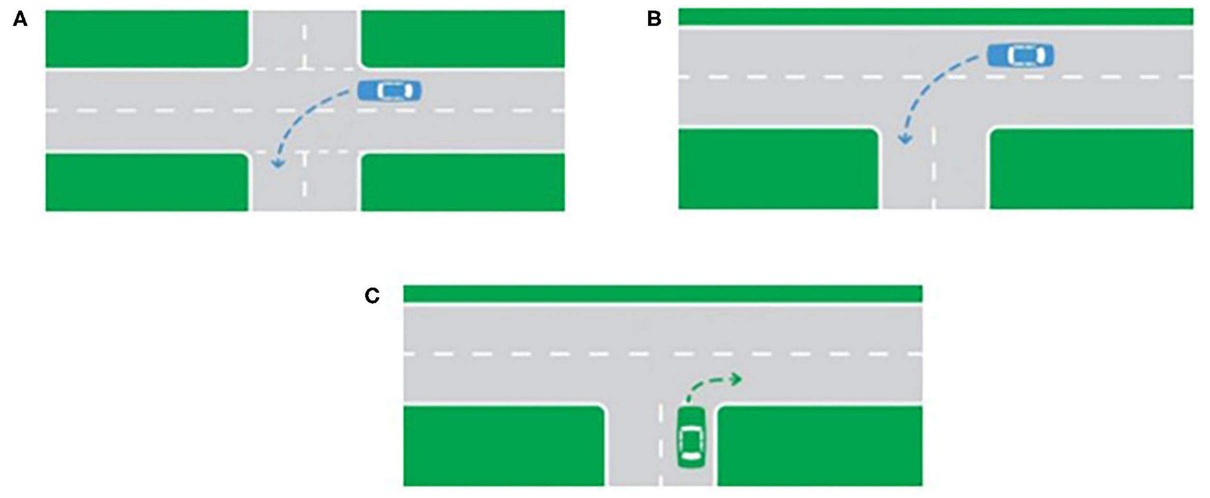


FIGURE 1 | (A) A left turn at a four-way intersection. **(B)** A left turn at a two-lane T junction. **(C)** A right turn at a T junction.

movements as: proper head movements = "2," partially proper head movements = "1," or improper head movements = "0." Checking the inter-rater reliability of the three research assistants in analyzing the same 30 intersections showed high reliability of 97%.

Experimental Design

The experimental design was a *A2<P3*I5*D4>* mixed design. The between-subjects independent variables included the age group (A: old-adults or older drivers). The within-subjects independent variables included the experiment's phase (P: 1, 2, or 3), the intersection type (I: 1–5, see data preparation section), and the travel direction (D: 1–4, see data preparation section). The dependent variables included the score of the head movements (proper head movements = "2," partially proper head movements = "1," or improper head movements = "0"), the number of unsafe events per hour for each one of the four types of safety outcomes (uFCW, FCW, HW, and LDW), and CWS riskiness grade (Mobileye, Inc.).

Procedure

The institutional review at Tel Aviv University board approved this study. Informed consent forms were obtained from all the volunteers prior to commencing the study. After signing the informed consent, each participant completed the questionnaires that were relevant for the screening tools, and only those who met the inclusion criteria were able to participate in the study. Since this study is focused on naturalistic driving, the research location was determined according to the participants' convenience in their community (e.g., a quiet room in a community center or the driving laboratory at the University). Next, the CWS was installed by expert technicians in each of the participants' own vehicle at his/her house. After verifying the system functioned properly, each driver was given a personal identification code number to type in before starting the vehicle. The technicians disassembled the systems after a period of 6 months, at the end of the study.

The study included three phases, and each phase lasted about 2 months. During the first (silent) phase, unsafe events were recorded without the use of active alerts. In the second phase (intervention), the CWS feedback was activated in all the participants' vehicles to examine its effectiveness. In the third phase, the feedback was silenced to examine behavioral change. In each phase, the drivers' head movements were also examined. A comparison was made between the three phases of the study for the two age groups, and also each subject was self-compared between the pre- and post-intervention phase. The study period included a variety of driving events (such as lane deviations and risk of rear-end collisions), under a variety of road or traffic situations (such as dense traffic in urban roads or inter-urban highways) and weather conditions.

Data Analysis

In order to assess the effectiveness of the feedback-based intervention at improving head movements' behavior, several statistical tests were carried out within the Generalized Linear Mixed Models (GLMM) framework. The head movements behavior analysis, included a logistic regression model with a logit link function. The independent variables that were included in the initial model included: age group (old-adults and older drivers), gender, phase (1, 2, or 3), and intersection type (1–5). The initial model also included three second-order interactions of intersection type*phase, age group*phase, and intersection type*age group, and a third-order interaction of phase*age group*intersection type. Participants were also included into the model as a random effect. The dependent variable included head movements' behavior that was coded as a binary variable (proper head movements behavior = 1 and Improper head movements behavior = 0). Notably, although we initially coded partially proper behavior as well, due to the small number of cases they were not included in the analysis. This logistic regression model was applied twice: once for the left travel direction and once for the right travel direction. A backwards elimination procedure was applied for each model such that non-significant interactions were removed in case they were not statistically significant.

In order to statistically examine the effect of the feedback-based intervention in reducing risk-related driving events we applied a repeated measures analysis within the GLM framework in SPSS. This model was applied twice. One model included the average number of CWS events per km driven per participant as the dependent variable and the other model included the average number of CWS events per hour driven per participant. Phase was included as the independent variable in both models.

In addition, Pearson correlations were used to examine the correlations between head movements behavior and unsafe driving events. Analysis of the research videos revealed that drivers behaved differently at traffic light intersections as opposed to unsignalized intersections because they were guided by the light signals and did not have to scan the environment themselves. Thus, signalized intersections were excluded from the study. As a result, the percentage of four-ways intersections and T junctions was low relative to other intersections in the sample as in Israel, traffic lights control most In order to statistically examine driving characteristics differences between the two age groups and across the three phases of the study, two linear regression models were applied within the framework of General Linear Model (GLM). This GLM was applied twice: once for number of kilometers driven and once for number of hours driven as dependent variables. The independent variables that were included in the initial models were: phase (1, 2, or 3), and age group (old-adults and older drivers). The models included a second-order interaction of age group*phase. Participants were included as a random effect of these types of intersections.

All analyses were carried out at a significant level of 5 percent. Next, all significant effects of the final model were further analyzed using *post-hoc* pairwise contrast comparisons analysis where the Bonferroni correction procedure for multiple comparisons was also applied whenever it was required.

RESULTS

The results chapter includes four main sections. The first section presents an examination of the feedback-based intervention efficacy on the study population (hypothesis 1). The second section presents the analyses related to the participants' head movements at intersections and differences between age groups (hypothesis 2). The third section presents the correlations between head movements at intersections and unsafe events (CWS) (hypothesis 3). Section four presents additional analyses related to driving exposure in terms of hours driven and distance traveled.

TABLE 1 | A summary of the final logistic regression model's fixed effects of head movements.

Fixed variables	*F*	df2	df1	sig
Corrected model	46.9	3,638	16	0.00
Age group	0.03	3,638	1	0.86
Gender	0.37	3,638	1	0.54
Phase	4.26	3,638	2	0.01
Intersection type	135.7	3,638	4	0.00
Intersection type*phase	7.9	3,638	8	0.00

$^*p < 0.05$, $^{**}p < 0.01$.

Section 1: Examining the Efficacy of the Feedback-Based Intervention on the Study Population

Hypothesis 1 estimated that the feedback-based intervention provided in phase 2 by the CWS will be found effective in improving the head movements of study population at intersections. In order to assess the effectiveness of the feedback-based intervention at improving head movements' behavior and reducing unsafe driving events, the study used a linear regression within the GLMM framework. As presented in **Table 1**, the phase variable was found significant [$F_{(2,3,638)} = 4.26, p < 0.01$]. **Table 1** presents a summary of the significant fixed effects that were included in the final model.

The final regression model for examining the head movements showed that the independent variables entered into the model found to be significant were phase, intersection type and their interaction. *Post-hoc* pairwise contrast comparisons of the significant interaction between the intersection type and the phases of the study are shown in **Figure 2**. The Y axis presents the estimated mean probability of proper head movements at intersections.

According to **Figure 2**, T-junctions, turns and four-way intersections drivers' head movements' behavior was better at phases 2 and 3 compared to phase 1 ($p < 0.01$). A significant difference was found between phases 1 and 3 ($p < 0.01$). There was no significant difference between phases 2 and 3 for these three types of intersections. However, two intersection types presented a decrease along the study phases. Roundabout and merges head movements' behavior was better at phase 1 compared to phase 3 ($p < 0.01$). At merges the decrease from phase 1 to phase 2 was significant ($p < 0.01$) but there was no significant difference between phases 2 and 3 and for roundabout there was no significant difference between phases 1 and 2 but the decrease from phase 2 to phase 3 was significant ($p < 0.01$).

In order to examine the effect of the feedback-based intervention in reducing risk-related driving events as mentioned in hypothesis 1, a repeated measures analysis within the GLM framework was applied. Both models didn't present any significant differences between the average number of CWS events (neither the model per hour nor the model per km driven) per participant. **Figure 3** illustrates driving patterns according to CWS unsafe events per hour before, during, and after the intervention.

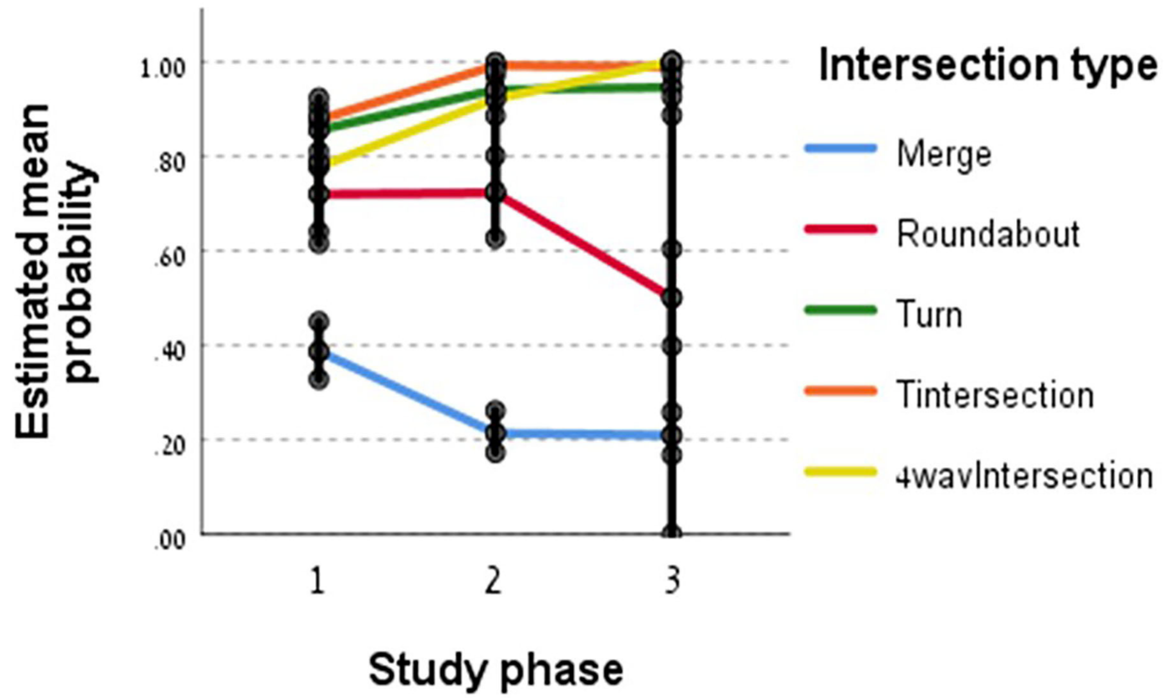


FIGURE 2 | Interaction between intersection types and study phases.

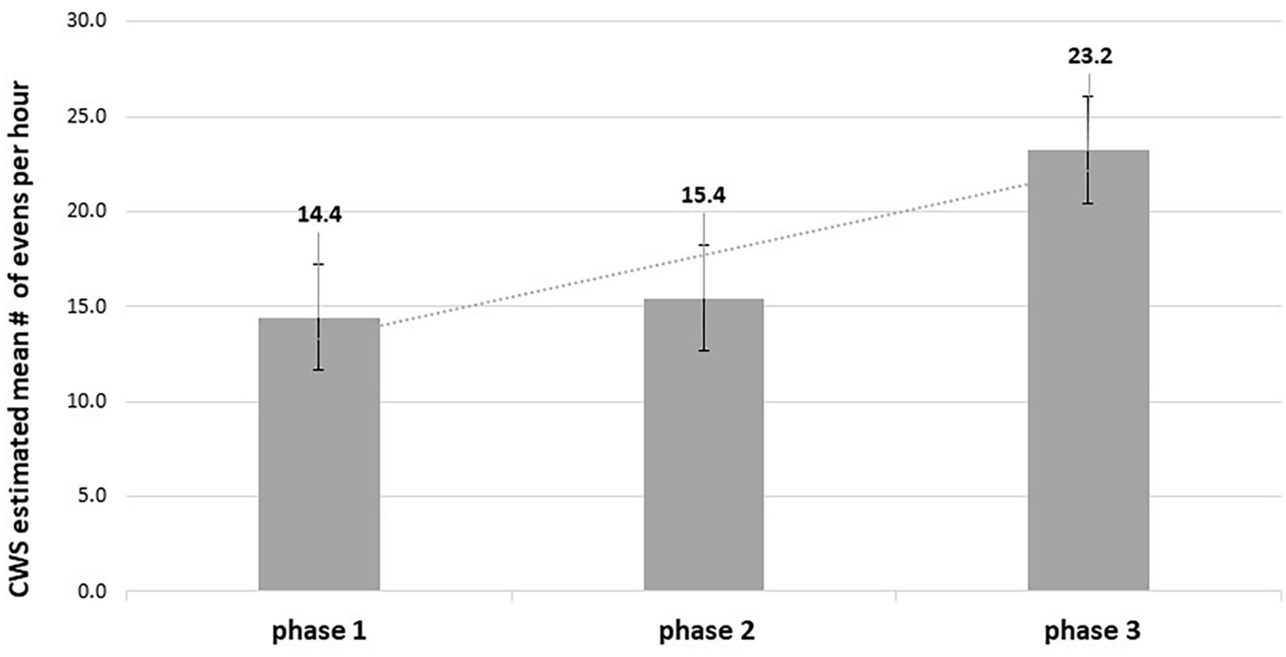


FIGURE 3 | CWS estimated mean number of events per hour throughout the study phases ($n = 40$).

The CWS mean number of unsafe events per hour refers to the whole drivers' population. According to **Figure 3** although the mean number of safety incidents per hour is noticeably increasing across the study' phases, there were no significant differences between the mean number of CWS unsafe events per hour across the phases of the study.

Section 2: Analysis of Head Movements' Behavior at Intersections

As noted in the method section, the dependent variable representing the quality of the head movements' behavior was coded into three levels: proper head movements' behavior = "2," partially proper head movements' behavior = "1," and improper head movements' behavior = "0." The Percentage of improper, partially proper and proper head movements were computed for each type of intersection (5 types altogether). The descriptive information regarding improper head movements' behavior is presented in **Table 2**.

As showed in **Table 2**, most improper head movements at intersections were associated with right or left merges, forward four-ways intersection, and roundabout to the left across the study phases.

Hypothesis 2 estimated that the old-adults group (55–64) will have better head movements at intersections than the older drivers group (+65). The purpose of the current analysis was to examine this assumption. In order to test whether the differences in the head movements' behavior between the old-adults and the older drivers are significant across the three study's phases, we used a logistic regression model within the GLMM framework as noted in the data analysis section. Notably, since the model of all 5 intersection types at the right direction did not yield significant effects, it is not presented here, and the study focuses on left direction only. As shown in **Table 1**, the final logistic regression model revealed that no effect was statistically significant for the age group and its interactions with the study phase or intersection type. In other words, there were no significant differences between the mean percentage of head movements' behavior of older and old-adult drivers at intersections across the three phases of the study, suggesting that hypothesis 2 was rejected.

Section 3: Correlations Between Head Movements at Intersections and Unsafe Events

Hypothesis 3 anticipated positive correlations between poor head movements' behavior at intersections as measured by IVC and hazardous driving-related events as obtained from CWS across the study phases. This analysis was aimed at examining whether the quality of the head movements' behavior at intersections was related to the number of risky events produced by the CWS. **Table 3** presents the correlations between head movements and the number of unsafe events. The CSW in the table represents the riskiness grade obtained by the total mean number of all unsafe events per hour.

Through examining the correlations between head movements' behavior and risky driving events, it seems that the statistically significant correlations are mainly found in phases 2 and 3. Moderate positive correlations were found for the improper head movements and negative correlations for the proper head movements at all CWS categories. The positive correlations indicate that as the percentage of the improper head movements increases, the number of unsafe events increases as well. Diversely, the significant negative correlations indicate that when the proper head movements rate increases, the number of unsafe events decreases.

Section 4: Driving Exposure (Total Number of Hours Driven and Distance Traveled)

Additional analyses were carried out to examine the differences between driving exposure of old adults and older drivers. The two age groups that participated in the study, old adults and older drivers, did not significantly differ in terms of gender and education variables as well as in driving history-related variables such as driving days per week and road accident history. The average age was different between the groups consistent with the experimental design. During the entire study period (6 months), the old-adults group drove a total of 9,568 km, while the older group drove a total of 7,071 km, however, the difference was not statistically significant. The final GLM model included one significant main effect of phase [$F_{(2,68)} = 5.957, p < 0.01$] and one significant second-order interaction of age group*phase [$F_{(2,68)} = 5.695, p < 0.01$]. **Figure 4** presents the average total number of kilometers driven by each group per phase (i.e., along a period of 2 months per phase).

Phases 1–3 represent the pre-intervention, intervention, and post-intervention phases, respectively. According to **Figure 4**, the older drivers appeared to travel a shorter distance than the old-adult drivers in phases 1 and 3. However, during the intervention phase, where the feedback from the CWS was activated, they traveled a more considerable distance than the old-adult drivers. *Post-hoc* pairwise contrast comparisons using the Bonferroni correction for multiple comparisons revealed that among the older drivers, phase 1 (EM = 1815.3 km, SE = 286.4) was not significantly different from phase 2 or 3 (EM = 3537.7 km, SE = 520.6; EM = 1717.6 km, SE = 385.8, for phase 2 and 3, respectively). Phase 2 was significantly different from phase 3, suggesting that drivers tended to drive many more kilometers during the intervention phase than the post-intervention phase. Among the old-adults, there was a significant difference between the total number of kilometers driven in phases 1 (EM = 3697.7 km, SE = 602) and 3 (EM = 2662.9 km, SE = 438.9). However, the intervention phase (EM = 3207 km, SE = 455.7) was not statistically significant in either phase 1 or phase 3.

To further characterize the participants' driving patterns, the study also examined the total number of actual driving hours, defined as the number of hours driving after reducing the total number of standing engine hours (when the engine is activated without driving). The study found that the old-adult group drove a total of ∼190 h, while the older drivers' group drove in total ∼156 h. The GLM model included one significant main effect of phase [$F_{(2,70)} = 7.428, p < 0.01$] and one significant second-order interaction between age group*phase [$F_{(2,70)} = 4.03$, $p < 0.05$]. **Figure 5** presents the total number of hours driven for each phase of the study.

Phases 1–3 represent the pre-intervention, intervention and post-intervention phases, respectively. According to **Figure 5**, the older drivers appeared to be driving fewer hours than the old-adults in phases 1 and 3. However, during the intervention phase, where feedback from the CWS was activated, they drove

TABLE 2 | Mean proportion (%) of improper head movements' cases at intersections in each of the study phases.

Phase	M-R	M-L	Ro-F	Ro-R	Ro-L	Tu-R	Tu-L	Tj-R	Tj-L	4W-F	4W-R	4W-L
1	30	30	5	1	9	2	0	0	1	22	0	0
2	28	25	3	4	12	3	3	0	0	21	1	0
3	31	26	4	1	13	3	2	0	0	20	0	0

R, Right; L, Left; F, Forward; M, Merge; Ro, Roundabout; Tu, Turn; Tj, T-junction; 4W, 4-ways intersection.
This table is based on 120 junctions per participants. The mean proportions are based on 40 participants in phase 1, 36 participants in phase 2, and 37 participants in phase 3. The sum of each row in this table equals 100%.

TABLE 3 | Correlations between head movements and unsafe driving.

Phase	Head movements	CWS[a]	HW[b]	uFCW[c]	LDW[d] (right)	LDW[d] (left)
1	Improper	−0.01	−0.03	−0.08	0.05	0.002
	Partially proper	−0.07	−0.04	0.05	−0.10	−0.03
	Proper	0.02	0.03	0.04	−0.01	0.01
2	Improper	0.06	0.00	−0.09	0.32	−0.07
	Partially proper	**0.48****	**0.38***	0.15	**0.36***	**0.56****
	Proper	−0.19	−0.11	0.06	**−0.41***	−0.08
3	Improper	0.20	−0.07	0.32	0.26	**0.38***
	Partially proper	0.29	0.26	−0.08	**0.38***	−0.06
	Proper	**−0.43***	−0.15	−0.26	**−0.54****	−0.31

$^{*}p < 0.05$ and $^{**}p < 0.01$.
[a]*Collision warning system (riskiness grade).*
[b]*Headway warning.*
[c]*Urban forward collision warning.*
[d]*Lane departure warning.*

more hours than the old-adults. *Post-hoc* pairwise contrast comparisons of the main effect revealed that the older group phase 1 ($EM = 39.1$ h, $SE = 5.3$) was not significantly different from phase 3 ($EM = 37.7$, $SE = 5.4$). However, phase 2 ($EM = 79.1$, $SE = 10.0$) was significantly different from phase 1 and 3, suggesting that drivers tended to drive many more hours during the intervention phase than the post-intervention phase. Among old-adults, there was a significant difference between the total number of hours driven in phases 1 ($M = 70.6$, $SE = 10.8$) and 3 ($M = 51.8$, $SE = 8.7$), with a smaller total number of hours driven in the post-intervention phase compared to the pre-intervention phase. There was no significant difference between the intervention phase ($M = 67.5$, $SE = 9.8$), and phases 1 and 3.

DISCUSSION

The present study was aimed at examining the influence of a collision warning system (CWS) visual or auditory feedback on older drivers' driving and head movements' behavior. The feedback-based intervention provided by CWS was found to be effective in improving study population's head movements' behavior at certain intersections such as T-junctions, turns and four-way intersections but was not effective at roundabout and merges, on the left direction.

The study findings showed that most of the improper head movements were associated with the types of intersections such as right or left merges, turning left at roundabouts and forward four-ways intersection. Possible explanations are divided between drivers' real road scanning difficulties or problems arising from excessively strict coding of improper head movements at these intersection types. At merges for example, it is possible that drivers were not scanning properly side roads that merge from the left because they knew they had the right-of-way forward and believed that the "responsibility" of road scanning was on the merging drivers.

Analyzing the IVC research videos showed that drivers rarely turn their heads sideways at road merges and therefore "gained" a high percentage of improper head movements. In the study of Lemonnier et al. (2020), they found most of the significant effects on oculomotor behaviors, but not on head orienting behaviors. In the same way, this may partly explain the high score of improper head movements in this specific type of intersection. Little is known about drivers' scanning behavior of merging side roads because most studies that measure drivers' road scanning behavior rarely focus on merges of side roads into the main road (Cheng et al., 2016). Moreover, the very few studies that investigated road scanning patterns at merging side roads focused on the scanning patterns of drivers who were merging into the main road rather than on drivers who were driving on the main road.

The current study also demonstrates that the proper head movement's behavior at roundabouts deteriorated from phase 1 and 2 to phase 3. We assume that the reason for the lack of

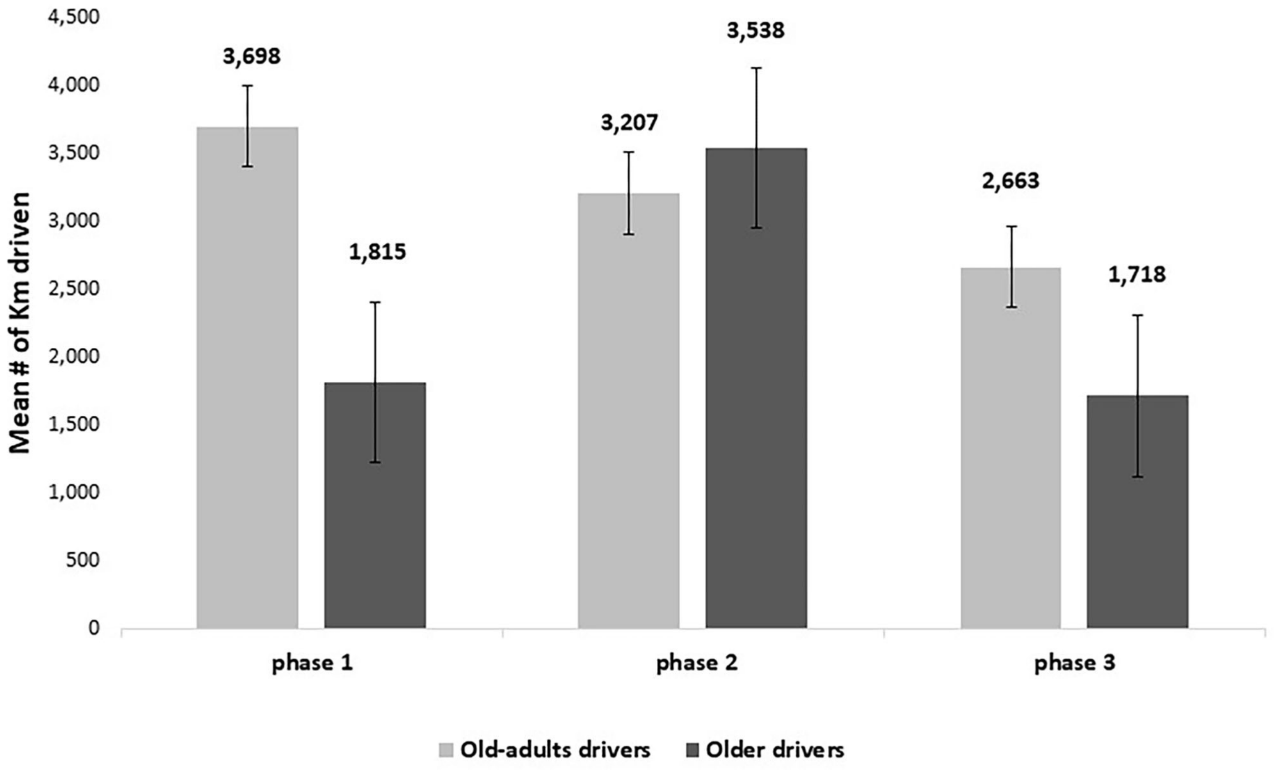


FIGURE 4 | Mean number of kilometers driven per age group.

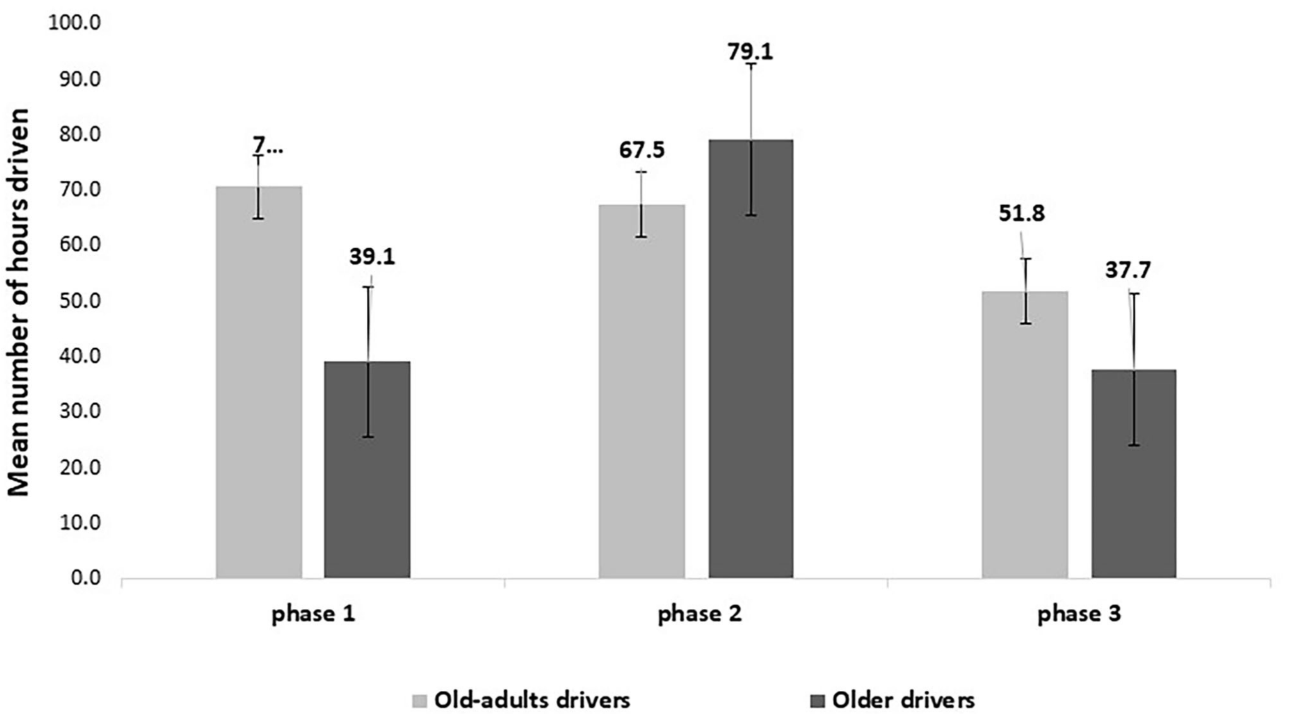


FIGURE 5 | Mean number of hours driven per age group.

effectiveness lies in the characteristics of the CWS feedback. The CWS feedback is not very effective at low speeds to begin with and most of the safety warning are activated only when driving speed is above 30 or 55 km/h (including FCW, HW, and LDW). Due to the fact that roundabouts are a known traffic calming countermeasure where lower speeds are generally observed (Zubaidi et al., 2020), the warnings were probably less active than the usual. Roundabouts are considered in the literature a good alternative to signalized or controlled intersections because of several advantages, such as the reduction of speed and fatal crashes and the enhancing of traffic capacity (Zubaidi et al., 2020). However, despite these advantages that roundabouts provide, it seems that crashes still occur. The recent study of Zubaidi et al. (2020) investigating the factors that contribute to injury severity sustained by drivers involved in crashes at roundabouts, found that one of the contributory factors was that vehicles did not wait to make a left turn. The current study findings are consistent with this research demonstrating a high percentage of improper head movements when turning left at the roundabouts. When turning left, the operational characteristics of roundabouts every time traffic approaches an entry point, should force drivers to slow down and regard the traffic entering the roundabout otherwise a possible conflict between road users could occur.

The study findings, showing a high percentage of improper head movements at four-ways intersections without a traffic light during the first phase, justify the global tendency to transform them to be controlled intersections or roundabouts (Zubaidi et al., 2020). The complex geometric structure of uncontrolled four-ways intersections can explain the high percentage of improper head movement behavior scores at phase 1 and the improvement throughout the phases as drivers needed the CWS intervention to assist their road scan behavior. According to the literature, four-ways intersections require extra visual scanning that includes looking in both directions and, in addition, performing a secondary scan by making another glance (Romoser and Fisher, 2009; Yamani et al., 2015; Samuel et al., 2016). In a review article by Samuel et al. (2016), they emphasize the importance and complexity of taking secondary glances at these types of intersections. Further research is needed to analyze the difficulties of older drivers in performing secondary glances at intersections using IVC and CWS to develop scanning behavior interventions, improve road scanning ability and thus reduce the risk of car accidents in older drivers (Romoser and Fisher, 2009; Samuel et al., 2016; Lococo et al., 2018).

Findings of the study revealed that the intervention phase improved the road scanning head movements of the study population at T-junctions and turns. According to Bao and Boyle (2009), older drivers do not utilize their full scanning range when compared to middle-aged drivers, and tend to check fewer areas before executing a maneuver through intersections, specifically during left and right turns. Similar to the results of simulator intervention studies (Romoser and Fisher, 2009; Pollatsek et al., 2012), the results of the current naturalistic study demonstrated evidence of improved older drivers' head search behavior at these specific intersections, considered dangerous intersections, with the advantages of using CWS technology.

In opposed to hypothesis 1, the intervention was not found to be effective in reducing drivers' involvement in hazardous driving-related events. The current study findings seem to contradict the conclusions of several studies that have investigated the effects of immediate feedback on improving driving performance by examining unsafe events measured by IVDR (Campbell et al., 2007; Toledo et al., 2008). Also, studies investigating the effectiveness of IVDR's immediate feedback in young drivers have shown significant improvement in speeding and additional non-significant improvement in acceleration and hard braking (Farmer et al., 2010; Farah et al., 2014). A retrospective study by Hickman et al. (2015) shows improvement in truck fleets following the introduction of the LDW system. These studies, however, investigated relatively young drivers (teens and young adults) compared to the current study older population. In fact, the current studies' results are in alliance with some of the recent studies concerning older drivers. Only one study was conducted on older Japanese taxi drivers and CWS. It was published only in the Mobileye's website[1] and this research concludes that once the CWS was installed, the number of accidents due to front-end collisions decreased by 85% down to zero and also significantly improved driving habits. A close examination of the Japanese study, reveals that they combined the feedback of collision avoidance systems with robust driver training by the fleet managers to improve driver's behavior further. It was found that although advanced safety features and automated vehicles offer great potential to improve road safety and the mobility of drivers, older drivers are skeptical about this technology and are least likely to rely on ADAS to improve their safety on the road (Robertson et al., 2017; Nielsen and Haustein, 2018).

Researches claim that the reason immediate negative feedback is not sufficiently effective over time is that drivers forget most of their near-accidents very quickly, and therefore it is worthwhile that drivers will be given retrospective feedback in addition to immediate feedback (Chapman et al., 2000). A retrospective feedback shows drivers a summary of their driving patterns to raise their awareness and motivation and make a real change in behavior for more extended time periods (Chapman et al., 2000). This argument is supported by findings from other domains such as gamification. Xie (2016), for example, have shown that when young drivers are driving in a driving simulator and receive meaningful feedback with game-design elements, it results in reduced distraction and motivates lasting behavioral changes. Two studies examined the long effects of retrospective feedback, which was conducted through tracking on a dedicated web site for drivers use, indicated an immediate improvement from the beginning of the study. In one study, the improvement was maintained throughout the 9 months of the driving monitoring (Musicant et al., 2007). In the second study, the improvement was maintained over a 4-month period, while in the fifth month, an increase in the number of undesired events above the baseline average was evidenced (Lotan and Toledo, 2006). Intervention studies from various fields have demonstrated the importance of active professional mediation for successful intervention and

[1]https://drive.google.com/file/d/1uNCPIwmqysQLen7Zidj5L4j93wYscwgk/view

achieving significant improvement (Lahav et al., 2008; Ratzon et al., 2009; Romoser and Fisher, 2009). The CWS is currently considered a "Stand-Alone" technology (meaning only the driver receives the technological feedback), and it is not enough to be used as an educational tool for older drivers. Therefore, additional mediation or follow-up is recommended for ensuring older drivers' safety.

In contrary to Hypothesis 2, the results of the study showed that there were no significant differences between old-adults and older drivers in all types of head movement's behavior—mainly proper and improper head movements throughout the three phases of the study. These findings are not in accordance with the literature. According to several studies, it has been shown that older drivers perform a reduced road scanning comparing to old-adults or experienced young drivers while driving at intersections (Bao and Boyle, 2009; Dukic and Broberg, 2012; Romoser et al., 2013). However, the present study compares older drivers to old-adults, aged 55–65, and maybe if the study had compared older drivers to younger drivers, the result would have been more similar to literature.

Visual distraction is considered among the major causes of road accidents (Khan and Lee, 2019). Therefore, the current study examined the relationship between head movements' behavior and risky driving. Analyzing the correlations between CWS unsafe events and head movements' behavior revealed that Hypothesis 3 was verified. It was found that mainly in phases 2 and 3, there were statistically significant moderate positive correlations between improper head movements' behavior at intersections as measured by IVC and hazardous driving-related events as obtained from CWS. Additionally, there were significant negative correlations with proper head movements' behavior and the riskiness grade (total mean number of unsafe events per hour) and LDW. Theses correlations proved that a connection between the two variables exists, and that might be a way to improve the older population's head movements' behavior at intersections by using current in-vehicle technology.

The connection between CWS unsafe events and head movements' behavior was not trivial because CWS technology monitors driving throughout the whole driving period and counts unsafe events in a wide range of driving situations activated from 30 km/h and above. However, the IVC road scan coding in the study focuses only on intersections where drivers tend to slow down naturally according to road infrastructure. The CWS scores consisted of the number of unsafe events recorded during the driving period in relation to the drivers' travel time and as noted in the method section. These scores included a variety of driving events (such as lane deviations and risk of rear-end collisions), under a variety of road or traffic situations (such as dense traffic in urban roads or inter-urban highways). Driving at intersections is just a small part of this variety of driving situations and conditions. Similar correlations between driving performance and visual distraction were reported by other researchers (Hirayama et al., 2013; Yang et al., 2015).

Additional results of the study showed a significant increase in both the number of kilometers traveled and the number of driving hours among the older drivers at the intervention phase, when the CWS was active, compared to the pre- and post-activation phases. In old-adults, however, there has been a marked decline in driving kilometers and the number of driving hours throughout the phases. Research data from phase 1 showing that older drivers have reduced driving kilometrage and driving hours are in line with known literature that older drivers conduct "self-regulation" on their driving to reduce the likelihood of accidents. Self-regulation is defined as making adjustments in driving to accommodate for changes in cognitive, sensory, and motor capabilities, such as shorter distance travel, avoiding rush hour travel, only driving in urban settings, avoiding nighttime travel, or stopping driving in general (Ekelman et al., 2009; Meng and Siren, 2012; Svancara et al., 2020). It seems that older drivers may benefit from the presence of CWS in their vehicles, as drivers with these systems appear willing to drive more and as a result maintain their mobility option to a greater extent than drivers without the technology. It is possible that driving with immediate feedback, alerting to possible dangers on the road, may have increased their confidence to drive even at times when they had previously tried to reduce driving. In the third phase, when the feedback "crutches" were removed, there was a return to the previous driving patterns as in the first phase. On the other hand, among old-adults that do not regulate their driving yet, the intervention phase has not caused a significant change, as seen in the older drivers.

These findings, along with the findings that older drivers tended to trust the CWS as manifested by the significantly higher number of hours they drove during the intervention phase present a complex picture. On the one hand, older drivers are more confident to drive with the help of a CWS which is a good thing as it supports their independent mobility. On the other hand, the CWS is not perfect and while it improves road scanning performance in some cases it impairs other cases. Thus, the inclusion of CWS and other in-vehicle technologies should be done with cautious making sure that the benefits are greater than the cost and does not compromise safety.

LIMITATIONS AND FURTHER RESEARCH

Since coding relied solely on head movements, if drivers depended only on their eye movements, UFOV, or wore sunglasses, an improper head movement was registered. However, the driver might have managed to road scan the area, but without considerable head movement. When considering the results, it is important to take into account that the analyses are based on a relatively small sample. Scoring on a 0–2 scale may have limited the sensitivity of this metric further. As was mentioned above, the present study compares older drivers to old-adults, aged 55–65, and maybe if the study had compared older drivers to younger drivers, the result would have been more similar to the literature. Further research is needed to better understand the right combination of immediate and retrospective feedback to maximize its impact on driving behavior. Further studies are warranted to examine the combined feedback on improving road scanning patterns of older drivers without compromising their safety.

CONCLUSIONS AND RECOMMENDATIONS

1. The immediate feedback of the CWS encouraged the older population to drive significantly more hours during the intervention phase and thus increased their involvement in everyday life. Being aware of the CWS's benefit can assist in making sure that older drivers continue enjoying a normal yet safe driving routine.
2. The results of this study showed that the immediate feedback provided by the CWS was partially effective in terms of participants' head movements at certain intersections (T-junctions, turns, and four-ways) but not in others (roundabout and merges). The feedback intervention was not effective in terms of reducing the number of CWS events across the study phases.
3. The CWS is currently considered a "Stand-Alone" technology (meaning only the driver receives the technological feedback), and it is not enough to be used as an educational tool for older drivers. Therefore, additional mediation or follow-up is recommended for ensuring older drivers' safety.
4. The combined feedback (immediate and retrospective) could allow drivers to receive an immediate response on unsafe driving events and "near accidents," and get a more detailed explanation of the meanings and consequences of their impaired driving behavior later on.
5. Combining the information from CWS's alerts and IVC with a telematics system will allow fleet managers/safety organizations/driving rehabilitators to analyze the driving patterns and habits of each older driver. With this hard data, they could support and train drivers who have bad driving habits as well as reward safe drivers' behavior. Retrospective feedback can be given to vehicle fleets by an authority, such as a fleet manager, through a feedback call/driving rating compared to other drivers in the company/providing reward to those who improve.
6. In regards to elderly drivers who drive in private vehicles, driving improvement can be rewarded through providing discounts when buying insurance (Pay as you drive), or mobile feedback messages to summarize a driving period.
7. Nevertheless, when introducing such aiding technologies to the vehicles of elderly drivers, caution should be exercised because these technologies have a complex-positive and negative effect that needs to be examined in further studies.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Ben-Gurion in the Negev, Israel Institutional Review Board (IRB). The patients/participants provided their written, informed consent to participate in this study. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was funded by a grant from the Research Fund on Insurance Matters affiliated with the Israel Insurance Association (R.A), a registered non-profit association, supporting research, and studies on traffic accidents.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.596278/full#supplementary-material

REFERENCES

Awadzi, K. D., Classen, S., Hall, A., Duncan, R. P., and Garvan, C. W. (2008). Predictors of injury among younger and older adults in fatal motor vehicle crashes. *Accid. Anal. Prevent.* 40, 1804–1810. doi: 10.1016/j.aap.2008.07.010

Bao, S., and Boyle, L. N. (2009). Age-related differences in visual scanning at median-divided highway intersections in rural areas. *Accid. Anal. Prevent.* 41, 146–152. doi: 10.1016/j.aap.2008.10.007

Bédard, M., Guyatt, G. H., Stones, M. J., and Hirdes, J. P. (2002). The independent contribution of driver, crash, and vehicle characteristics to driver fatalities. *Accid. Anal. Prevent.* 34, 717–727. doi: 10.1016/s0001-4575(01)00072-0

Bromberg, S., Oron-Gilad, T., Ronen, A., Borowsky, A., and Parmet, Y. (2012). The perception of pedestrians from the perspective of elderly experienced and experienced drivers. *Accid. Anal. Prevent.* 44, 48–55. doi: 10.1016/j.aap.2010.12.028

Campbell, J. L., Richard, C. M., Brown, J. L., and McCallum, M. (2007). *Crash Warning System Interfaces: Human Factors Insights and Lessons Learned.* NHTSA Technical Report, Administration, Washington, DC.

Carr, D. B., and Grover, P. (2020). The role of eye tracking technology in assessing older driver safety. *Geriatrics* 5:36. doi: 10.3390/geriatrics5020036

Caserta, R. J., and Abrams, L. (2007). The relevance of situation awareness in older adults' cognitive functioning: a review. *Eur. Rev. Aging Phys. Activity* 4, 3–13. doi: 10.1007/s11556-007-0018-x

Chapman, P., Roberts, K., and Underwood, G. (2000). "A study of the accidents and behaviours of company car drivers," in *Behavioural Research in Road Safety X*, ed G. B. Grayson (Crowthorne: Transport Research Laboratory).

Cheng, Y., Gao, L., Zhao, Y., and Du, F. (2016). Drivers' visual characteristics when merging onto or exiting an urban expressway. *PLoS ONE* 11:e0162298. doi: 10.1371/journal.pone.0162298

Classen, S. (2010). Special issue on older driver safety and community mobility. *Am. J. Occupat. Ther.* 64, 211–214. doi: 10.5014/ajot.64.2.211

Dellinger, A. M., Kresnow, M., White, D. D., and Sehgal, M. (2004). Risk to self versus risk to others: how do older drivers compare to others on the road? *Am. J. Prevent. Med.* 26, 217–221. doi: 10.1016/j.amepre.2003.10.021

Dukic, T., and Broberg, T. (2012). Older drivers' visual search behaviour at intersections. *Transport. Res. Part F* 15, 462–470. doi: 10.1016/j.trf.2011.10.001

Ekelman, B., Stav, W., Baker, P., O'Dell-Rossi, P., and Mitchell, S. (2009). "Chapter 14: community mobility," in *Functional Performance in Older Adults, 3th Edn,* eds B. Bonder, V. Dal Bello-Haas, and M. Wagner (Philadelphia, PA: F.A. Davis Company), 332–385.

Farah, H., Musicant, O., Shimshoni, Y., Toledo, T., Grimberg, E., Omer, H., and Lotan, T. (2014) Can providing feedback on driving behavior and training on parental vigilant care affect male teen drivers and their parents? *Accid. Anal. Prevent.* 69, 62–70. doi: 10.1016/j.aap.2013.11.005

Farmer, C. M., Kirley, B. B., and McCartt, A. T. (2010). Effects of in-vehicle monitoring on the driving behavior of teenagers. *J. Safety Res.* 41, 39–45. doi: 10.1016/j.jsr.2009.12.002

Fisher, D. L., Horrey, W. J., Caird, J. K., and Trick, L. M. (2016). *Handbook of Teen and Novice Drivers: Research, Practice, Policy, and Directions.* doi: 10.1201/9781315374123

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-6

Green, M. (2002). The invisible pedestrian. *OHandS Canada* 18, 44.

Hickman, J. S., Guo, F., Camden, M. C., Hanowski, R. J., Medina, A., and Mabry, J. E. (2015). Efficacy of roll stability control and lane departure warning systems using carrier-collected data. *J. Safety Res.* 52, 59–63. doi: 10.1016/j.jsr.2014.12.004

Hirayama, T., Mase, K., and Takeda, K. (2013). Analysis of temporal relationships between eye gaze and peripheral vehicle behavior for detecting driver distraction. *Int. J. Veh. Technol.* 2013:285927. doi: 10.1155/2013/285927

Hostetter, R. S., McGee, H. W., Crowley, K. W., Sequin, E. L., and Dauber, G.W. (1986). *Improved Perception-Reaction Time Information for Intersection Sight Distance*. Report No. FHWA/RD-87/015.

Hubele, N., and Kennedy, K. (2018). Forward collision warning system impact. *Traffic Injury Prevent.* 19, S78–S83. doi: 10.1080/15389588.2018.1490020

Karthaus, M., and Falkenstein, M. (2016). Functional changes and driving performance in older drivers: assessment and interventions. *Geriatrics* 1:12. doi: 10.3390/geriatrics1020012

Khan, M. Q., and Lee, S. (2019). Gaze and eye tracking: techniques and applications in ADAS. *Sensors* 19:5540. doi: 10.3390/s19245540

Lahav, O., Apter, A., and Ratzon, N. (2008). A comparison of the effects of directive visuomotor intervention versus nondirective supportive intervention in kindergarten and elementary-school children. *J. Neural Transm.* 115, 1231–1939. doi: 10.1007/s00702-008-0086-8

Land, M., and Tatler, B. (2009). "Driving," in *Looking and Acting: Vision and Eye Movements in Natural Behaviour,* eds M. Land and B. Tatler (New York, NY: Oxford University Press), 117–141.

Langford, J., and Koppel, S. (2006). Epidemiology of older driver crashes - Identifying older driver risk factors and exposure patterns. *Transport. Res. Part F* 9, 309–321. doi: 10.1016/j.trf.2006.03.005

Lees, M. N., and Lee, J. D. (2007). The influence of distraction and driving context on driver response to imperfect collision warning systems. *Ergonomics* 50, 1264–1286. doi: 10.1080/00140130701318749

Lemonnier, S., Désiré, L., Brémond, R., and Baccino, T. (2020). Drivers' visual attention: a field study at intersections. *Transport. Res. Part F* 69, 206–221. doi: 10.1016/j.trf.2020.01.012

Lococo, K. H., Staplin, L., and TransAnalytics, L. L. C. (2018). *Visual Scanning Training For Older Drivers: A Literature Review (No. DOT HS 812 514).* United States. National Highway Traffic Safety Administration.

Lotan, T., and Toledo, T. (2006). "An In-Vehicle Data Recorder for evaluation of driving behavior and safety," in *Preprints of the 85th Transportation Research Board Annual Meeting* (Washington, DC).

Maltz, M., and Shinar, D. (2004). Imperfect in-vehicle collision avoidance warning systems can aid drivers. *Hum. Factors* 46, 357–366. doi: 10.1518/hfes.46.2.357.37348

McGwin, G. Jr., Chapman, V., and Owsley, C. (2000). Visual risk factors for driving difficulty among older drivers. *Accid. Anal. Prevent.* 32, 735–744. doi: 10.1016/S0001-4575(99)00123-2

Meng, A., and Siren, A. (2012). Cognitive problems, self-rated changes in driving skills, driving-related discomfort and self-regulation of driving in old drivers. *Accid. Anal. Prevent.* 49, 322–329. doi: 10.1016/j.aap.2012.01.023

Metz, B., and Krueger, H. P. (2010). Measuring visual distraction in driving: the potential of head movement analysis. *IET Intell. Trans. Syst.* 4, 289–297. doi: 10.1049/iet-its.2009.0116

Musicant, O., Lotan, T., and Toledo, T. (2007). "Safety correlation and implication of an In-Vehicle Data Recorder on driver behavior," in *Preprints of the 86th Transportation Research Board Annual Meeting* (Washington, DC).

National Highway Traffic Safety Administration (2020). *Older Population: 2018 Data* (Traffic Safety Facts. Report No. DOT HS 812 928). National Highway Traffic Safety Administration, National Center for Statistics and Analysis.

Nielsen, T. A. S., and Haustein, S. (2018). On sceptics and enthusiasts: what are the expectations towards self-driving cars? *Trans. Policy* 66, 49–55. doi: 10.1016/j.tranpol.2018.03.004

Park, S. J., Subramaniyam, M., Kim, S. E., Hong, S., Lee, J. H., and Jo, C. M. (2017). "Older driver's physiological response under risky driving conditions— overtaking, unprotected left turn," in *Proceedings of the AHFE 2016 International Conference on Digital Human Modeling and Simulation, Vol. 481*, ed V. Duffy (Cham: Springer International Publishing), 107–114. doi: 10.1007/978-3-319-41627-4_11

Pollatsek, A., Romoser, M. R. E., and Fisher, D. L. (2012). Identifying and remediating failures of selective attention in older drivers. *Curr. Direct. Psychol. Sci.* 21, 3–7. doi: 10.1177/0963721411429459

Pomidor, A. (2015). *Clinician's Guide to Assessing and Counseling Older Drivers.* Washington, DC: National Highway Traffic Safety Administration.

Ratzon, N., Lahav, O., Cohen- Hamsi, S., Metzger,. Y., Efraim, D., and Bart, O. (2009). Comparing different short-term service delivery methods of visual-motor treatment for first grade students in mainstream schools. *Res. Dev. Disabil.* 30, 1168–1176. doi: 10.1016/j.ridd.2009.03.008

Robertson, R. D., Meister, S. R., Vanlaar, W. G. M., and Mainegra Hing, M. (2017). Automated vehicles and behavioural adaptation in Canada. *Transport. Res. Part A* 104, 50–57. doi: 10.1016/j.tra.2017.08.005

Romoser, M. R., and Fisher, D. L. (2009). The effect of active versus passive training strategies on improving older drivers' scanning in intersections. *Hum. Factors* 51, 652–668. doi: 10.1177/0018720809352654

Romoser, M. R., Pollatsek, A., Fisher, D. L., and Williams, C. C. (2013). Comparing the glance patterns of older versus younger experienced drivers: Scanning for hazards while approaching and entering the intersection. *Transport. Res. Part F* 16, 104–116. doi: 10.1016/j.trf.2012.08.004

Samuel, S., Yamani, Y., and Fisher, D. L. (2016). Large reductions are possible in older driver crashes at intersections. *Clin. Exper. Optometry* 99, 419–424. doi: 10.1111/cxo.12443

Strayer, D. L., Drews, F. A., and Johnston, W. A. (2003). Cell phone-induced failures of visual attention during simulated driving. *J. Exp. Psychol.* 9, 23–32. doi: 10.1037/1076-898X.9.1.23

Svancara, A. M., Villavicencio, L., Kelley-Baker, T., Horrey, W. J., Molnar, L. J., Eby, D. W., et al. (2020). The relationship between in-vehicle technologies and self-regulation among older drivers. *Geriatrics* 5:23. doi: 10.3390/geriatrics5020023

Toledo, T., Musicant, O., and Lotan, T. (2008). In-vehicle data recorders for monitoring and feedback on drivers' behavior. *Transport. Res. Part C* 16, 320–331. doi: 10.1016/j.trc.2008.01.001

Van Houten, R., and Retting, R. A. (2001). increasing motorist compliance and caution at stop signs. *J. Appl. Behav. Anal.* 34, 185–193. doi: 10.1901/jaba.2001.34-185

Vespa, J., Armstrong, D. M., and Medina, L. (2018). *Demographic Turning Points for the United States: Population Projections for 2020 to 2060*. Washington, DC: US Department of Commerce, Economics and Statistics Administration, US Census Bureau.

Wang, Z., Zlatanova, S., and Steenbruggen, J. (2016). "Traffic incident management in the presence of hazards," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Iv-4/w1*, (Copernicus Publications) 45–49.

Xie, J. Y. (2016). "Gaming to safety: The design and evaluation of feedback gamification for mitigating driver distraction," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 60, 1884–1888. doi: 10.1177/1541931213601429

Yamani, Y., Samuel, S., Gerardino, L. R., Zafian, T., and Fisher, D. L. (2015). "Navigating intersections: examining age-related differences in visual scanning on a driving simulator," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Vol. 59* (Los Angeles, CA: Sage CA, SAGE Publications.ISO 690), 1636–1640.

Yang, Y., Sun, H., Liu, T., Huang, G. B., and Sourina, O. (2015). "Driver workload detection in on-road driving environment using machine learning," in *Proceedings of ELM-2014, Vol. 2* (Berlin; Heidelberg: Springer), 389–398.

Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., et al. (1982). Development and validation of a geriatric depression screening scale: a preliminary report. *J. Psychiatr. Res.* 17, 37–49. doi: 10.1016/0022-3956(82)90033-4

Zubaidi, H. A., Anderson, J. C., and Hernandez, S. (2020). Understanding roundabout safety through the application of advanced econometric techniques. *Int. J. Transport. Sci. Technol.* 9, 309–321. doi: 10.1016/j.ijtst.2020.03.001

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

BRIEF RESEARCH REPORT
published: 25 March 2021
doi: 10.3389/fpsyg.2021.619104

Check for updates

Skin Conductance Responses of Learner and Licensed Drivers During a Hazard Perception Task

Theresa J. Chirles[1], Johnathon P. Ehsani[1], Neale Kinnear[2] and Karen E. Seymour[3,4,5]*

[1]Center for Injury Research and Prevention, Bloomberg School of Public Health, Health Policy and Management Department, Johns Hopkins University, Baltimore, MD, United States, [2]Transportation Research Laboratory, Wokingham, United Kingdom, [3]School of Medicine, Department of Psychiatry, Johns Hopkins University, Baltimore, MD, United States, [4]Bloomberg School of Public Health, Department of Mental Health, Johns Hopkins University, Baltimore, MD, United States, [5]National Institutes of Health, Center for Scientific Review, Bethesda, MD, United States

OPEN ACCESS

Edited by:
Oren Musicant,
Ariel University, Israel

Reviewed by:
William Horrey,
AAA Foundation for Traffic Safety, United States
Domen Novak,
University of Wyoming, United States
Mariaelena Tagliabue,
University of Padua, Italy

****Correspondence:***
Theresa J. Chirles
tchirle1@jhmi.edu

Specialty section:
This article was submitted to Performance Science, a section of the journal Frontiers in Psychology

Received: *19 October 2020*
Accepted: *22 February 2021*
Published: *25 March 2021*

Citation:
Chirles TJ, Ehsani JP, Kinnear N and Seymour KE (2021) Skin Conductance Responses of Learner and Licensed Drivers During a Hazard Perception Task. Front. Psychol. 12:619104. doi: 10.3389/fpsyg.2021.619104

Background: While advanced driver assistance technologies have the potential to increase safety, there is concern that driver inattention resulting from overreliance on these features may result in crashes. Driver monitoring technologies to assess a driver's state may be one solution. The purpose of this study was to replicate and extend the research on physiological responses to common driving hazards and examine how these may differ based on driving experience.

Methods: Learner and Licensed drivers viewed a Driving Hazard Perception Task while electrodermal activity (EDA) was measured. The task presented 30 Event (hazard develops) and 30 Non-Event (routine driving) videos. A skin conductance response (SCR) score was calculated for each participant based on the percentage of videos that elicited an SCR.

Results: Analysis of the SCR score during Event videos revealed a medium effect (d = 0.61) of group differences, whereby Licensed drivers were more likely to have an SCR than Learner drivers. Interaction effects revealed Licensed drivers were more likely to have an SCR earlier in the Event videos compared to the end, and the Learner drivers were more likely to have an SCR earlier in the Non-Event videos compared to the end.

Conclusion: Our results support the viability of using SCR during driving videos as a marker of hazard anticipation differing based on experience. The interaction effects may illustrate situational awareness in licensed drivers and deficiencies in sustained vigilance among learner drivers. The findings demand further examination if physiological measures are to be validated as a tool to inform driver potential performance in an increasingly automated driving environment.

Keywords: electrodermal activity, autonomous vehicles, driving experience, hazard perception, young drivers

INTRODUCTION

Advanced driver assistance systems (ADAS) have the potential to drastically reduce vehicle crash injury and death but may be accompanied with possible setbacks. Recent experience demonstrates that overreliance on this technology poses a separate set of risks where drivers may be unable to regain vehicle control in situations of technology failure (Krompier, 2017;

Vogelpohl et al., 2019). One approach to addressing this problem is to augment safety by monitoring driver state (e.g., drowsiness, workload, and levels of vigilance) by using physiological measurements (Balters and Steinert, 2017; Lohani et al., 2019), but a model to understand the complexity of the relationship between physiological measures, individual driver cognitive state, and the implications for driving behavior and performance is far from complete (Balters and Steinert, 2017).

One psychophysiological measure of autonomic arousal utilized to monitor driver state is electrodermal activity (EDA). EDA is a measure of neuronally mediated autonomic changes in the electrical properties of the skin, and has been shown to be a sensitive index of sympathetic nervous system activity (Braithwaite et al., 2013; Dawson et al., 2017). Tonic skin conductance levels and phasic skin conductance responses (SCR), are elements of EDA that have long been used in the driving literature to measure workload, risk of accident (Hulbert, 1957; Taylor, 1964; Helander, 1978), as well as levels of stress and tension (Michaels, 1960; Healey and Picard, 2005). A recent study by Darzi et al. (2018) found decreased tonic skin conductance levels to be indicative of sleep deprivation. However, if driver state monitoring is to become a successful intervention to facilitate the safe interplay of driver assistance technology and driver manual takeover, psychophysiological monitoring models must not only incorporate cognitive states but also how individual responses may vary based on experience (Collet and Musicant, 2019) and the acquisition of critical driving skills.

For example, a critical skill that develops with driving experience is hazard perception (Quimby et al., 1986; McKenna and Crick, 1991; Horswill and McKenna, 2004). Hazard perception is the learned ability to detect, predict, recognize, and respond to developing hazards (Horswill and McKenna, 2004; Wetton et al., 2011; Crundall et al., 2012; Crundall, 2016) and has been associated with crash risk (McKenna and Crick, 1991). Kinnear et al. (2013) found that compared to novice drivers, experienced drivers were twice as likely to demonstrate an SCR when watching videos containing a driving hazard. The videos used in this study were validated to distinguish between novice and experienced drivers as part of the development of the United Kingdom hazard perception test. The difference between novices and experienced drivers was in the period leading up to the hazardous event, termed the "anticipatory period." Subsequent hazard perception and SCR research have found similar results (Tagliabue and Sarlo, 2015; Barnard and Chapman, 2016; Tagliabue et al., 2017) but have been conducted outside the United States.

A potential reason for these differing autonomic responses to driving hazards between novice and experienced drivers emerges from literature suggesting the role of somatic experience on decision-making. Specifically, evidence suggest that this learning not only occurs from explicit knowledge of reward/punishment schedules, but also from affect-based somatic signals (i.e., pulse rate blood flow, pupil response, etc.) experienced by the driver (Damasio, 1994; Phelps et al., 2014; Petracca, 2020). Known as the somatic marker hypothesis (Damasio, 1994), this theory has potential implications for novice vs. experienced drivers suggesting that prior positive or negative experience results in the formation of a gut feeling or "somatic marker" (i.e., a physiological response) which in turn biases the options available for decision-making when encountering a similar situation in the future. This "feeling-based" system for decision-making complements and operates in parallel to the rational decision-making process, which if it were operating in isolation would take too long to reach complex decisions. However, decision speed and accuracy could be ecologically viable if facilitated using feedback from the autonomic and the somatic nervous systems *via* the emotion circuitry in the brain (Bechara and Damasio, 2005). Taken together, these biologically based decision-making theories have relevance for driver assessment of hazards, a decision-making process which needs to occur rapidly. Drivers who have progressed past the novice (or learner) stage would have a larger library of experience to draw from, allowing their feeling-based appraisal to identify potential risks earlier, and bias a behavioral response to anticipate and avoid the impending hazard.

The purpose of this study was to replicate and extend research investigating measures of autonomic arousal (i.e., EDA) during the viewing of driving hazards for young drivers differing in experience levels in the United States context. As the stimuli developed for previous studies were from the United Kingdom, they could not be readily used in the United States context due to the differences in the driver position in the vehicle and the direction of travel lanes. Thus, we developed a novel Driving Hazard Perception Task (Ehsani et al., 2020) and measured SCR during videos where a hazard occurred (Event), and videos of routine driving (Non-Event). Videos were extracted from real-world driving captured as part of a large-scale naturalistic driving study. While driving simulators more closely mimic the on-road driving task, the focus here was autonomic responses to developing hazards, rather than driving task performance. Videos including naturally occurring cues therefore provided the stimuli necessary for this study. The use of videos for this purpose has been demonstrated previously (Kinnear et al., 2013) and is commonly used for hazard perception testing. We hypothesized that more experienced drivers (Licensed) would have a greater likelihood of SCR than Learner drivers during Event videos, and both Learner and Licensed drivers would have a greater likelihood of SCR during the Event videos compared to the Non-Event videos. By replicating and extending the literature supporting the finding of heightened autonomic arousal during hazard anticipation induced by experiential learning, we would be providing valuable information to the developers of driver state monitoring systems.

MATERIALS AND METHODS

Participants

Participants were recruited through flyer, email, website announcements, and in person from the Baltimore metro region. To be included in the study, participants needed to be between the ages of 16–20 years and have a valid driver's license or learner's permit and speak English fluently. To be included in the Learner group, drivers held a valid learner's permit and had driven less than 1,000 miles as assessed by self-reported mileage. To be included in the Licensed group, participants had a valid non-commercial driver's license for a minimum of 2 years and had driven more than 3,000 miles in the past 12 months as assessed by self-reported annual mileage. Exclusionary criteria included: (1) a history of neurologic disorder (e.g., epilepsy, cerebral palsy, traumatic brain injury, and Tourette syndrome), (2) a history of visual impairment, (3) the inability to read English fluently, and (4) the presence of psychiatric illness or neurodevelopmental disorders assessed *via* The Mini International Neuropsychiatric Interview for Children and Adolescents (MINI-KID). All study procedures were approved by the Johns Hopkins Medicine Institutional Review Board.

Study Visit

Participants came to a single study session after passing an initial phone screening interview. Participants were introduced to the physiological recording equipment after informed consent and eligibility for the study confirmed with the MINI-KID. After the physiological recording equipment was placed on the participant and the quality of the data collection was verified, a 2-min baseline measurement was collected (the participant sat quietly and was task free) before the commencement of The Driving Hazard Perception Task that included 60 30-s videos [30 Event videos (hazard develops) and 30 Non-Event videos (routine driving)]. The development and details about the task may be found in Ehsani et al. (2020) and in the **Supplementary Material**. After the task, the recording equipment was removed, and participants completed demographics and medication history questionnaires. The participants were provided with a $50 gift card as compensation.

Calculation of SCR Score

Measurement windows in the Event videos were defined from the first frame the hazard appeared on the screen to 3 s after the driver was required to perform an evasive action (see **Figure 1**). The window included the evasive action due to the delayed response of SCR (Braithwaite et al., 2013). These windows approximated the *anticipatory period* from the study of Kinnear et al. (2013). In Non-Event videos, non-hazardous occurrences were randomly selected from general driving clips with timing that corresponded to the event videos. Descriptions of the frames chosen to define the *measurement windows* are in **Supplementary Table S1** (**Supplementary Material**). To avoid learning effects, the *Onset Time* of these *measurement windows* were staggered so that in the Event and Non-Event videos, *Early Onset videos* had the window at the beginning, *Middle Onset videos* had the window in the middle, and the *Late Onset videos* had the window toward the end of the video.

For an SCR to be included in the data, the phasic component increase of the EDA signal was required to be equal to or exceed 0.03 μS (Braithwaite et al., 2013), and the waveform onset was initiated within *measurement window*, and the peak was achieved within 10 s of the waveform onset. While Kinnear et al. (2013) used a 0.05 μS threshold, we opted to use the current acceptable threshold due to improvements in technology. These analyses were performed using *AcqKnowledge* Biopac Basic Scripting Software, 5.0, and the results were visually inspected for quality control. Participants were monitored during the data acquisition for behavior that might induce an SCR artifact (e.g., yawning, deep breaths, and body movement). There were no SCRs attributed to these artifacts during the *measurement windows*.

Repeating the method in Kinnear et al. (2013), we calculated an SCR score. This dependent measure reflects non-responses as well as responses. To calculate an *SCR Score*, the number of Event and Non-Event video clips where the participant exhibited an SCR response during the *measurement window* were summed. To do this, Event periods were coded (using Matlab) as a 0 or 1, depending if an SCR occurred within the time frame. On rare occasions, a participant demonstrated two SCRs during the time frame, but it was still coded as "1." The following equation was used to calculate each participant's SCR score for Event and Non-Event videos:

$$SCR\ Score(\%) = \frac{no.\,of\ \ clips\ with\ SCR}{total\ no.\,of\ \ clips} \times 100$$

This score represented the proportion of clips within each video type that elicited an SCR.

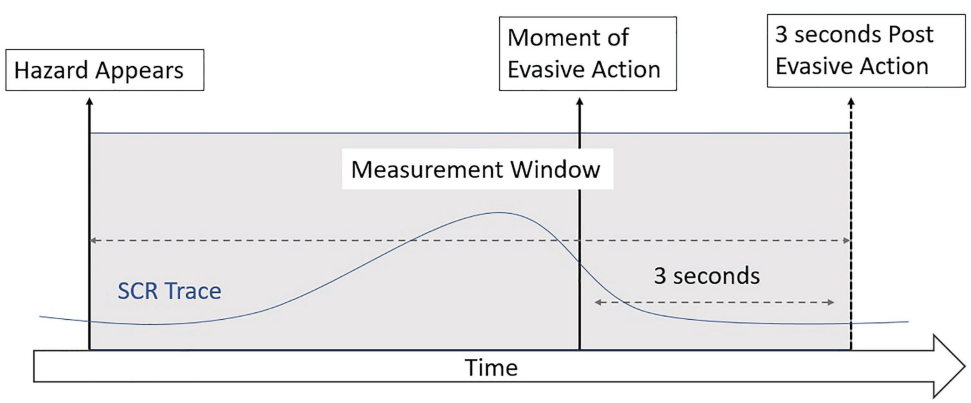


FIGURE 1 | *Measurement windows* for event videos.

Data Analysis

All analyses were conducted using IBM Statistics SPSS v26. To compare Learner and Licensed drivers on SCR score, a mixed model ANOVA was used with Group as the between subject variable and Video Type as the within subject factor. We reported main effects and interactions from this analysis. *Post hoc* pairwise comparisons were examined with Bonferroni corrections applied for multiple comparisons. Due to our small sample, when values of p approached significance, we calculated effect sizes. Effect sizes were assessed using Cohen's d with small, medium, and large effect sizes as Cohen's d 0.3–0.5, 0.5–0.8, and $\geq$0.8, respectively (Cohen, 1992).

Data were also examined for influences due to age, sex, and the task design effects of trial order and *Onset Time*.

RESULTS

Participant Demographics

Participants included 41 drivers aged 16–20-years old. Three participants were excluded from analyses after reporting regular use of medications known to blunt SCR response. For the remaining 38 participants, 20 were Learner drivers who reported holding a United States learner's permit for 1.29 ± 1.05 years (mean ± *SD*) and driving less than 900 miles of self-reported miles (mean ± *SD*: 293 ± 306). The Licensed drivers (n = 18) had a United States driver's license (n = 16) or International license (n = 2) for at least 2 years (mean ± *SD*: 2.74 ± 0.62 years) and had more than 3,000 self-reported driving miles in the past 12 months (mean ± *SD*: 4,766 ± 1,708). All participants had normal or corrected to normal vision and did not report colorblindness.

The male:female ratio for the groups was: Licensed 11:7; Learner 6:14; [$\chi^2(1) = 3.71$, $p = 0.054$]. Licensed drivers were slightly older than Learner drivers, $F(1,36) = 5.79$, $p = 0.021$. We had no missing data. Licensed and Learner drivers differed significantly in number of miles driven, $F(1,36) = 132.78$, $p < 0.001$ such that Licensed drivers reported a higher number of miles driven compared to Learner drivers. Additionally, Licensed drivers reported having driven for more years than Learner driver, $F(1,36) = 26.00$, $p < 0.001$. Participant demographics are presented in **Table 1**.

Analysis of SCR Score

Preliminary analyses revealed no significant main effects of age or sex, and no interaction with other variables; thus, we omitted sex from the model, but we included age as a covariate as there were group differences. There was no apparent video order effect, but a mixed model ANOVA (Onset Time: Early, Mid, and Late) × 2 (Event) × 2 (Group) revealed a significant three-way interaction [$F(1,36) = 5.669$, $p = 0.023$, $r^2 = 0.136$, observed power 0.639] indicating group had a different effect on SCR score depending on Onset Time and Event Type.

Do Learner and Licensed Drivers Experience Differences in Psychophysiological Reactions to a Driving Hazard?

The results of a 2 × 2 mixed model ANOVA, with Video Type as the within subject factor (Event or Non-Event) and Group as the between subject factor (Learner or Licensed), revealed a Group effect that approached significance [$F(1,36) = 3.623$, $p = 0.065$, $r^2 = 0.094$, observed power 0.457, $d = 0.64$]; the medium effect size indicating Licensed drivers were more likely to have an SCR response than Learner drivers ($\text{mean}_{\text{Lic}} \pm SD = 32.06 \pm 16.18$; $\text{mean}_{\text{Learn}} \pm SD = 21.71 \pm 16.12$). There was no significant interaction effect [$F(1,36) = 0.474$, $p = 0.496$, $r^2 = 0.013$, observed power 0.103], suggesting that the group effect was consistent across Event and Non-Event videos.

Pairwise comparisons further probing the Group effect approached significance for the Event Videos [$F(1,37) = 3.524$, $p = 0.069$, $d = 0.61$], the medium effect indicating Licensed drivers more likely to have an SCR response than Learner drivers in the Event videos ($\text{mean}_{\text{LicEvent}} \pm SD = 34.59 \pm 17.96$; $\text{mean}_{\text{LearnEvent}} \pm SD = 24.24 \pm 16.02$; see **Figure 2**). The difference between groups did reach significance in the *Early Onset* Event Videos, [$F(1,37) = 6.259$, $p = 0.017$; $\text{mean}_{\text{LearnEvent}} \pm SD = 22.22 \pm 18.73$; $\text{mean}_{\text{LicEvent}} \pm SD = 41.36 \pm 27.96$] indicating a greater likelihood of an SCR in the Licensed drivers compared to Learner drivers if the hazard appeared early (**Figure 3**, letter a).

There was no Group effect for the Non-Event Videos [$F(1,37) = 1.691$, $p = 0.202$, $d = 0.42$], yet in the *Late Onset* Non-Event Videos, there was a significant difference between groups [$F(1,37) = 05.393$, $p = 0.026$; $\text{mean}_{\text{LearnNonEvent}} \pm SD = 15.00 \pm 12.10$; $\text{mean}_{\text{LicNonEvent}} \pm SD = 25.46 \pm 15.60$]

TABLE 1 | Demographics and driving history.

	Learner (n = 20)		Licensed (n = 18)		Group comparisons	
	Mean(*SD*)	Range	Mean(*SD*)	Range	*F*-statistic	*p* value
Age (years)	18.37(1.78)	16.12–20.87	19.47(0.79)	18.04–20.78	5.79	0.021
Sex*	6M/14F		11M/7F		3.71	0.054
Miles driven**	293 (306)	0–900	4,766(1708)	3,000–9,240	132.78	<0.001
Time driving***	1.29(1.05)	0.06-3.02	2.74(0.62)	2.02–4.17	26.00	<0.001

*Pearson Chi-Square value reported.
**Miles driven: self-reported total mileage for learner drivers; self-reported mileage in the past 12 months for licensed drivers.
***Time driving: number of years from permit (for learner) or license (for licensed) issued date and time of study assessment.

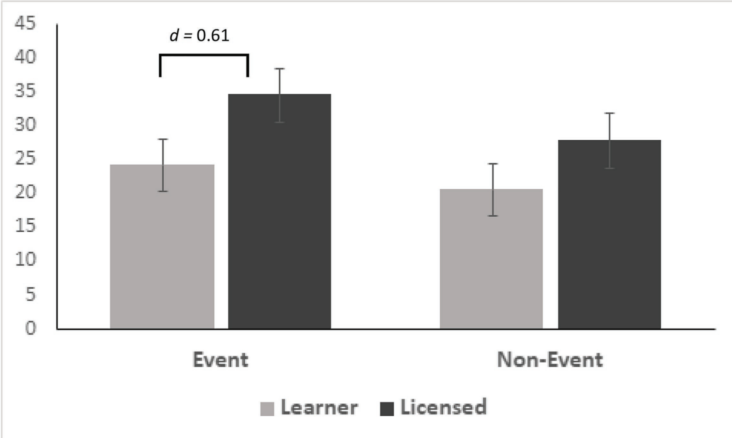


FIGURE 2 | Mean skin conductance response (SCR) score by Group and Video Type with SE Bars.

indicating the Licensed drivers were more likely to have an SCR than Learner drivers during routine driving if the *measurement window* was toward the end of the video (**Figure 3**, letter b).

There was a significant difference between *Early* and *Late Onset* Event Videos in the Licensed group ($\text{mean}_{\text{LicEarly}} \pm SD = 41.36 \pm 27.96$; $\text{mean}_{\text{LicLate}} \pm SD = 27.02 \pm 16.39$; $p = 0.024$), indicating that the Licensed drivers were more likely to have an SCR response if the hazard developed earlier in the video compared to the end of the video (**Figure 3**, letter c). However, for the Learner group, there was a significant difference between *Early and Late Onset* Non-Event videos ($\text{mean}_{\text{LearnEarly}} \pm SD = 26.50 \pm 20.33$; $\text{mean}_{\text{LearnLate}} \pm SD = 15.00 \pm 12.10$; $p = 0.018$), indicating the Learner drivers were more likely to have an SCR response earlier in the Non-Event videos compared to the end of the Non-Event video (**Figure 3**, letter d).

Are Drivers More Likely to Show an SCR in Event Videos Compared to Non-Event Videos?

There was not a significant effect of Video Type [$F(1,35) = 0.109$, $p = 0.744$, $r^2 = 0.003$, observed power 0.062, $d = 0.29$], indicating a similar likelihood across all videos to have an SCR response ($\text{mean}_{\text{Event}} \pm SD = 29.14 \pm 17.54$; $\text{mean}_{\text{Non-Event}} \pm SD = 24.08 \pm 17.20$).

Yet, in *Early Onset videos*, there was a significant difference between video type in the Licensed group [$F(1,17) = 5.776$, $p = 0.028$; $\text{mean}_{\text{LicEvent}} \pm SD = 41.36 \pm 27.96$; $\text{mean}_{\text{LicNonEvent}} \pm SD = 28.33 \pm 22.82$] indicating Licensed drivers were more likely to have an SCR when a hazard developed compared to routine driving (**Figure 3**, letter e) when the *measurement window* was at the beginning of the video.

In the *Late Onset* videos, there was a significant difference in the Learner group between video types [$F(1,19) = 5.794$, $p = 0.026$, $r^2 = 0.234$, observed power = 0.627; $\text{mean}_{\text{LearnEvent}} \pm SD = 24.55 \pm 19.81$; $\text{mean}_{\text{LearnNonEvent}} \pm SD = 15.00 \pm 12.10$] indicating Learner driver were more likely to have an SCR when a hazard developed than compared to routine driving (**Figure 3**, letter f) when the *measurement window* was at the end of the video.

DISCUSSION

The purpose of this study was to examine differences in the autonomic responses (SCR) of Learner and Licensed drivers in response to video stimuli in the United States, replicating past studies performed in the United Kingdom. These psychophysiological differences will provide insights into how experience should be incorporated in the algorithms monitoring driver cognitive state in this increasingly automated driving environment. Overall, Learner drivers did demonstrate fewer SCRs than the Licensed drivers during The Driving Hazard Perception Task, and Event videos appeared to discriminate between Learner and Licensed drivers. The medium effect size suggests the relationship between driving group and SCR score in event videos was meaningful ($d = 0.61$) but the sample was underpowered to reach statistical significance. This finding is consistent with previous research examining differences in autonomic arousal between novice and experienced drivers (Kinnear et al., 2013) in a controlled setting. The addition of Non-Event videos to the task was novel, and Licensed drivers were more likely to exhibit an SCR in and Event video than Non-Event if the hazard developed early. Conversely, Learner drivers were more likely to demonstrate an SCR in the Event videos if the hazard developed late.

The design of this study differs from the experiment that it sought to replicate in one critical aspect, and this may explain underpowered results compared to the clear picture of greater autonomic responses in experienced drivers in Kinnear et al. (2013). The stimuli used for this study were derived from naturalistic driving dashcam footage as opposed to the professionally filmed hazard perception clips that were

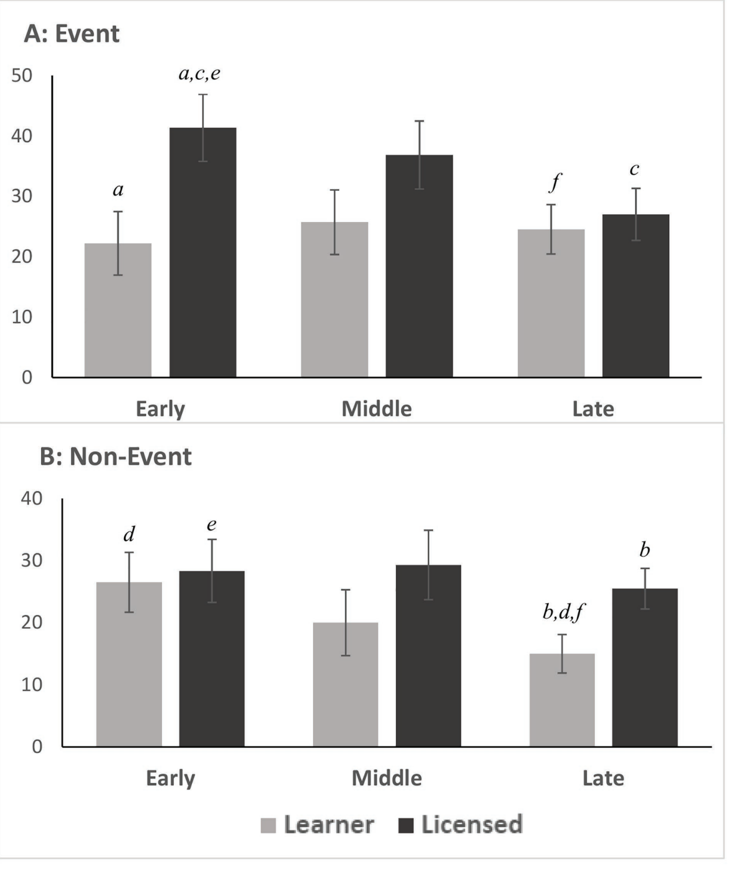


FIGURE 3 | Skin conductance response score by Onset Time and Group. SE bars. Matched letters indicate statistically significant differences in mean score. Panel **(A)**: in the Event videos, **(a)** Licensed drivers had a greater SCR score in the *Early Onset* compared to Learner drivers; **(c)** Licensed drivers had a greater SCR score in *Early Onset* compared to *Late Onset*. Panel **(B)**: in the Non-Event videos, **(b)** Licensed drivers had a greater SCR score in the *Late Onset* compared to Learner drivers; **(d)** Learner drivers had a greater SCR score in the *Early Onset* compared to *Late Onset*. Across Panels: **(e)** Licensed drivers had a greater SCR score in the Event *Early Onset* compared to Non-Event *Early Onset*; **(f)** Learner drivers had a greater SCR score in the Event *Late Onset* compared to the Non-Event *Late Onset*.

used in the United Kingdom study. The United Kingdom clips were developed during the design of the official hazard perception test and went through a detailed validation process (Grayson and Sexton, 2002). As such, they included validated examples of developing hazards and included defined timing windows, allowing differentiation between "anticipatory" (or precursory) and "event" areas in the time window of the defined hazard. Differentiation between the anticipatory and event periods could not be as clearly made in this study because the experimental stimuli lacked an extended build up period. This lack of definition between the anticipatory and the event stages of the hazards may have masked differences between novice and more experienced drivers and points to the importance of the experimental stimuli in measuring hazard perception.

This study also found an influence of *Onset Time* that had not been previously examined. This may also be an experiment artifact related to the specific stimuli, but the findings do suggest some important factors as it relates to autonomous driving and driver monitoring. Licensed drivers had a pattern of *decreased* likelihood of producing an SCR when the hazard developed later in the video, yet the Learner drivers maintained similar responses regardless of the timing of the hazard. This finding provides evidence that may be indicative of situational awareness as it relates to hazard *prediction*. As more experienced drivers have more time to observe the environmental and behavioral stimuli in a developing scene, they are better able to predict possible behaviors, and thus less likely to elicit an SCR when a predicted hazard occurs. Learners, on the other hand, were equally "surprised" when the hazard developed regardless of the time spent observing the situation. This theory coincides with current work indicating hazard prediction is the subcomponent of hazard perception that differentiates experienced and novice drivers (Crundall, 2016; Ventsislavova et al., 2019). The similar pattern of *decreased* likelihood of producing an SCR later in the video was observed in the Learner drivers in the Non-Event videos. The Licensed drivers, in contrast, had a consistent likelihood of an autonomic response regardless of the timing of the *measurement window*.

We have labeled these videos as Non-Event, but the routine driving captured by the dashcam will inherently include situational cues to which more experienced drivers may respond. While the Licensed drivers consistently attend to these potential hazard cues throughout an "uneventful" video clip, the patterns observed in the Learner drivers may be physiological evidence of previously identified novice deficiencies in lack of awareness and sustained vigilance.

This interpretation of differences in situational awareness and sustained vigilance is presented with caution for there are other possible interpretations. While there was an overall lower reactivity of Learner drivers, it may be these drivers needed a longer period to discriminate between Event and Non-Event videos, and thus video type differences were only seen in the *Late Onset* videos for this group due to sustained vigilance. In contrast, the Licensed driver decreased reactivity in Event videos as time progresses may be indicative of decreased sensitivity rather than prediction. Regardless, the timing of the *measurement window* in a naturalistic driving video does need to be investigated as it points to different levels of experience may predispose drivers to hazard detection vulnerabilities that manifest at different stages on the driving task. These investigations should include several task versions with different timing windows for the same stimuli. This research design would clearly examine the influence of timing independent from possible stimuli specific responses that is a limitation of the current study. Additionally, this would improve the input to driving monitoring algorithms determining the appropriate wait time between driver hazard orientation and expected SCR.

While this study describes our groups as differing in experience, exposure and experience are not the same thing. An individual driving on the same routes is not likely to be as experienced as an individual who is driving on different road types and in varied traffic scenarios. However, epidemiological evidence suggests that exposure and experience (and crash risk) are related (Elvik, 2006). As there is no established measure of experience, participants were screened for inclusion in the study based on their exposure as a proxy of experience. Additionally, there was a significant difference in age between the two groups, as well as a lower age range of the Learner group, but age was used as covariate in the analyses.

In conclusion, this experimental study used a novel Driving Hazard Perception Task and measured SCR during videos where a hazard occurred (Event), and videos of routine driving (Non-Event). A medium effect size suggests that videos containing a hazard (Event) appeared to discriminate between novice and more experienced drivers, but the sample was underpowered to reach statistical significance. The confounding influence of *Onset Time* may be an additional factor influencing the findings. The decreased likelihood of SCR as the videos progress may be an indicator of situational awareness that needs further investigation. While physiological measures such as SCR may be useful for research and real-time state measurement relating to automated technologies and situational awareness, more needs to be understood with regard to experimental design and use of stimuli for validating such standards. Regardless, our research provides evidence that SCRs relative to hazard perception do differ based on experience, and this needs to be included in the model of driver state monitoring to properly understand the physiological signal and to facilitate safe integration of ADAS technologies in vehicles.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Johns Hopkins Medicine Institutional Review Board (IRB# 170352). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

TC wrote the manuscript and was responsible for data collection, analysis, and interpretation. TC also assisted in the development of the novel driving hazard perception task. JE assisted in the writing of the manuscript, data interpretation, and was responsible for the research concept. NK assisted in the writing of the manuscript, data interpretation, and laid the foundational work for this paper. KS assisted in the writing of the manuscript, data interpretation, led the development of the driving hazard perception task, and was responsible for the research concept. All authors contributed to the article and approved the submitted version.

FUNDING

This research was made possible by funding from the Johns Hopkins Science of Learning Institute (1604040036).

ACKNOWLEDGMENTS

We would like to thank Dr. John Desmond, Dr. Mitchell Doucette, Leland Brigham, Andrew Hellinger, Karolyne Kim, Edward Llinas, Christopher Rodman, Ariane Sharifi, and Charissa Zhou for their valuable contributions during the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.619104/full#supplementary-material

REFERENCES

Balters, S., and Steinert, M. (2017). Capturing emotion reactivity through physiology measurement as a foundation for affective engineering in engineering design science and engineering practices. *J. Intell. Manuf.* 28, 1585–1607. doi: 10.1007/s10845-015-1145-2

Barnard, M. P., and Chapman, P. (2016). Are anxiety and fear separable emotions in driving? A laboratory study of behavioural and physiological responses to different driving environments. *Accid. Anal. Prev.* 86, 99–107. doi: 10.1016/j.aap.2015.10.021

Bechara, A., and Damasio, A. R. (2005). The somatic marker hypothesis: a neural theory of economic decision. *Games Econ. Behav.* 52, 336–372. doi: 10.1016/j.geb.2004.06.010

Braithwaite, J. J., Watson, D. G., Jones, R., and Rowe, M. (2013). A guide for analysing electrodermal activity (EDA) and skin conductance responses (SCRs) for psychological experiments. *Psychophysiology* 49, 1017–1034.

Cohen, J. (1992). Statistical power analysis. *Curr. Dir. Psychol. Sci.* 1, 98–101. doi: 10.1111/1467-8721.ep10768783

Collet, C., and Musicant, O. (2019). Associating vehicles automation with drivers functional state assessment systems: a challenge for road safety in the future. *Front. Hum. Neurosci.* 13:131. doi: 10.3389/fnhum.2019.00131

Crundall, D. (2016). Hazard prediction discriminates between novice and experienced drivers. *Accid. Anal. Prev.* 86, 47–58. doi: 10.1016/j.aap.2015.10.006

Crundall, D., Chapman, P., Trawley, S., Collins, L., van Loon, E., Andrews, B., et al. (2012). Some hazards are more attractive than others: drivers of varying experience respond differently to different types of hazard. *Accid. Anal. Prev.* 45, 600–609. doi: 10.1016/j.aap.2011.09.049

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Putnam.

Darzi, A., Gaweesh, S. M., Ahmed, M. M., and Novak, D. (2018). Identifying the causes of drivers' hazardous states using driver characteristics, vehicle kinematics, and physiological measurements. *Front. Neurosci.* 12:568. doi: 10.3389/fnins.2018.00568

Dawson, M. E., Schell, A. M., and Filion, D. L. (2017). "The electrodermal system" in *Handbook of psychophysiology*. eds. J. T. Cacioppo, L. G. Tassinary and G. G. Berntson (Cambridge University Press), 217–243.

Ehsani, J. P., Seymour, K. E., Chirles, T., and Kinnear, N. (2020). Developing and testing a hazard prediction task for novice drivers: a novel application of naturalistic driving videos. *J. Saf. Res.* 73, 303–309. doi: 10.1016/j.jsr.2020.03.010

Elvik, R. (2006). Laws of accident causation. *Accid. Anal. Prev.* 38, 742–747. doi: 10.1016/j.aap.2006.01.005

Grayson, G. B., and Sexton, B. F. (2002). The development of hazard perception testing. Transport Research Laboratory.

Healey, J. A., and Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Trans. Intell. Transp. Syst.* 6, 156–166. doi: 10.1109/TITS.2005.848368

Helander, M. (1978). Applicability of drivers' electrodermal response to the design of the traffic environment. *J. Appl. Psychol.* 63, 481–488. doi: 10.1037/0021-9010.63.4.481

Horswill, M. S., and McKenna, F. P. (2004). "Drivers' hazard perception ability: situation awareness on the road" in *A cognitive approach to situation awareness*. eds. S. Banbury and S. Tremblay (Aldershot, UK: Ashgate), 155–175.

Hulbert, S. F. (1957). Drivers' GSRs in traffic. *Percept. Mot. Skills* 7, 305–315. doi: 10.2466/pms.1957.7.3.305

Kinnear, N., Kelly, S. W., Stradling, S., and Thomson, J. (2013). Understanding how drivers learn to anticipate risk on the road: a laboratory experiment of affective anticipation of road hazards. *Accid. Anal. Prev.* 50, 1025–1033. doi: 10.1016/j.aap.2012.08.008

Krompier, J. (2017). Safety first: the case for mandatory data sharing as a federal safety standard for self-driving cars. *Univ. Ill. J. Law Technol. Policy* 2, 439–468.

Lohani, M., Payne, B. R., and Strayer, D. L. (2019). A review of psychophysiological measures to assess cognitive states in real-world driving. *Front. Hum. Neurosci.* 13:57. doi: 10.3389/fnhum.2019.00057

McKenna, F., and Crick, J. (1991). "Experience and expertise in hazard perception." in *Behavioural research in road safety. Proceedings of a Seminar Held in at Nottingham University*. September 26–27, 1990.

Michaels, R. M. (1960). Tension responses of drivers generated on urban streets. *Highw. Res. Board Bull.* 271, 29–44.

Petracca, E. (2020). Two and a half systems: the sensory-motor system in dual-process judgment and decision-making. *J. Neurosci. Psychol. Econ.* 13, 1–18. doi: 10.1037/npe0000113

Phelps, E. A., Lempert, K. M., and Sokol-Hessner, P. (2014). Emotion and decision making: multiple modulatory neural circuits. *Annu. Rev. Neurosci.* 37, 263–287. doi: 10.1146/annurev-neuro-071013-014119

Quimby, A., Maycock, G., Carter, I., Dixon, R., and Wall, J. (1986). Perceptual abilities of accident involved drivers. Research Report-Transport and Road Research Laboratory, 27.

Tagliabue, M., Gianfranchi, E., and Sarlo, M. (2017). A first step toward the understanding of implicit learning of hazard anticipation in inexperienced road users through a moped-riding simulator. *Front. Psychol.* 8:768. doi: 10.3389/fpsyg.2017.00768

Tagliabue, M., and Sarlo, M. (2015). Affective components in training to ride safely using a moped simulator. *Transp. Res. F* 35, 132–138. doi: 10.1016/j.trf.2015.10.018

Taylor, D. H. (1964). Driver's galvanic skin response and the risk of accident. *Ergonomics* 7, 439–451. doi: 10.1080/00140136408930761

Ventsislavova, P., Crundall, D., Baguley, T., Castro, C., Gugliotta, A., Garcia-Fernandez, P., et al. (2019). A comparison of hazard perception and hazard prediction tests across China, Spain and the UK. *Accid. Anal. Prev.* 122, 268–286. doi: 10.1016/j.aap.2018.10.010

Vogelpohl, T., Kühn, M., Hummel, T., and Vollrath, M. (2019). Asleep at the automated wheel—sleepiness and fatigue during highly automated driving. *Accid. Anal. Prev.* 126, 70–84. doi: 10.1016/j.aap.2018.03.013

Wetton, M. A., Hill, A., and Horswill, M. S. (2011). The development and validation of a hazard perception test for use in driver licensing. *Accid. Anal. Prev.* 43, 1759–1770. doi: 10.1016/j.aap.2011.04.007

Disclaimer: This work was prepared while KS was employed at Johns Hopkins University and Kennedy Krieger Institute. The opinions expressed in this article are the author's own and do not reflect the view of the National Institutes of Health, the Department of Health and Human Services, or the United States government.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 07 June 2021
doi: 10.3389/fpsyg.2021.617889

Grip Force on Steering Wheel as a Measure of Stress

Yotam Sahar†, Tomer Elbaum†, Michael Wagner†, Oren Musicant†, Tehila Hirsh and Shraga Shoval†*

Department of Industrial Engineering and Management, Ariel University, Ariel, Israel

OPEN ACCESS

Edited by:
*Silvia Riva,
St Mary's University, Twickenham,
United Kingdom*

Reviewed by:
*Javier M. Antelis,
Monterrey Institute of Technology
and Higher Education (ITESM),
Mexico
Pierluigi Cordellieri,
Sapienza University of Rome, Italy*

***Correspondence:**
*Yotam Sahar
yotam.sahar@msmail.ariel.ac.il*

†ORCID:
*Yotam Sahar
orcid.org/0000-0002-2499-3321
Tomer Elbaum
orcid.org/0000-0002-4455-6153
Michael Wagner
orcid.org/0000-0002-8448-7906
Oren Musicant
orcid.org/0000-0002-6639-0262
Shraga Shoval
orcid.org/0000-0002-0582-4821*

Specialty section:
*This article was submitted to
Performance Science,
a section of the journal
Frontiers in Psychology*

Received: *15 October 2020*
Accepted: *09 March 2021*
Published: *07 June 2021*

Citation:
*Sahar Y, Elbaum T, Wagner M,
Musicant O, Hirsh T and Shoval S
(2021) Grip Force on Steering Wheel
as a Measure of Stress.
Front. Psychol. 12:617889.
doi: 10.3389/fpsyg.2021.617889*

Driver performance is crucial for road safety. There is a relationship between performance and stress such that too high or too low stress levels (usually characterized by stressful or careless driving, respectively) impair driving quality. Therefore, monitoring stress levels can improve the overall performance of drivers by providing either an alert or intervention when stress levels are sub-optimal. Commonly used stress measures suffer from several shortcomings, such as time delays in indication and invasiveness of sensors. Grip force is a relatively new measure that shows promising results in measuring stress during psychomotor tasks. In driving, grip force sensor is non-invasive and transparent to the end user as drivers must continuously grip the steering wheel. The aim of the current research is to examine whether grip force can be used as a useful measure of stress in driving tasks. Twenty-one participants took part in a field experiment in which they were required to brake the vehicle in various intensities. The effects of the braking intensity on grip force, heart rate, and heart rate variability were analyzed. The results indicate a significant correlation between these three parameters. These results provide initial evidence that grip force can be used to measure stress in driving tasks. These findings may have several applications in the field of stress and driving research as well as in the vehicle safety domain.

Keywords: grip force, stress, steering wheel, driving, heart rate variability, heart rate, psychomotor tasks, physiologic indices

INTRODUCTION

"Will I have to learn how to drive?" asked the 10 year-old daughter of one of the authors recently. Although it is a common belief that autonomous cars will take over in the next few years, the more probable answer is that she would have to learn how to drive unless she is willing to rely solely on public transportation. According to Litman (2019), it is not until the 2050s that fully autonomous cars, in which no human involvement is required (known as level 5 in the autonomous driving scale defined by the Society of Automotive Engineers International), will be commonly used. Until such time, driver performance will remain critical in road safety.

Road accident investigation (Hendricks et al., 2001) and observational studies (Dingus et al., 2016) indicate that about 90% of all road accidents result from human error. To improve road safety, it is essential to recognize the factors affecting driver performance, specifically, factors that can be moderated to improve driver performance and road safety. This paper focuses on one such factor—temporal driver stress.

Stress is defined as "...a real or interpreted threat to the physiological or psychological integrity of an individual that results in physiological and/or behavioral responses" (McEwen, 2000). According to the transactional model, stress is the outcome of appraisals of demands and personal competence, together with a coping strategy that mediates between external demands (Lazarus and Folkman, 1984). Appraisal processes generate various outcomes or stress symptoms: physiological, emotional, and behavioral (Matthews, 2001). McGrath (1976) stated that stress results from an interaction between three elements: perceived demand, perceived ability to cope, and perceived importance of coping with the demand.

The concept of "stress" is often used as a synonym to the "mental workload" concept (Staal, 2004). Accordingly, the mental workload is also referred to as a transactional concept since it represents an interaction between mental capacities and task demands (Dehais et al., 2020). Furthermore, some stress definitions hold that stress represents a higher mental workload (Brookhuis and De Waard, 2010; Hou et al., 2015).

An additional definition by Mulder and Moray (1979) suggests that mental workload is "...an inferred construct that mediates between task difficulty, operator skill, and observed performance" (Mulder and Moray, 1979; p. 443). Thus, based on Mulder and Moray's (1979) definition of mental workload and the mentioned definition of stress by McGrath (1976), the main difference between mental workload and stress stems from the perceived ability to cope with the demands, namely, unlike mental workload (Mulder and Moray, 1979), stress is caused by the perceived consequences of failing to cope with the demands (McGrath, 1976).

Indeed the confusion between the terms "workload" and "stress" is an entangled issue, as these terms are not yet adequately defined nor unambiguously differentiated in the literature. Furthermore, the manifestations through the sympathetic nervous system of stress and workload are similar and may be indistinguishable (Alsuraykh et al., 2019). This confusion will not be resolved within the framework of the present study, and henceforth we shall use only the term "stress" for simplicity.

One of the common descriptions of the relationship between performance and stress is based on findings made more than a century ago by Yerkes and Dodson (1908), later described as an "inverted U-shaped curve." According to the inverted U-shaped curve, the upper and lower levels of stress yield unsatisfactory performance, while the mid-level produces the best performance (Hancock, 1989; Hancock and Szalma, 2008). Concerning driving, higher stress levels are harmful to driver performance (Qu et al., 2016). At the other end of the scale, very low stress, termed by Hancock and Szalma (2008) as "under-stimulation," was found to impair driver performance (Joosen et al., 2017).

The construct of stress is divided into chronic and acute stress (Segerstrom and Miller, 2004). Chronic stress refers to a continuous state beyond a specific driving situation. Acute stress refers to a single event of short duration or a "micro-event" (Meyer et al., 2010). In the context of driving, short-duration events that may cause stress are unexpected events that, in turn, require sudden and unplanned reactions (Davies and Underwood, 2000). Stress-inducing driving events require two main maneuvers from the driver: manipulating the steering wheel and braking. Studies on driver stress have used manipulations such as driving through a labyrinth or slalom to force the driver to manipulate the steering wheel (Zontone et al., 2020) and pedestrians or other objects erupting into the road to force the driver to brake intensely (Daviaux et al., 2020).

Acute stress during driving causes a high mental workload (Wiberg et al., 2015) and adverse effects (Frasson et al., 2014) that may decrease driver performance (Brookhuis and De Waard, 2010; Rastgoo et al., 2019). Adding automation would not necessarily provide drivers with a less effortful working environment (Botzer et al., 2016). However, detecting acute stress during driving may allow various interventions that would reduce potential risks. An example of such an application is stress-adaptive car systems that modify the parameters of in-vehicle driver-aiding systems based on the driver's stress levels (Collet and Musicant, 2019). Another application is in-car just-in-time stress management interventions (e.g., mild temperatures and music, bio-feedback interfaces, and chatbots) administered when the stress levels are too high (Balters et al., 2019).

Acute stress is manifested physiologically by the sympathetic nervous system, which stimulates the body's "fight or flight" response. This response is antagonistic to the parasympathetic nervous system, which reduces stress (Contrada and Baum, 2011). These reactions can be measured in many ways, such as maximal heart rate (HR) (Kudielka et al., 2004), power spectra in specific frequency bands of the heart rate variability (HRV) signal (Allen et al., 2014), galvanic skin response (GSR) (Al-Fudail and Mellar, 2008), eye-related measures (Matthews et al., 2015), and cortisol levels (Yamaguchi et al., 2006).

In HRV analysis, the cardiac signal is divided into three components: VLF (very low frequency, 0–0.04 Hz), LF (low frequency, 0.04–0.15 Hz), and HF (high frequency, 0.15–0.4 Hz) (Malik, 1996). The LF measure reflects the sympathetic system (and therefore is related to stress), while the HF measure reflects the parasympathetic system (Sztajzel, 2004). The LF/HF ratio indicates the balance between the sympathetic and parasympathetic divisions of the autonomic nervous system and is used as a measure of stress as well (Kristal-Boneh et al., 1995; McCraty et al., 1995).

The measures mentioned above suffer from several practical shortcomings. Cortisol level analysis is not suitable during task performance since it is not easily measured continuously. GSR, HR, and HRV measurements may be inconvenient for use in realistic driving scenarios (Healey and Picard, 2005) and may even be considered obtrusive (Dinges et al., 2005). These measures also suffer from delays in the measurement (the time gap between the stressful event and the observed response). GSR's delays are between 2 and 11 s long (Kucera et al., 2004; Dawson et al., 2007; Bruun, 2018), and valid analysis of changes in HRV may require a continuous signal of 4–5 min in duration (Nickel and Nachreiner, 2003). Cortisol measurement reacts to stressors with a delay of several minutes, sometimes up to half an hour (Kirschbaum and Hellhammer, 1994).

Eye-related measures, such as pupil dilation (Palinko et al., 2010), fixation duration (Matthews et al., 2015), saccade rate, and gaze shifts (Tomer et al., 2018), as well as saccadic range (May et al., 1990), were reported as indices of mental workload. There is limited evidence for ocular measures as stress indicators, and most findings are concerning pupil dilation (Pedrotti et al., 2014). Though a stressor's administration leads to pupil dilation, the pupil's size is susceptible to light intensity and requires an illumination-controlled environment—not practical for non-lab applications (Pedrotti et al., 2014). While eye closure level is useful in measuring drowsiness (Grimberg et al., 2020), it is also not useful for stress measurement. Finally, GSR and HRV do not always correlate strongly with stress, neither induced (e.g., Rohleder et al., 2006; Zhai and Barreto, 2006) nor measured by well-established measures, such as cortisol level (e.g., Healey and Picard, 2005).

Therefore, it may be useful to develop additional stress measures that provide solutions to the issues discussed earlier. One such possible measure is grip force, found to be capable of measuring stress in a prompt and non-invasive manner.

Continuous and repeated stress measurements using non-invasive methods have been of great interest in recent years. Hernandez et al. (2014) measured the amount of force applied to a computer keyboard and a mouse. Although not suitable for use in driving tasks, it shows that hand muscle tonus measurement has the potential to be an indication of stress. Wahlström et al. (2002) examined the effect of stressors (e.g., time pressure and verbal provocation) on various factors, including the grip force upon a computer mouse. Grip force increased when stressors were used. However, this effect was attributed both to stress and the mouse operation's speed. In another research that used grip force on a computer mouse, Liao et al. (2006) found greater grip force in response to higher time pressure. It should be noted that these studies manipulated the mental workload rather than the stress, as no direct implications for low-performance outcomes were involved. These two studies also used static tasks (e.g., math problems and typing tasks), making it difficult to extend their findings to other contexts such as driving tasks.

Wagner et al. (2015) examined the feasibility of grip force as a measure of stress in tracking tasks. Grip force was higher in the presence of stress. This study provides initial evidence of distinguishing between stressful and non-stressful conditions during physical tracking tasks by measuring the grip force. Recently, these findings were successfully reconstructed (Botzer et al., 2020). Mühlbacher-Karrer et al. (2017) used the measurement of a driver's grip force on a steering wheel as part of a stress estimation system. However, the grip force's contribution to the stress level calculation was only 10%. Another limitation of this study is that it was conducted only in a simulator and not in actual driving, ignoring the effect of other factors on grip force other than stress (e.g., vehicle accelerations). Additionally, the mere driving an actual car in an experiment is known to induce a state of stress (Balters et al., 2019), as the consequences of one's performance are tangible, unlike participating in an experiment conducted in a simulator.

We aim to study the relationship between grip force and other more common indices of stress, namely, physiological indices and performance indices during short-duration driving events. The purpose of the study, hopefully its main contribution, is to provide initial empirical evidence to grip force as an index of stress in driving tasks. To this end, one should define the appropriate driving events that elicit acute stress. In previous research, forced changes in driving behavior were found to cause stress (Ross and Burnett, 2001; Lee and Winston, 2016; Saxena, 2017). Such changes may result from the road conditions and unexpected factors that force emergency maneuvers (i.e., a braking sign or a figure bursting into the road). Specifically, stopping in response to a STOP sign during driving and the necessity to brake have been found to induce stress in an experimental context (Min et al., 2002; Collet et al., 2014; Prasolenko et al., 2017; Sugiono et al., 2019).

Thus, in this study, we used diverse heart rate measurements to show that braking events lead to higher stress as manifested in psychophysiological changes of heart measurements. By manipulating stress-inducing driving events and measuring their effect on the heart and also grip force measurements, we aim to verify the use of grip force as a valid measure of stress during driving.

In the current research, grip force and heart rate measures (HRV and HR) were recorded while driving and braking in various intensities (in response to a STOP sign), as described in "Experimental Setup and Methods." We hypothesized that intense braking during driving affects grip force and elicits correlations between grip force, HRV, and HR measures. Thus, grip force data were analyzed as a function of braking intensity, and correlations among grip force and heart rate measures (HRV and HR) were calculated to validate the grip force measure against an accepted measure of stress (see section "Results"). HRV and HR data were also analyzed as a function of braking intensity, serving as a manipulation check.

EXPERIMENTAL SETUP AND METHODS

Participants

Twenty-one participants took part in this study. Due to technical issues (failure in recording the data of one participant), we used the data from 20 participants. All participants were bachelor course students. All participants were male, between the ages of 24 and 34 (average 28.45, SD 2.18), and had a private car driver's license for at least 4 years.

Before the experiment, the participants underwent a safety briefing, including a description of the experimental task, and completed an informed consent statement. A safety supervisor, positioned in the front passenger seat, was responsible for maintaining safety during the experiment.

Apparatus

The experiment was conducted using an instrumented Kia Nero (hereafter, the "Mobile-Lab"; **Figure 1**). The Mobile-Lab[1] is equipped with sensors for monitoring the vehicle and road environment, including inertial measurement units (Lidar, GPS antennas, and several cameras), and sensors for monitoring indices of the driver using the Mindware Mobile Impedance

[1] http://www.ariel.ac.il/wp/mobile-lab/

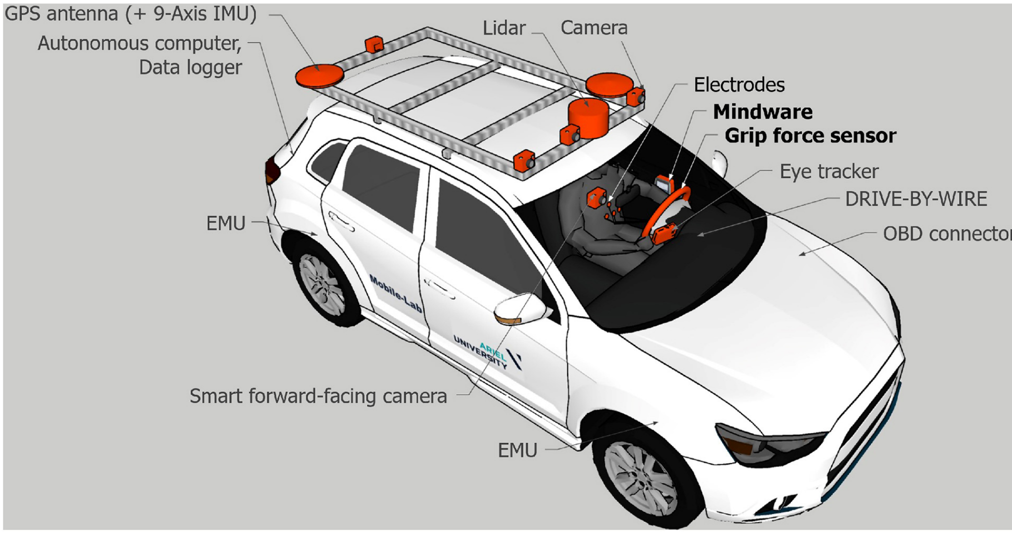


FIGURE 1 | The Mobile Lab, equipped with GPS and sensors, Mindware and grip force sensors, as well as additional equipment that was not used in the current research.

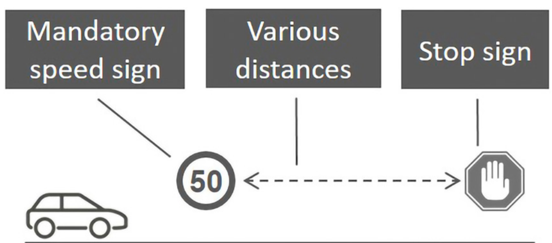


FIGURE 2 | The general layout of the experimental driving session. Experimental manipulations: mandatory speed sign (one of two speeds: 50 or 60 km/h) and STOP sign (at various distances: 15, 20, 25, 30, 35, or 40 m). Each participant performed the driving sessions under all combinations of speeds and braking distances at random order.

Cardiograph (heart activity measurement system Model 50–2303-00, 2014, with a sampling rate of 500 Hz and 24-bit ADC digitization). Cardiac data were recorded from electrodes affixed to the participant's chest. A grip force self-developed measurement system was used (utilizing a force-sensitive resistor sensor, sampled by an Arduino UNO R3 board). Both systems were equipped with three-axis accelerometers.

Procedure

The participants performed 12 experimental driving sessions. Each session involved driving along a straight path of approximately 200 m at one of two mandatory speeds (50 or 60 km/h) and braking at varied mandatory distances (15, 20, 25, 30, 35, or 40 m), as shown in **Figure 2**. Each participant performed the driving sessions under all combinations of speed and braking distances (two speeds × six distances = 12 conditions) in random order. Each participant performed two training sessions of about 2 min (during which they got acquainted with the experimental path) and 12 experimental driving sessions (one for each condition), which lasted nearly 20 min overall. After each experimental driving session, the participant was instructed to leave the vehicle stationary for 15 s.

Data Preparation and Analysis Procedures

To analyze the data from both acquisition systems used (Mindware and the self-developed grip force measurement system), first we synchronized the data (see section "Data Synchronization"). After the data synchronization, we used the acceleration data to identify the peak deceleration for each driving session. Each session was characterized by a static phase (of at least 15 s), an acceleration phase, and a deceleration phase (as demonstrated in **Figures 3**, **4**), in contrast to other acceleration data (e.g., data from the training sessions) which were less organized. Later, HRV and HR heart rate measures and grip force measures were calculated (see "Heart Rate Measures Calculation" and "Grip Force Data Preparation and Calculation") and analyzed (see section "Data Analysis"). **Figure 5** summarizes the flow of these processes.

Data Synchronization

Cardiac activity data and grip force data were synchronized *post factum*, according to the accelerometers' data (from both measuring systems), using a dynamic time wrapping (DTW) algorithm. Both data acquisition systems (Mindware for HRV data and the self-developed grip force measurement system) were equipped with three-axis accelerometers (X, Y, and Z), which were fixed to the vehicle's chassis. Acceleration data were recorded by each system in a synchronized manner with the physiological data (HRV and HR data at the Mindware system and grip force data at the grip force measurement system).

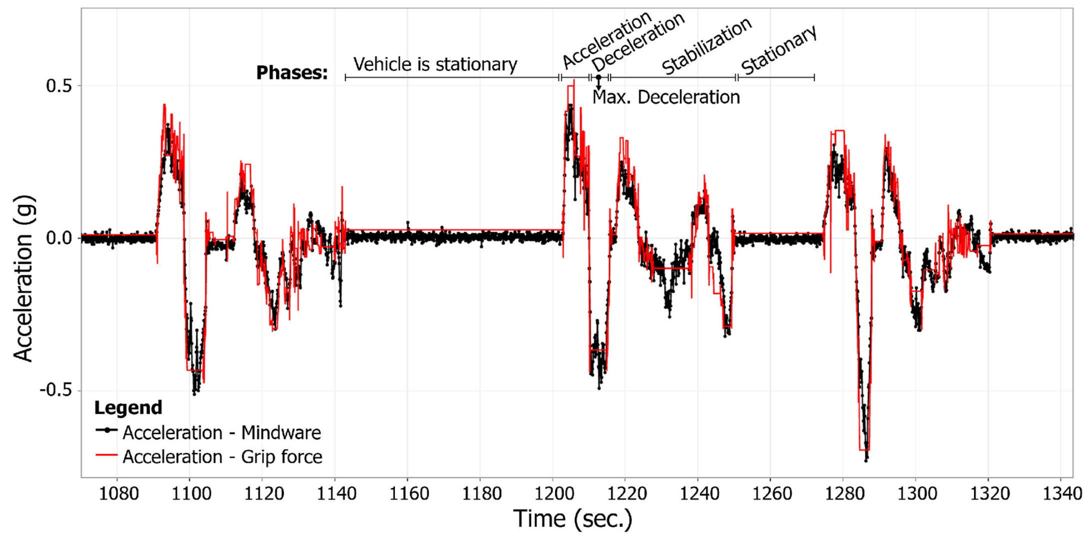


FIGURE 3 | Synchronized acceleration vectors of both accelerometers (of the Mindware system—in black—and the self-developed grip force measurement system—in red) in the direction of the vehicle's travel, during three successive experimental driving sessions, of a single participant.

First, the grip force system's sampling rate was uneven and ranged between 80 and 120 ms (8–12 Hz). Since a pre-condition of the DTW procedure is that "the data should be sampled at equidistant points in time" (Senin, 2008), a standard method to deal with this requirement is resampling the data as has been done in the current study. The resampled grip force data had a 10 Hz sampling rate. The Mindware system had a sampling rate of 500 Hz.

Next, for each system separately, a unified vector of the three axes was calculated ($\sqrt{X^2+Y^2+Z^2}$), using a 1-s sliding window. Finally, a DTW algorithm was used to synchronize these acceleration vectors from both systems. A similar synchronization method has been used by Mantilla et al. (2017) to detect temporal synchronization. DTW was proven to be a robust distance measure for time series, enabling the matching of similar plots even if they are out of phase in the time axis (Keogh and Ratanamahatana, 2005).

Heart Rate Measures Calculation

To properly calculate the HRV LF measure and LF/HF ratio, a minimum sliding window size of 30 s is required (De Rivecourt et al., 2008; Wang et al., 2009). Typically, an HRV window size is between 20 s and several minutes. For example, Mulder et al. (2009) used a 300-s window, whereas Healey and Picard (2005) used 100- and 300-s windows. It should be noted that the decision about the window size is often arbitrary.

A small window size of 30 s is sufficient if combined with a short-time Fourier transform (Li et al., 2019). Therefore, in the current research, a sliding window with a window size of 30 s was implemented to the cardiac data for computing the HRV LF and HF measures and the LF/HF ratio as well as for heart rate. The center of the window was determined according to the braking event's maximal deceleration. Since there were pauses of 15 s after each braking event (as described in "Procedure"), there could not have been any other experimental effect on the physiological signals during the entire window size other than the effect of the forced braking event itself. In addition, to account for the chi-square distribution of HRV and heart rate values (Van Roon et al., 2004), a natural log transformation was applied to these measures.

Grip Force Data Preparation and Calculation

Grip force data under the threshold of 5 N (newton) was considered mostly white noise due to its proximity to the lower boundary of the grip force sensitivity. Accordingly, grip force data below 5 N was excluded. Grip force data was collected and resampled at a 10 Hz rate (as detailed in "Data Synchronization"). Grip force measures were calculated to explore various aspects of grip force in relation to the other measures. The grip force measures calculated were mean, maximum (max), and standard deviation (sd).

These grip force measures were calculated using a 2-s time window, centered around each braking event's peak deceleration. Due to this study's preliminary nature, there are no widely accepted guidelines to rely on for grip force window size in stress measurement. In defining the time window, we referred to the preliminary findings from an ongoing study regarding this issue, which shows an initial inclination toward the use of a narrow time window of fewer than 5 s in calculating grip force as a measurement of stress (Botzer et al., 2020). Additionally, we based this decision on the indication received from an analysis detailed and illustrated in **Appendix A**.

Based on psychophysical and physiological reports and stress models (Liu and Ulrich, 2014), exponential logarithmic or sigmoid transformation functions, rather than linear functions, are expected. For example, electromyography is reported to have a logarithmic transfer function (Rezazadeh et al., 2012). Therefore, a natural log transformation was applied to the grip

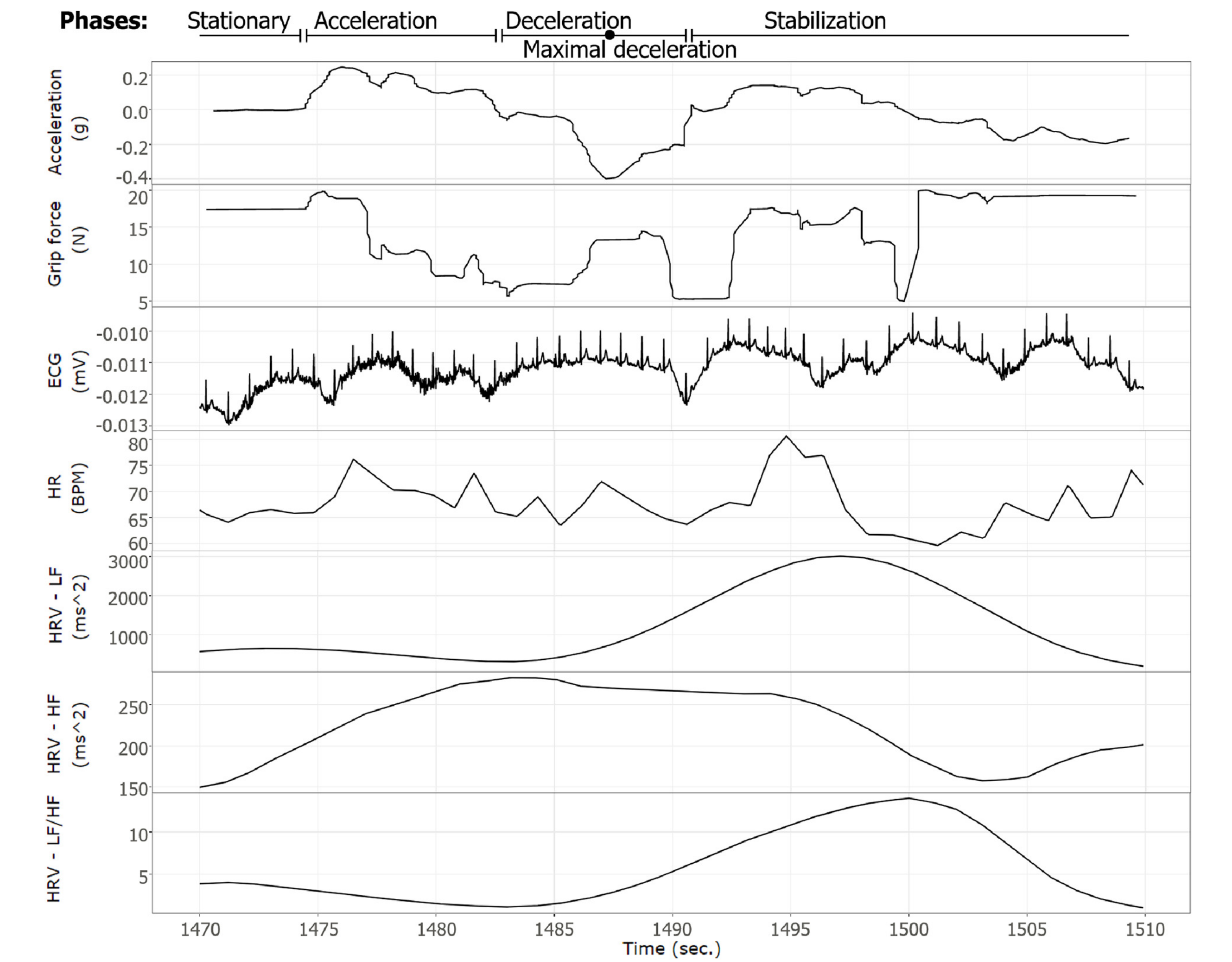


FIGURE 4 | An example of a braking event as expressed by the various measures recorded. A single braking event of participant number 1. The *X*-axis represents time (s), plots (from top to bottom): acceleration (g), grip force (N), raw ECG signal (mV), HR (BPM), HRV—LF (ms^2), HRV—HF (ms^2), and HRV—LF/HF ratio.

force calculated measures since the higher end of grip force is limited by the maximal grip strength (for each individual).

Data Analysis

The linear mixed model (LMM) was selected to analyze the effects of braking intensity on the various physiological measures, primarily due to its suitability to repeated-measures designs (Peat and Barton, 2014). In this method, within-subject correlations are modeled using the covariance structure, built on the variance around the outcome measurement at each time point and on the correlations between measurements taken at different times from the same participant (Peat and Barton, 2014).

A meta-analysis method was used to analyze the correlation of grip force measures with HR and HRV measures. The correlations between these measures for each participant served as the meta-analysis input. This procedure enabled the consideration of inter-personal variance (for further description, see "Results").

RESULTS

To examine our hypothesis that compelled braking during driving elicits correlating measured patterns of grip force and heart rate, we first explored the braking events' effects on grip force. All three LMMs were fitted to the data with the assumption of a linear relationship in order to study the nature of the relationship between the three grip force measures (i.e., Ln transformation of mean, max, and sd of grip force) and the *D* parameter for maximal deceleration (i.e., the intensity of braking events). Maximal deceleration (D) was included in the model as the predictor and grip force measures as the predicted variables (see rows 1–3 in **Table 1** for a formal description of these three-mixed effect models and **Figures 6A–C** for their visual depiction).

Significant main effects were found for maximal deceleration on all three grip force transformations ($p < 0.001$; see rows 1–3 in **Table 1**). Based on the models' coefficients (rows 1–3 in **Table 1**

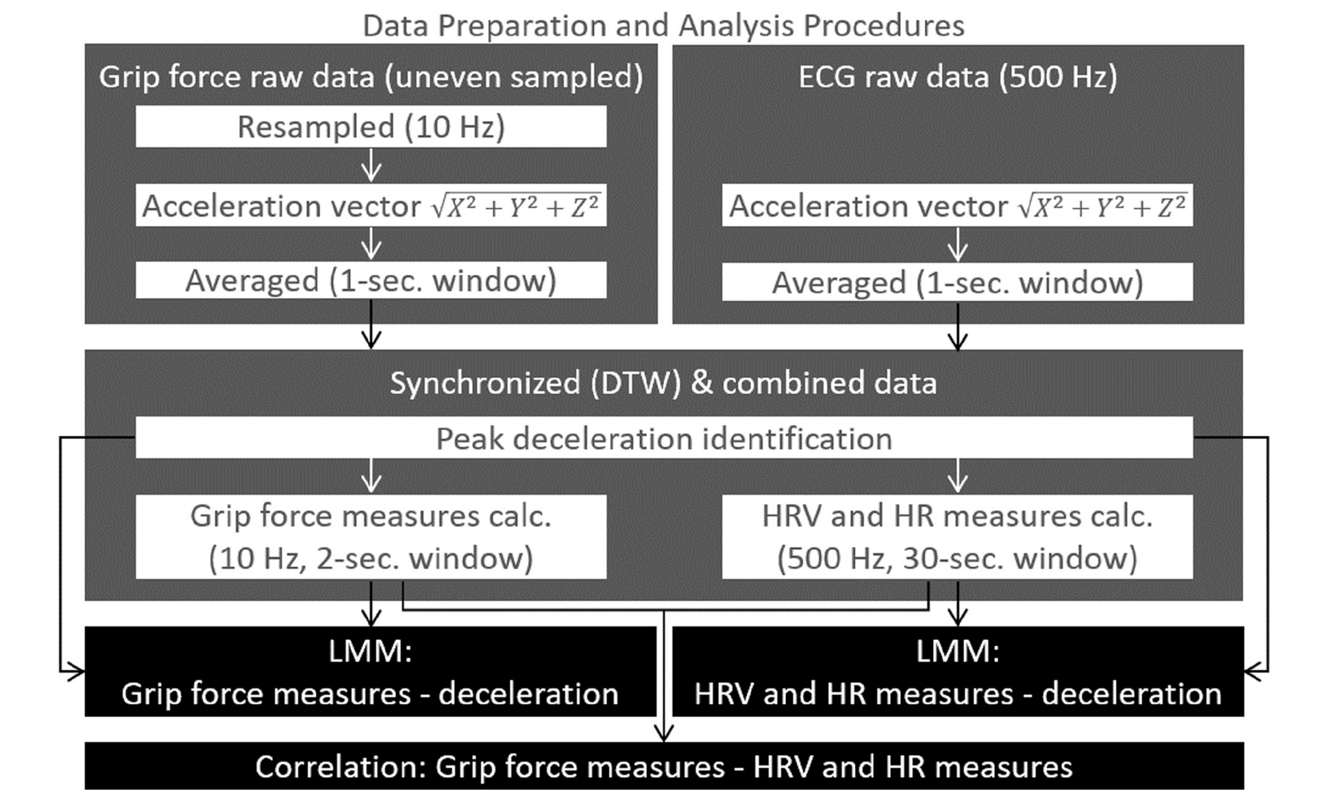


FIGURE 5 | Data preparation and analysis procedures: a workflow description of the data preparation procedures, data synchronization procedure, and the data analyses used.

and **Figures 6A–C**) and in accordance with our hypothesis, the findings indicate that higher deceleration (braking intensity) predicted greater grip force (i.e., mean and maximum grip force) and larger changes in grip force (i.e., grip force sd).

An additional four-LMM analysis was performed to investigate whether braking events elicit stress as manifested by HRV and HR measures (the models are described in rows 4–7 of **Table 1**). The four additional LMMs include the *D* parameter for maximal deceleration (i.e., braking intensity) as the predictor and Ln-transformed HRV (LF, HF, and LF/HF ratio) and HR measures as predicted variables (see rows 4–7 in **Table 1** for a formal description of these three mixed-effect models).

The additional LMM analysis showed a main effect for maximal deceleration on heart rate ($p < 0.001$; row 6 in **Table 1** and **Figure 6D**), and a moderate trend toward significance was also found on HRV LF/HF ratio ($p = 0.069$; row 4 in **Table 1** and **Figure 6E**). The LMMs for maximal deceleration on HRV

TABLE 1 | Summary of linear mixed model analyses for various models.

Model	β_0 (SE)	β_1 (SE)	*F*	*P*	Adj. R^2
Grip force					
$Ln(meanGF) = \beta_0 + \beta_1(D) + b_i + \varepsilon$	6.527 (0.044)	0.244 (0.061)	16.4	0.0001***	0.465
$Ln(maxGF) = \beta_0 + \beta_1(D) + b_i + \varepsilon$	6.512 (0.042)	0.318 (0.061)	26.54	< 0.0001***	0.375
$Ln(sdGF) = \beta_0 + \beta_1(D) + b_i + \varepsilon$	−6.724 (1.182)	10.401 (1.896)	30.963	< 0.0001***	0.209
HR and HRV					
$Ln(LF/HF) = \beta_0 + \beta_1(D) + b_i + \varepsilon$	1.333 (0.235)	0.21 (0.37)	3.33308	0.0693·	0.19
$Ln(LF) = \beta_0 + \beta_1(D) + b_i + \varepsilon$	5.852 (0.257)	−0.15 (0.354)	1.2628	0.2624	0.453
$Ln(HR) = \beta_0 + \beta_1(D) + b_i + \varepsilon$	4.438 (0.038)	0.097 (0.025)	41.89	< 0.0001***	0.882
$Ln(HF) = \beta_0 + \beta_1(D) + b_i + \varepsilon$	4.546 (0.264)	−0.262 (0.331)[a]	0.6404	0.4244	0.51

*mean GF, mean grip force; maxGF, maximum grip force; sdGF, standard deviation of the grip force; LF, HF, and LF/HF, heart rate variability measures; HR, heart rate, D, maximal deceleration (−g) during the braking event; b_i, random effect parameter for driver i; ε, error term. [a]HRV HF's negative value represents the inhibitory pattern of the parasympathetic system during stressful situations. It is assumed that $b_i \sim N(0, \sigma_b)$. Significance codes: 0 ≤ "***" < 0.001 < "**" < 0.01 < "*" < 0.05 < "." < 0.1 < " " ≤ 1.*

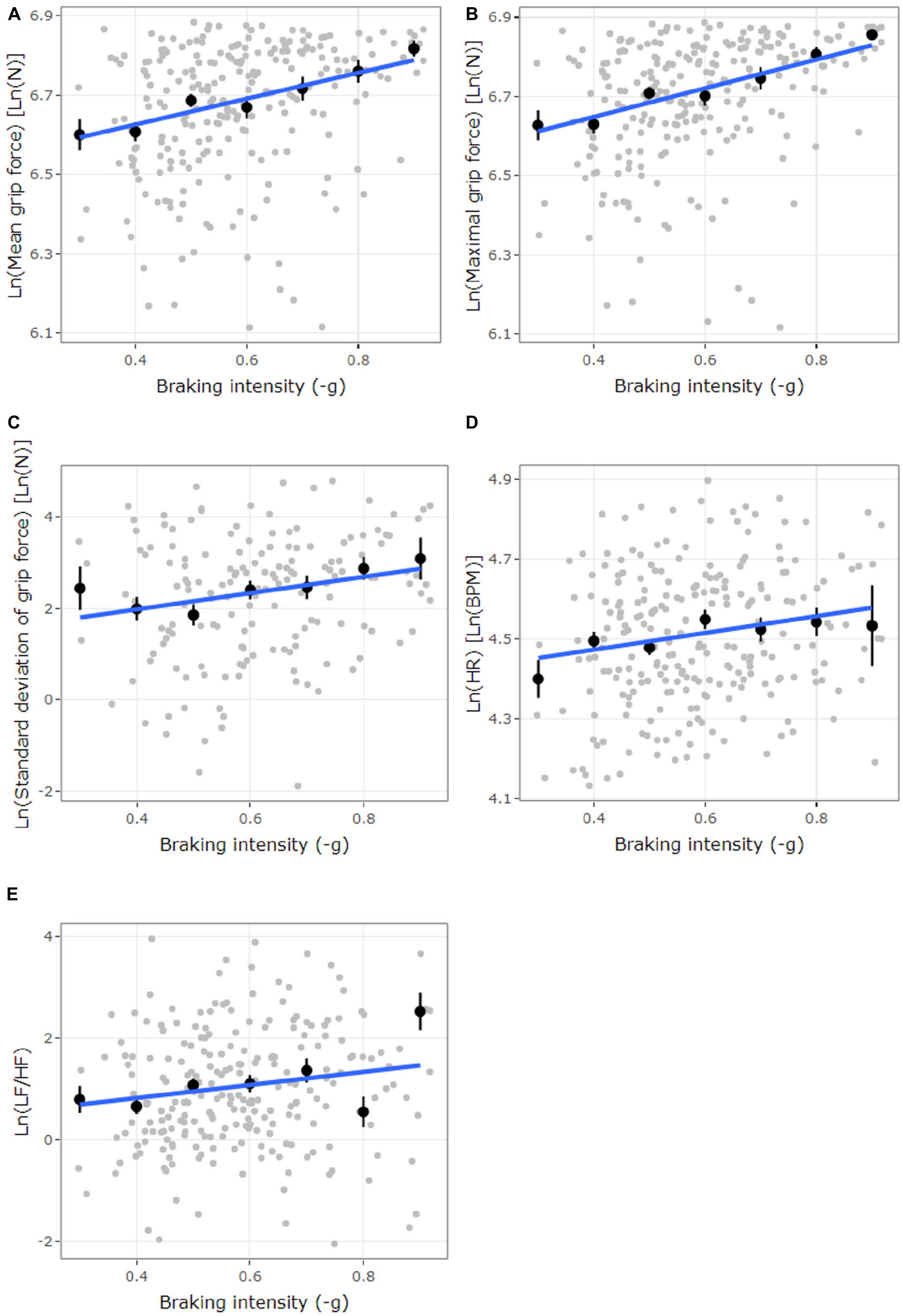


FIGURE 6 | The plots illustrate the significant LMM models described in **Table 1**. Higher *x*-axis values (-g) represent a higher braking intensity in this figure. **(A–C)** Ln-transformed grip force [Ln(N)], **(A)** mean grip force, **(B)** maximal grip force, and **(C)** standard deviation of grip force, as a function of braking intensity (–g). **(D)** Ln-transformed HR [Ln(BPM)] as a function of braking intensity (–g). **(E)** Ln-transformed HRV LF/HF as a function of braking intensity (–g). Gray dots represent observations. Black dots represent the mean of each binned group of observations (according to deceleration), with 95% confidence interval. The blue line represents smoothed conditional means using lm smoothing.

LF and HF were not significant (p = 0.262 and p = 0.424, respectively). Based on the additional models' coefficients for maximal deceleration on HRV LF/HF ratio and HR (rows 4 and 6 in **Table 1** and **Figures 6D,E**) and in accordance with the assertion presented in the introduction (i.e., that braking events induce stress), higher declaration (braking intensity) predicted greater HRV LF/HF ratio and HR.

Our hypothesis addressed the association between grip force and HRV and HR as prevalent measures of stress. Specifically, the hypothesis aimed to serve as an additional association between grip force and stress. To test this hypothesis, Pearson correlations were calculated separately for each participant, followed by a meta-analysis procedure. This integration of the two procedures (i.e., separate correlations followed by a meta-analysis) was designed to include the participants as a random effect, partly similar to using LMM. Separate correlations were conducted between heart measures (HRV and HR) and the grip force's central tendency indices (i.e., mean and maximum grip force) as mentioned above.

The meta-analysis procedure was applied for the separate correlations, using the "meta" R package. The meta-analyses examined all possible correlations between each heart measure and each grip force measure. The effect sizes were transformed into standard values using Fisher's r to z transformation (Rosenthal, 1991). The z-transformed score has a standard error of $^1/_{\sqrt{(n-3)}}$, where n is the number of braking events for each participant. The inverse of this error was used as a weight for each individual z-transformed score so that participants with smaller standard errors were given more emphasis. After this weighting, all participants' values were aggregated by averaging their z-transformed scores. Rosenthal (1991) suggests this as a conservative procedure. Finally, z-transformed scores were translated back to r values. The meta-analysis was applied to the grip force's central tendency indices and heart rate measures (HRV LF, HRV HF, HRV LF/HF ratio, and HR). Additionally, natural log transformation was used for all heart rate and grip force measures. **Table 2** contains the descriptions of the correlations and the values of the meta-analysis' coefficients.

Six of the eight correlations were significant, and the correlations' direction was consistent with our hypothesis (i.e., significant positive correlations for grip force with HRV LF/HF ratio and with HR; significant negative correlations for grip force with HRV HF negative values represent the inhibitory pattern of the parasympathetic system during stressful situations). The direction of the correlation between mean grip force and HRV LF was not consistent with the hypothesis, and the correlation between max grip force and HRV LF was not significant (p = 0.892). Although four of the six significant correlations were highly significant ($p < 0.001$), their effect size was relatively small (i.e., smaller than 0.3, Cohen, 1992).

DISCUSSION

This study's main aim was to examine the feasibility of detecting driver stress by grip force measurement in actual driving scenarios. Accordingly, the study's main goal of indicating that grip force can serve as a measure of stress in driving tasks was mostly achieved.

The assertion that braking as a response to a STOP sign elicits stress has a vast support (e.g., Min et al., 2002; Collet et al., 2014; Prasolenko et al., 2017; Sugiono et al., 2019). Accordingly, the LMM analyses conducted in the current study revealed that, during braking as a response to a STOP sign, maximal deceleration had a highly significant effect on the HR measure. In addition, a moderate trend toward significance was found regarding maximal deceleration's effect on HRV LF/HF ratio measure. Since these measures (HRV LF/HF ratio and HR) are referred to as stress measures (Kristal-Boneh et al., 1995; McCraty et al., 1995; Sztajzel, 2004; Allen et al., 2014), this finding offers additional support for braking as a stress-inducing driving event.

Braking intensity had a highly significant effect on all grip force measures. This finding, combined with the re-confirmed assertion that braking events induce stress, leads to a possible deduction that grip force constitutes an indication of stress. However, unlike HR and HRV, grip force may also be affected during braking by the task itself. Therefore, another possible explanation for consideration is that, during braking events, grip force may have been affected by the braking task solely or by a joint effect of stress and the braking task.

The analyses also showed correlations of HRV HF and LF/HF ratio with grip force's transformations. The correlations of HRV HF with grip force's transformations had a negative direction, consistent with the parasympathetic system's inhibitory pattern during stressful situations (Hall et al., 2004; Hjortskov et al., 2004; Vuksanović and Gal, 2007). The correlations of HRV LF/HF ratio with grip force's transformations had a positive direction as can be expected since HRV LF/HF ratio is known to increase during stressful situations. These findings may serve as further modest validation of grip force as a measure of stress. It should be noted that these correlations were weak. However, this may result from the different acquisition systems used (Milstein and Gordon, 2020). Additionally, weak correlations among different physiological measures of mental states are not an unfamiliar phenomenon (Contrada and Baum, 2011).

In this study, the participants had to brake in response to a STOP sign, which is known to induce stress, a response that was also found here as expressed by the effect of braking intensity on HR and on HRV LF/HF ratio. Grip force's magnitude was also affected by the intensity of these braking events, a finding that lends partial support to our hypothesis that grip force is an indication of stress during driving events. This hypothesis received further support by the correlations of grip force and HR and HRV measures. Therefore, it is feasible that grip force can be used as a measure of stress in braking events during actual driving. These findings may contribute to further investigations needed to establish this relatively new measure of stress, specifically in driving contexts.

As found in the current study, a stressor's physiological response during driving can be detected using grip force upon a steering wheel, even with a small time window of 2 s. Compared to other more established measures of stress, such as HRV, which

TABLE 2 | Summary of meta-analyses of correlations, for all $k = 20$.

	Measures in correlation	Pearson's r	Fisher's z	p	95% CI
1.	Ln(meanGF)–Ln(LF/HF)	0.1108	7.66	< 0.0001***	[0.0826, 0.1388]
2.	Ln(maxGF)–Ln(LF/HF)	0.0622	4.29	< 0.0001***	[0.0338, 0.0905]
3.	Ln(meanGF)—Ln(LF)	−0.0513	−3.53	0.0004***	[−0.0796, −0.0229]
4.	Ln(maxGF)–Ln(LF)	−0.002	−0.14	0.8921	[−0.0304, 0.0265]
5.	Ln(meanGF)–Ln(HR)	0.0293	2.02	0.0436*	[0.0008, 0.0577]
6.	Ln(maxGF)–Ln(HR)	0.0395	2.72	0.0065**	[0.0110, 0.0679]
	HRV HF negative values represent the inhibitory pattern of the parasympathetic system during stressful situations				
7.	Ln(meanGF)–Ln(HF)	−0.1631	−11.33	< 0.0001***	[−0.1907, −0.1353]
8.	Ln(maxGF)–Ln(HF)	−0.09	−6.21	< 0.0001***	[−0.1181, −0.0617]

*meanGF, mean grip force; maxGF, maximum grip force; sdGF, standard deviation of the grip force; LF, HF, and LF/HF, heart rate variability measures; HR, heart rate. Significance codes: 0 ≤ "***" < 0.001 < "**" < 0.01 < "*" < 0.05 < "." < 0.1 < " " ≤ 1.*

require much larger time windows, grip force's narrow window can enable a "real-time" assessment of the effect of stressful situations on the driver.

According to the summation of the current research findings, grip force may be considered as one of the measures of stress in mobile environments such as vehicles. By measuring the driver's stress level in "real time," various interventions can be employed to prevent calamities from occurring due to inferior human performance under stressful conditions. The usage of non-invasive measures of stress that do not interfere with the user's experience allows access to the information in the realistic environment of vehicles despite the limitations inherent in it.

Measuring stress in a vehicle is beneficial not only for human-controlled cars but also for self-driving vehicles. It is clear that, in such scenarios, the grip force measurement will not be on the steering wheel but at different grip points in the vehicle or on mobile devices held by the passengers, such as smartphones and tablets. Information about the passengers' stress levels may aid the vehicle's control system to adjust its conduct to minimize stress and thus achieve a better user experience. Moreover, by measuring grip force exerted on a surface of a non-operation means, stress measurement may reflect a purer indication of stress, without possible influences of task-conducting-related grip force.

The current research has some limitations. First, due to the difficulty to differentiate stress from related terms (Alsuraykh et al., 2019), it should be mentioned that the manipulation used in the current study (i.e., a STOP sign as a mandatory stopping position) may not have been experienced solely as stress by the participants. Second, this study examined a limited range of potential stress-eliciting driving events. Additional driving events such as lane crossing, overtaking, or driving in heavy traffic should also be evaluated in a similar manner to gain more comprehensive insights regarding the potential of driving incidents being used as stressors. This may also help clarify whether grip force was a result of stress elicited during braking.

The third limitation is the use of a homogeneous population, constituted of male students only, with a limited range of ages. This uniform sample limits this research's external validity, and further research should be conducted using more heterogenic samples. Finally, the "noisy" signal of the measured grip force (as illustrated in **Figure 4**) may interfere with the analysis of the state of the driver. Therefore, other data processing methods (e.g., fast Fourier transform) should be considered if real-time assessment is required (Zak et al., 2020).

CONCLUSION

The current research's primary purpose and contribution are to provide initial empirical evidence on the extent to which grip force may serve as an additional index of stress in driving tasks and its validation using HR and HRV measures. Heart measures were affected by braking, a finding which is consistent with the findings of previous studies and which establishes the assertion that braking events induce stress. Variations in grip force as an outcome of these stress-inducing braking events support its suitability for stress measurement in driving scenarios. The correlation of grip force and heart measures strengthens the statement that, similar to heart measures, grip force is an appropriate measure of stress.

The ability to identify a specific change in stress during a driving scenario using a non-invasive measurement tool which is transparent to the end user has the potential of introducing in-car just-in-time stress management interventions. It may also help develop a stress-adaptive car system that may adjust its conduct according to the driver's current level of stress. Future investigations may aid in describing the relationship of grip force and stress in driving as well as in other tasks.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ariel University Human Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YS with the guidance of MW and SS, conceptualized the study and chose the theoretical framework. YS, OM, TE, and TH designed and conducted the experiment, chose the methods, and collected the data. OM, TE, and TH obtained the ethics committee's approval. YS, OM, and TE analyzed the data. YS wrote the manuscript. MW, SS, OM, and TE read and revised the manuscript several times. All authors contributed to the article and approved the submitted version.

REFERENCES

Al-Fudail, M., and Mellar, H. (2008). Investigating teacher stress when using technology. *Comput. Educat.* 51, 1103–1110. doi: 10.1016/j.compedu.2007.11.004

Allen, A. P., Kennedy, P. J., Cryan, J. F., Dinan, T. G., and Clarke, G. (2014). Biological and psychological markers of stress in humans: focus on the Trier Social Stress Test. *Neurosci. Biobehav. Rev*. 38, 94–124. doi: 10.1016/j.neubiorev.2013.11.005

Alsuraykh, N. H., Wilson, M. L., Tennent, P., and Sharples, S. (2019). "How stress and mental workload are connected," in *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*, (New York: Association for Computing Machinery), 371–376.

Balters, S., Bernstein, M., and Paredes, P. E. (2019). "On-road stress analysis for in-car interventions during the commute," in *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, (New York: Association for Computing Machinery).

Botzer, A., Meyer, J., and Parmet, Y. (2016). Effects of cues in a binary categorization task on dual-task performance, mental workload, and effort. *J. Exp. Psychol.* 22:350. doi: 10.1037/xap0000095

Botzer, A., Sahar, Y., Wagner, M., and Elbaum, T. (2020). **Analyzing Individuals' Grip Force Over Short Intervals in a Joystick-Controlled Task with and without a Stress Manipulation. Manuscript submitted for publication*.

Brookhuis, K. A., and De Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. *Accid. Anal. Prev*. 42, 898–903. doi: 10.1016/j.aap.2009.06.001

Bruun, A. (2018). "It's not complicated: a study of non-specialists analyzing GSR sensor data to detect UX related events," in *Proceedings of the 10th Nordic Conference on Human-Computer Interaction*, (New York: Association for Computing Machinery), 170–183.

Cohen, J. (1992). "Things I have learned (so far)," in *Annual Convention of the American Psychological Association, 98th, Aug, 1990, Boston, MA, US; Presented at the aforementioned conference*, (Washington, DC: American Psychological Association).

Collet, C., and Musicant, O. (2019). Associating vehicles automation with drivers functional state assessment systems: a challenge for road safety in the future. *Front. Hum. Neurosci.* 13:131. doi: 10.3389/fnhum.2019.00131

Collet, C., Salvia, E., and Petit-Boulanger, C. (2014). Measuring workload with electrodermal activity during common braking actions. *Ergonomics* 57, 886–896. doi: 10.1080/00140139.2014.899627

Contrada, R. J., and Baum, A. eds (2011). *The Handbook of Stress Science: Biology, Psychology, and Health*. Berlin: Springer.

Daviaux, Y., Bonhomme, E., Ivers, H., de Sevin, É., Micoulaud-Franchi, J. A., Bioulac, S., and Altena, E. (2020). Event-Related Electrodermal Response to Stress: results From a Realistic Driving Simulator Scenario. *Hum. Fact.* 62, 138–151. doi: 10.1177/0018720819842779

Davies, M. N. O., and Underwood, G. (2000). Cognition and stress. *Encycl. Stress* 1, 478–483.

Dawson, M. E., Schell, A. M., and Filion, D. L. (2007). The electrodermal system. *Handb. Psychophysiol.* 2, 200–223.

De Rivecourt, M., Kuperus, M. N., Post, W. J., and Mulder, L. J. M. (2008). Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight. *Ergonomics*, 51, 1295–1319. doi: 10.1080/00140130802120267

Dehais, F., Lafont, A., Roy, R., and Fairclough, S. (2020). A Neuroergonomics Approach to Mental Workload, Engagement and Human Performance. *Front. Neurosci*. 14:268. doi: 10.3389/fnins.2020.00268

Dinges, D. F., Rider, R. L., Dorrian, J., McGlinchey, E. L., Rogers, N. L., Cizman, Z., and Metaxas, D. N. (2005). Optical computer recognition of facial expressions associated with stress induced by performance demands. *Aviat. Space Environ. Med*. 76, B172–B182.

Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., and Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proc. Natl. Acad. Sci*. 113, 2636–2641. doi: 10.1073/pnas.1513271113

Frasson, C., Brosseau, P. O., and Tran, T. H. D. (2014). "Virtual environment for monitoring emotional behaviour in driving," in *International Conference on Intelligent Tutoring Systems*. eds S. Trausan-Matu., K.E. Boyer., M. Crosby, K. Panourgia, (Cham: Springer), 75–83 doi: 10.1007/978-3-319-07221-0_10

Grimberg, E., Botzer, A., and Musicant, O. (2020). Smartphones vs. in-vehicle data acquisition systems as tools for naturalistic driving studies: a comparative review. *Safe. Sci.* 131:104917. doi: 10.1016/j.ssci.2020.104917

Hall, M., Vasko, R., Buysse, D., Ombao, H., Chen, Q., Cashmere, J. D., and Thayer, J. F. (2004). Acute stress affects heart rate variability during sleep. *Psychos. Med.* 66, 56–62. doi: 10.1097/01.psy.0000106884.58744.09

Hancock, P. A. (1989). A dynamic model of stress and sustained attention. *Hum. Fact.* 31, 519–537. doi: 10.1177/001872088903100503

Hancock, P. A., and Szalma, J. L. (Eds.). (2008). *Performance under stress*. Farnham: Ashgate Publishing Ltd.

Healey, J., and Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Trans. Intell. Transp. Syst.* 6, 156–166. doi: 10.1109/tits.2005.848368

Hendricks, D. L., Freedman, M., and Fell, J. C. (2001). *The Relative frequency of Unsafe Driving Acts in Serious Traffic Crashes (No. DOT-HS-809-206)*. Washington, DC: National Highway Traffic Safety Administration.

Hernandez, J., Paredes, P., Roseway, A., and Czerwinski, M. (2014). "Under Pressure: Sensing Stress of Computer Users," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, (New York: ACM), 51–60.

Hjortskov, N., Rissén, D., Blangsted, A. K., Fallentin, N., Lundberg, U., and Søgaard, K. (2004). The effect of mental stress on heart rate variability and blood pressure during computer work. *Eur. J. Appl. Physiol.* 92, 84–89. doi: 10.1007/s00421-004-1055-z

Hou, X., Liu, Y., Sourina, O., and Mueller-Wittig, W. (2015). "CogniMeter: EEG-based emotion, mental workload and stress visual monitoring," in *2015 International Conference on Cyberworlds (CW)*, (New York: IEEE), 153–160

Joosen, P., Exadaktylos, V., Taelman, J., and Berckmans, D. (2017). "The effect of Individual Stress Zones on car-racing performance," in *2017 IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, (New York: IEEE), 79–82.

Keogh, E., and Ratanamahatana, C. A. (2005). Exact indexing of dynamic time warping. *Knowl. inform. Syst.* 7, 358–386. doi: 10.1007/s10115-004-0154-9

Kirschbaum, C., and Hellhammer, D. H. (1994). Salivary cortisol in psychoneuroendocrine research: recent developments and applications. *Psychoneuroendocrinology*, 19, 313–333. doi: 10.1016/0306-4530(94)90013-2

Kristal-Boneh, E., Raifel, M., Froom, P., and Ribak, J. (1995). Heart rate variability in health and disease. *Scand. J. Work Environ. Health* 21, 85–95.

Kucera, P., Goldenberg, Z., and Kurca, E. (2004). Sympathetic skin response: review of the method and its clinical use. *Bratisl. Lekars. Listy* 105, 108–116.

Kudielka, B. M., Schommer, N. C., Hellhammer, D. H., and Kirschbaum, C. (2004). Acute HPA axis responses, heart rate, and mood changes to psychosocial stress

(TSST) in humans at different times of day. *Psychoneuroendocrinology*, 29, 983–992. doi: 10.1016/j.psyneuen.2003.08.009

Lazarus, R.S., and Folkman, S. (1984). *Stress, Appraisal and Coping*. Springer: New York.

Lee, Y. C., and Winston, F. K. (2016). Stress induction techniques in a driving simulator and reactions from newly licensed drivers. *Transp. Res. Part F* 42, 44–55. doi: 10.1016/j.trf.2016.06.019

Li, K., Rüdiger, H., and Ziemssen, T. (2019). Spectral analysis of heart rate variability: time window matters. *Front. Neurol.* 10:545. doi: 10.3389/fneur.2019.00545

Liao, W., Zhang, W., Zhu, Z., Ji, Q., and Gray, W. D. (2006). Toward a decision-theoretic framework for affect recognition and user assistance. *Int. J. Hum. Comput. Stud.* 64, 847–873. doi: 10.1016/j.ijhcs.2006.04.001

Litman, T. (2019). *Autonomous Vehicle Implementation Predictions*. Canada: Victoria Transport Policy Institute.

Liu, D., and Ulrich, M. (2014). *Listen to Your Heart: Stress Prediction Using Consumer Heart Rate Sensors*. URL: http://cs229.stanford.edu/proj2013/LiuUlrich-ListenToYourHeart-StressPredictionUsingConsumerHeartRateSensors.pdf.

Malik, M. (1996). Heart rate variability: standards of measurement, physiological interpretation, and clinical use: task force of the European Society of Cardiology and the North American Society for Pacing and Electrophysiology. *Anna. Noninvas. Electrocardiol.* 1, 151–181. doi: 10.1111/j.1542-474x.1996.tb00275.x

Mantilla, J., Oudre, L., Barrois, R., Vienne, A., and Ricard, D. (2017). "Template-DTW based on inertial signals: Preliminary results for step characterization," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2267-2270)*, (New York: IEEE).

Matthews, G. (2001). *A Transactional Model of Driver Stress*. Florida: University of Central Florida, 133-163

Matthews, G., Reinerman-Jones, L. E., Barber, D. J., and Abich IV, J. (2015). The psychometrics of mental workload: multiple measures are sensitive but divergent. *Hum. Fact.* 57, 125–143. doi: 10.1177/0018720814539505

May, J. G., Kennedy, R. S., Williams, M. C., Dunlap, W. P., and Brannan, J. R. (1990). Eye movement indices of mental workload. *Acta Psychol.* 75, 75–89. doi: 10.1016/0001-6918(90)90067-p

McCraty, R., Atkinson, M., Tiller, W. A., Rein, G., and Watkins, A. D. (1995). The effects of emotions on short-term power spectrum analysis of heart rate variability. *Am. J. Cardiol.* 76, 1089–1093. doi: 10.1016/s0002-9149(99)80309-9

McEwen, B. S. (2000). Definitions and concepts of stress. *Encycl. Stress* 3, 508–509.

McGrath, J. E. (1976). Stress and behavior in organizations. *Handb. Industr. Organiz. Psychol.* 1351:1396.

Meyer, K., Rasch, T., and Schnotz, W. (2010). Effects of animation's speed of presentation on perceptual processing and learning. *Learn. Instruct.* 20, 136–145. doi: 10.1016/j.learninstruc.2009.02.016

Milstein, N., and Gordon, I. (2020). Validating Measures of Electrodermal Activity and Heart Rate Variability Derived From the Empatica E4 Utilized in Research Settings That Involve Interactive Dyadic States. *Front. Behav. Neurosci.* 14:148.doi: 10.3389/fnbeh.2020.00148

Min, B. C., Chung, S. C., Park, S. J., Kim, C. J., Sim, M. K., and Sakamoto, K. (2002). Autonomic responses of young passengers contingent to the speed and driving mode of a vehicle. *Int. J. Industr. Ergon.* 29, 187–198. doi: 10.1016/s0169-8141(01)00059-2

Model 50–2303-00 (2014). *Model 50–2303-00*. Gahannah, OH: Mindware Technologies.

Mühlbacher-Karrer, S., Mosa, A. H., Faller, L. M., Ali, M., Hamid, R., Zangl, H., and Kyamakya, K. (2017). A driver state detection system—Combining a capacitive hand detection sensor with physiological sensors. *IEEE Trans. Instrum. Measur.* 66, 624–636. doi: 10.1109/tim.2016.2640458

Mulder, G., and Moray, N. (1979). *Mental Workload: Its Theory and Measurement*. New York: Springer.

Mulder, L. J. M., Dijksterhuis, C., Stuiver, A., and De Waard, D. (2009). Cardiovascular state changes during performance of a simulated ambulance dispatchers' task: potential use for adaptive support. *Appl. Ergon.* 40, 965–977. doi: 10.1016/j.apergo.2009.01.009

Nickel, P., and Nachreiner, F. (2003). Sensitivity and diagnosticity of the 0.1-Hz component of heart rate variability as an indicator of mental workload. *Hum. Fact.* 45, 575–590. doi: 10.1518/hfes.45.4.575.27094

Palinko, O., Kun, A. L., Shyrokov, A., and Heeman, P. (2010). "Estimating cognitive load using remote eye tracking in a driving simulator," in *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, (New York: ACM), 41-144.

Peat, J., and Barton, B. (2014). *Medical Statistics: A Guide to SPSS, Data Analysis and Critical Appraisal*. Hoboken: John Wiley & Sons.

Pedrotti, M., Mirzaei, M. A., Tedesco, A., Chardonnet, J. R., Mérienne, F., Benedetto, S., and Baccino, T. (2014). Automatic stress classification with pupil diameter analysis. *Int. J. Hum. Comput. Interact.* 30, 220–236. doi: 10.1080/10447318.2013.848320

Prasolenko, O., Burko, D., and Halkin, A. (2017). Galvanic skin response as a estimation method of the driver's emotional state. *Am. J. Sci. Engin. Technol.* 2, 50–56.

Qu, W., Zhang, Q., Zhao, W., Zhang, K., and Ge, Y. (2016). Validation of the driver stress inventory in china: relationship with dangerous driving behaviors. *Accid. Anal. Prev.* 87, 50–58. doi: 10.1016/j.aap.2015.11.019

Rastgoo, M. N., Nakisa, B., Rakotonirainy, A., Chandran, V., and Tjondronegoro, D. (2019). A critical review of proactive detection of driver stress levels based on multimodal measurements. *ACM Comput. Surv.* 51:88.

Rezazadeh, I. M., Firoozabadi, M., Hu, H., and Golpayegani, S. M. (2012). Co-adaptive and affective human-machine interface for improving training performances of virtual myoelectric forearm prosthesis. *IEEE Trans. Affect. Comput.* 3, 285–297. doi: 10.1109/t-affc.2012.3

Rohleder, N., Wolf, J. M., Maldonado, E. F., and Kirschbaum, C. (2006). The psychosocial stress-induced increase in salivary alpha-amylase is independent of saliva flowrate. *Psychophysiology* 43, 645–652 doi: 10.1111/j.1469-8986.2006.00457.x

Rosenthal, R. (1991). *Meta-Analytic Procedures for Social Research*, Canada: Sage Publications.

Ross, T., and Burnett, G. (2001). Evaluating the human–machine interface to vehicle navigation systems as an example of ubiquitous computing. *Int. J. Hum. Comput. Stud.* 55, 661–674. doi: 10.1006/ijhc.2001.0495

Saxena, N. (2017). *Modelling the Effect of the Number of Stop-&-gos on the Route Choice Behaviour of Car Drivers*, Ph D. thesis, (Sydney: The University of New South Wales).

Segerstrom, S. C., and Miller, G. E. (2004). Psychological stress and the human immune system: a meta-analytic study of 30 years of inquiry. *Psychol. Bull.* 130:601. doi: 10.1037/0033-2909.130.4.601

Senin, P. (2008). *Dynamic time warping algorithm review*. Honolulu, USA: Information and Computer Science Department University, 855, 1–23.

Staal, M. A. (2004). *Stress, Cognition, and Human Performance: A Literature Review and Conceptual Framework*. California: Ames Research Center.

Sugiono, S., Widhayanuriyawan, D., Andriani, D. P., and Prasetya, R. P. (2019). Investigating the Influence of Distance between Cars on the Driver Psychophysiology during Braking Using Eeg: a Case Study on Driving in Indonesia. *Acta Neuropsychol.* 17, 329–339. doi: 10.5604/01.3001.0013.6189

Sztajzel, J. (2004). Heart rate variability: a noninvasive electrocardiographic method to measure the autonomic nervous system. *Swiss Med. Weekly* 134, 514–522.

Tomer, E., Lupu, T., Golan, L., Wagner, M., and Braw, Y. (2018). Eye tracking as a mean to detect feigned cognitive impairment in the Word Memory Test. *Appl. Neuropsychol.* 27, 49–61 doi: 10.1080/23279095.2018.1480483

Van Roon, A. M., Mulder, L. J., Althaus, M., and Mulder, G. (2004). Introducing a baroreflex model for studying cardiovascular effects of mental workload. *Psychophysiology* 41, 961–981. doi: 10.1111/j.1469-8986.2004.00251.x

Vuksanović, V., and Gal, V. (2007). Heart rate variability in mental stress aloud. *Med. Engin. Phys.* 29, 344–349. doi: 10.1016/j.medengphy.2006.05.011

Wagner, M., Sahar, Y., Elbaum, T., Botzer, A., and Berliner, E. (2015). Grip Force as a Measure of Stress in Aviation. *Int. J. Aviat. Psychol.* 25, 157–170. doi: 10.1080/10508414.2015.1162632

Wahlström, J., Hagberg, M., Johnson, P., Svensson, J., and Rempel, D. (2002). Influence of time pressure and verbal provocation on physiological and psychological reactions during work with a computer mouse. *Eur. J. Appl. Physiol.* 87, 257–263. doi: 10.1007/s00421-002-0611-7

Wang, X., Ding, X., Su, S., Li, Z., Riese, H., Thayer, J. F., and Snieder, H. (2009). Genetic influences on heart rate variability at rest and during stress. *Psychophysiology* 46, 458–465. doi: 10.1111/j.1469-8986.2009.00793.x

Wiberg, H., Nilsson, E., Lindén, P., Svanberg, B., and Poom, L. (2015). Physiological responses related to moderate mental load during car driving in field conditions. *Biol. Psychol*. 108, 115–125. doi: 10.1016/j.biopsycho.2015.03.017

Yamaguchi, M., Wakasugi, J., and Sakakima, J. (2006). "Evaluation of driver stress using biomarker in motor-vehicle driving simulator." in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, (New York: IEEE), 1834-1837.

Yerkes, R. M., and Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *J. Comp. Neurol. Psychol*. 18, 459–482. doi: 10.1002/cne.920180503

Zak, Y., Parmet, Y., and Oron-Gilad, T. (2020). "Subjective Workload Assessment Technique (SWAT) in Real Time: Affordable Methodology to Continuously Assess Human Operators' Workload,"in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, (New York: IEEE), 2687-2694.

Zhai, J., and Barreto, A. (2006). Stress detection in computer users based on digital signal processing of noninvasive physiological variables. *Proc. IEEE Annu. Int. Conf*. 2006, 1355–1358.

Zontone, P., Affanni, A., Bernardini, R., Del Linz, L., Piras, A., and Rinaldo, R. (2020). Stress evaluation in simulated autonomous and manual driving through the analysis of skin potential response and electrocardiogram signals. *Sensors* 20:2494. doi: 10.3390/s20092494

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX A—DEFINING GRIP FORCE'S TIME WINDOW

To define the proper time window in the calculation of the grip force measures, we explored the following time windows: 2, 5, 10, 20, and 30 s. We have conducted the meta-analysis procedure described before (as detailed in the "Data Analysis" and "Results" sections) for each of these time windows.

For each combination of the three Ln-transformed grip force transformations (mean, maximum, and standard deviation) and the four Ln-transformed heart rate measures (HRV LF, HRV HF, HRV LF/HF ratio, and HR), we have calculated the Pearson correlation coefficients.

The following plot (see **Figure A1**) represents the confidence intervals of the Pearson correlation coefficients for each of these time windows.

As indicated in the plot, for the 2-s time window, the lower end of the confidence interval is larger than zero. From this, it seems that the choice of the 2-s time window is reasonable.

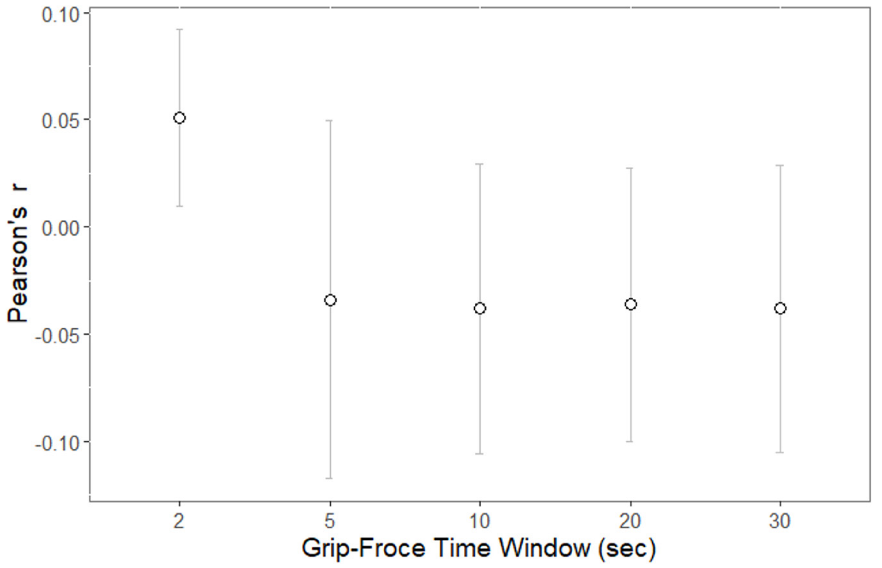


FIGURE A1 | Confidence intervals of the Pearson correlation coefficient (*r*) for meta-analyses of Ln-transformed grip force's measures (mean, maximum, and standard deviation) with Ln-transformed heart rate measures (HRV LF, HRV HF, HRV LF/HF ratio, and heart rate) for each grip force's time window (2, 5, 10, 20, and 30 s).