A PSYCHOPHYSIOLOGICAL INSIGHT INTO DRIVER STATE DURING HIGHLY AUTOMATED DRIVING

by

Vishnu Radhakrishnan

Submitted in accordance with the requirements
for the degree of Doctor of Philosophy

The University of Leeds
Institute for Transport Studies
June 2022

Intellectual Property and Publications

The candidate confirms that the work submitted is his/her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Four publications have been produced from research that was undertaken as part of this thesis. Each publication is listed below with a full reference and details of its location within this thesis. In the case of all publications listed the candidate was solely responsible for the production of the content with named authors providing support through review and modification only. The work in Chapters 2 to 5 of the thesis has appeared in publications as follows:

The work in Chapter 2 of the thesis is a working paper, appearing as follows:

Radhakrishnan, V., Solernou, A. & Merat, N. (May 2022). Motion artefact detection and correction for EDA signals in ambulatory environments. University of Leeds.

The candidate has contributed significantly to conceptualisation, data analysis, formulating algorithms, writing codes and validating the data. The candidate wrote the paper under the guidance of *AS* and *NM*. *AS* assisted with signal processing, idea formulation and debugging the software codes.

The work in Chapter 3 of the thesis has appeared in publication as follows:

Radhakrishnan, V., Merat, N., Louw, T., Lenné, M. G., Romano, R., Paschalidis, E., Hajiseyedjavadi, F., Wei, C., Boer, E. R. (2020). Measuring drivers' physiological response to different vehicle controllers in Highly Automated Driving (HAD): opportunities for establishing real-time values of driver discomfort. Published in Information, 11(8), 390. DOI: https://doi.org/10.3390/info11080390.

The candidate contributed substantially to the design of this study, which was conducted as a part of the HumanDrive project. The candidate collected, analysed, and interpreted the data, and wrote the paper under the guidance of *NM*, *TL* and *ML. RR*, *EP*, *FH*, *CW* and *EB* were part of the HumanDrive project, and contributed to the design, development and procurement of funding for the project, as well as assisted in data collection. *NM*, *TL* reviewed the manuscript. The candidate, *NM* and *TL* gave final approval for the version submitted for publication.

The work in Chapter 4 of the thesis has appeared in publication as follows:

Radhakrishnan, V., Merat, N., Louw, T., Gonçalves, R. C., Torrao, G., Lyu, W., Puente Guillen, P., & Lenné, M. G. (2022). *Physiological indicators of driver workload during car-following scenarios and takeovers in highly automated driving*. Published in: Transportation Research Part F: Traffic Psychology and Behaviour, 87, 149–163. https://doi.org/10.1016/J.TRF.2022.04.002.

The candidate contributed substantially to design of this study, which was conducted as a part of the L3Pilot project. The candidate collected, analysed, and interpreted the data, and wrote the paper under the guidance of *NM*, *TL* and *ML. RG and GT* assisted with the data collection and participant recruitment. *PG*, *WL* were part of the L3Pilot project, and contributed to procurement of funding, assistance in data collection and proof-reading. *NM*, *TL* and *ML* reviewed the manuscript. The candidate, *NM* and *TL* gave final approval for the version submitted for publication.

The work in Chapter 5 of the thesis has been submitted for publication as follows:

Radhakrishnan, V., Louw, T., Gonçalves, R. C., Torrao, G., Lenné, M. G. & Merat, N. (Under review). Using pupillometry and gaze-based metrics as objective indicators of drivers' mental workload during Highly Automated Driving (HAD). Transportation Research Part F: Traffic Psychology and Behaviour.

The candidate contributed substantially to the design of this study, which was conducted as a part of the L3Pilot project. The candidate collected, analysed, and interpreted the data, and wrote the paper under the guidance of *NM*, *TL* and *ML. RG and GT* assisted with the data collection and participant recruitment. *NM*, *TL* and *ML* reviewed the manuscript. The candidate, *NM* and *TL* gave final approval for the version submitted for publication.

Acknowledgements

I would like to thank my principal supervisor Natasha Merat for her excellent mentorship and guidance through the course of this project. I am truly grateful for the help, support and encouragement I received at every step of this project. I would like to thank my second supervisor Tyron Louw, for the friendship, guidance, support, and helping me to bridge the gap between engineering and psychology. To Michael Lenne, my industry supervisor, for his unwitting humour, cricket references, brilliant mind, incredible guidance and support throughout my project. Overall, I could not have asked for a better team to guide me through this incredible journey. I would also like to thank Seeing Machines Inc., for co-funding my PhD and providing me with this incredible opportunity. I would also like to thank the incredible members of the UoLDS team, and all the members of the HFS research group, for all their help and support.

I would like to thank my family, especially my dad, mom and sister, for supporting me emotionally, financially and encouraging me in every way possible, to fulfil my dreams.

I was lucky to be surrounded by some incredible friends. I would like to thank Alberto, for being a truly amazing selfless and wonderful friend, especially travelling to Amsterdam at 5am, to return my residency card, just so that I could get back to the UK for my MSc Viva. Without you, I would not be standing where I am now. Alice and Nathan, you were there for me unconditionally even in the darkest of times, and were always rooting for me. Nick, from football chats, to great food and amazing wines, you have been a rock. Davide and Patrick, you were my brothers at work, we started the journey together, and have been through all the ups and downs together, thank you for always being there. Mikey and Sophie, we have been a family, and without your support and encouragement, I don't think I could have gotten through the pandemic period. To Arjun and Lalu, for being such amazing and supportive friends. Finally, Sappho and Hobbs, your unconditional love, the licks, snuggles, and constant emotional support, I could not have done this without you both.

I would like to thank Debbie for proof-reading my introduction chapter, providing me with great feedback on grammar and punctuations, and in the process, making me a better writer. Thank you for all your help and support, and I am truly indebted to you. I would also like to thank Royal Literary Fund (RLF) for providing me with an incredible opportunity to improve my academic writing.

Last but not the least, to Janet, you have been there from day one. Always cheering me, taking my side, helping me, supporting me and just being the wonderful

person that you are. I truly could not have done this without you. Thank you for your enduring love and affirmation, at every step of this journey.

Abstract

The aim of this research was to investigate and validate the usage of physiological measures as an objective indicator of driver state in dynamic driving environments, and understand if such a methodology can be used to measure driver discomfort, and high workload. The work addressed questions relating to: (i) detecting and removing motion artefacts from electrodermal activity (EDA) signals in dynamic driving environments; (ii) primary factors contributing to driver discomfort during automation, measured in terms of their physiological state; (iii) understanding changes in drivers' workload levels at different stages of automation, as indicated by electrocardiogram (ECG) and EDA-based measures and; (iv) how drivers' attentional demands and workload levels are affected at different stages of automation, measured using eye tracking-based metrics. A series of experiments were developed to manipulate drivers' discomfort and workload levels. The analysis around driver discomfort focused on automated driving, whereas drivers' workload levels were investigated during automation, and during resumption of control from automation, in a series of car-following scenarios. Our results indicated that phasic EDA was able to pick up discomfort experienced by the driver during automation, and correlated to drivers' subjective ratings of discomfort. Narrower roads, higher resultant acceleration forces and how the automated vehicle negotiated different road geometries all influenced driver discomfort. We observed that drivers' workload levels were captured by ECG and EDA-based signals, with phasic component of EDA signal being more sensitive to short term variations in driver workload. Similar results were observed in drivers' pupil diameter values, as well as subjective ratings of workload. Factors such as engagement in a non-driving related task (NDRT), presence of a lead vehicle while maintaining a short time headway, and takeovers, all seemed to increase drivers' workload levels. Future work can build on this research by incorporating sensor fusion of ECG and EDA-based data, along with eye tracking, to help improve the accuracy and capabilities of future driver state monitoring systems.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

A number of abbreviations and acronyms are used throughout the this thesis. Some of common terms used in this thesis are expanded below:

ACF	Automated car-following
ADAS	Advanced driver assistance systems
ANOVA	Analysis of variance
ANS	Autonomic nervous system
DDT	Dynamic driving task
ECG	Electrocardiogram
EDA	Electrodermal Activity
EDR	ECG-derived respiration rate
FFT	Fast Fourier Transforms
HAD	Highly automated driving
HR	Heart Rate
HRV	Heart Rate Variability
MCF	Manual car-following
NDRT	Non-driving related tasks
ODD	Operational design domain
OOTL	Out of the loop
PSNS	Parasympathetic nervous system
RMSSD	Root mean squared of successive differences
SCL	Skin conductance level
SCR	Skin conductance response
SDLP	Standard deviation of lateral position
SNS	Sympathetic nervous system
THW	Time headway
TOR	Takeover request

To the most amazing human I've known: ***Vinaya Nair (1950-2017)***

I miss you every day of my life Peramma.

1 GENERAL INTRODUCTION

1.1 Introduction

In the past decade, there has been an increase in the implementation of Advanced Driving Assistance Systems (ADAS), which support the driver in lateral and/or longitudinal control of the vehicle. Some examples of such ADAS features include adaptive cruise control (ACC) or lane keeping assistance systems. Manufacturers, technology companies as well as researchers, have been working towards developing higher levels of vehicle automation. The main motivation for implementing higher levels of vehicle automation in the market is its hypothesised provision of increased safety (ERTRAC, 2017), mobility (Trommer et al., 2016), accessibility (Alessandrini et al., 2015), efficiency (Steck et al., 2018) and comfort (Beggiato et al., 2019).

We have observed the emergence of advanced SAE Level 2 (see section 1.2.1; SAE International, 2021) driving automation systems such as Tesla's Autopilot (Tesla, 2021), General Motor's Super Cruise (General-Motors, 2021) and Ford's BlueCruise (Ford-Motors, 2021). GM's Super Cruise and Ford's BlueCruise offer hands-free driving, within certain pre-defined environments, such as on motorways and under certain speed limits. However, all such systems require the driver to constantly pay attention to the drive, and supervise the operation of the automated system. Recently, Honda released the first SAE L3 conditional driving automation system called Traffic Jam Assist, where drivers are not required to monitor the drive under certain conditions ,such as driving on the motorway in congested traffic, for speeds of up to 50 kmph (Honda EU, 2021).

With automation, the role of a driver changes, from that of an active user to a passive observer (Parasuraman & Riley, 1997). As the driver takes on a supervisory role, their set of tasks required for the safe operation of the vehicle changes, and calls for a better understanding of how this role change can affect performance and safety of the vehicle. Automating the driving task, especially as seen in currently available SAE L2 and SAE L3 automated systems, would still require the driver to resume manual control of the vehicle, if and when prompted by the automated system, or in

case of automation failures. While automation of the driving task can increase driver safety, assuming the human driver is replaced by an infallible machine, it can also result in behavioural changes of the driver, that can negatively affect their driving performance and safety of the vehicle (Heikoop et al., 2016). Driving performance after a transition of control from the automated system to the driver, which is generally quantified in terms of the lateral and longitudinal control of the vehicle with respect to road boundaries and safety margins, using measures such as standard deviation of lane position (SDLP), lateral acceleration or braking (Gold et al., 2018), is not only affected by driver state during the transition itself, but also when automation is engaged (Parasuraman et al., 2008; Zeeb et al., 2016). Therefore, monitoring drivers' state, which, according to Gonçalves & Bengler (2015), is defined as a set of physical and mental conditions that can affect drivers' capabilities at the driving task in a specific instant, can help improve performance and safety by helping the system to assist or provide additional support to the driver. Certain driver states, such as discomfort or high workload, can negatively affect wider acceptance, as well as performance and safety of the vehicle, for when they resume manual control (Gonçalves & Bengler, 2015; Hartwich et al., 2018; Zeeb et al., 2016).

Increased driver comfort is one of the selling arguments for automated driving systems (Carsten & Martens, 2019). During manual driving, apart from the vehicle specifications around noise, vibration and harshness, the driver modulates comfort by having full control of the vehicle, and in turn, controlling aspects that affect comfort such as safety margins or acceleration and jerk forces (Wertheim & Hogema, 1997). However, in automated driving, the control of the driving task is taken away from the driver, and this loss of control, as well as the inability to anticipate the actions of the vehicle, can increase driver discomfort (Cahour, 2008). A limited number of studies, such as the KomfoPilot study by Beggiato et al. (2019), have explored factors influencing driver discomfort during automated driving. It is also important to monitor their comfort levels in real-time, to enable the automated system to adapt and provide the driver with a more comfortable ride. Therefore, further research is warranted on understanding how different driving styles, incorporated by the automated system to negotiate different road environments, affects driver comfort. This can improve driving experience and wider acceptance and implementation of such features.

Automation can result in the driver being in an out of the loop (OOTL) state (Endsley & Kiris, 1995), which refers to a state where drivers are not monitoring the driving environment, and may or may not be in physical control of the vehicle, and can result in driving performance decrements once they resume manual control (Merat et al., 2018). This OOTL state is further aggravated when drivers engage in other non-driving related tasks (NDRT) during automation, the likelihood of which increases with an increase in the level of automation (Carsten et al., 2012), and potentially increases drivers' workload levels (Merat et al., 2012). Studies have shown that both a sudden and unexpected high workload (overload), such as negotiating a takeover in complex traffic situation with high traffic density (Radlmayr et al., 2014), and low workload (underload) due to the diminished task demands brought about by automation (Young & Stanton, 2002a), can result in performance decrements if and when the driver has to resume manual control of the vehicle. Additionally, driver workload prior to a transition of control to manual driving, such as high workload due to engagement in a demanding NDRT during automation, can affect driving performance at a later stage, for example, during a transition of control to manual driving (Parasuraman et al., 2008; Zeeb et al., 2016). Therefore, it is of value to have real-time assessment of drivers' workload levels prior to (that is, when automation is activated) and during transitions, to help guide the automated system into providing improved support and assistance to the driver, for example, warning the driver of dangerous underload or overload conditions (Merat et al., 2012).

Continuous monitoring of drivers' state, during automation and during transition of control to the driver, is therefore extremely important for ensuring safe operation of the vehicle. Current automation systems such as GM's Super Cruise or Ford's BlueCruise use camera-based eye tracking to monitor drivers' attentional state (that is, whether or not they are monitoring the drive), objectively and in real-time. State-of-the-art driver monitoring systems that use fixed-base eye trackers such as the Seeing Machines Guardian Backup-driver Monitoring System (BdMS; Seeing Machines Inc, 2022) are able to monitor drivers' attention to the road in real-time, even when the driver is wearing a mask or sunglasses. These systems work in a wide range of lighting conditions and are non-invasive. However, eye tracking-based metrics, such as pupil diameter and direction of gaze, have their limitations, and as a set of stand-alone metrics, are incapable of objectively and accurately quantifying driver states that can

negatively affect driving performance, such as discomfort or high workload. For example, pupil diameter, which is indicative of drivers' workload levels, is greatly affected by lighting conditions or brightness levels in the driving environment (Mathôt, 2018). Additionally, if the drivers' eyes are occluded from the view of the fixed-base eye tracker, it cannot provide insight into drivers' attentional state. Therefore, further investigation is required into the validity of additional physiological measures to complement eye tracking based metrics, for reliable assessment of driver state.

Physiological signals such as electrodermal activity (EDA) and electrocardiogram (ECG) have been used in the past as an indicator of drivers' state (Foy & Chapman, 2018; Mehler et al., 2012; Patel et al., 2011). However, the majority of the studies that have used these physiological signals have been conducted in manual driving, with less availability of results from automated driving. Moreover, the majority of studies on physiological signals have been conducted in laboratory environments (Cho et al., 2017; Hjortskov et al., 2004; Patel et al., 2011; Shimomura et al., 2008), and, within the driving context, many have used fixed-base driving simulators (Beggiato et al., 2019; Foy & Chapman, 2018; Mehler et al., 2009). Physiological signals of EDA and ECG are extremely sensitive to motion and noise artefacts. A better understanding of how such signals are affected by motion artefacts in a dynamic driving environment, and whether they can be successfully removed, is required. Additionally, validation of whether such signals can be used to correctly assess driver states such as discomfort or high workload, within automated driving context, is required.

This thesis aimed to address this issue by investigating whether physiological measures of EDA and ECG can be used to complement eye tracking-based metrics in driver state monitoring. This can aid in understanding and measuring driver states, such as discomfort or workload, that can negatively affect wider acceptance of automation features, as well as driving performance after a transition, in real-time. This research also investigated how driver state was affected during different automated driving scenarios, in a full motion-based simulator environment. Given that this thesis focuses on automated driving, the following section considers how the role of a driver evolves during automation of the driving task.

1.2 History of automation and driving

In general, automation refers to “operating or acting, or self-regulating, independently, without human intervention” (Nof, 2009, p. 14). According to Sheridan & Parasuraman (2005), automation involves: “the mechanisation and integration of the sensing of environmental variables (by artificial sensors); data processing and decision making (by computers); mechanical action (by motors or devices that apply forces on the environment) and/or; information action by communication of processed information to people” (p. 90). The word automation originates from the Greek word automatos, meaning acting by itself, or by its own will, or spontaneously (Nof, 2009). Williams (2009) suggested that automation is a way for the human to extend the capabilities of their tools and machines. Towards the late 1950s, automation was viewed as a substitution for human efforts and decisions, by a combination of mechanical, pneumatic, hydraulic, electrical and electronic devices (Nof, 2009).

Within the transportation sector, automation generally refers to the transfer of responsibilities of some or all of the control tasks and sub-tasks, from the human operator to an automated system. An early example of automation in transportation was seen in the aviation industry. Innovative technologies, such as fly-by-wire controls which replaced manual flight controls with electronic interfaces, and Flight Management System, are used by pilots to assist with flight planning, navigation, performance management and flight-progress monitoring (Banks et al., 2019; Nadine B. Sarter & Woods, 1992). The aviation industry has pioneered the technological and engineering developments in automation within the transportation industry (Banks et al., 2019), and Stanton & Marsden (1996) warn that driving automation could yield similar human factors challenges as some of those observed in the aviation industry. These include pilots affected by dangerous underload or overload conditions during automation, or over-reliance on the automated system without fully understanding its limitations (Nadine B. Sarter & Woods, 1992; Wilson, 2002; Young et al., 2007). However, it should be noted that pilots and road vehicle drivers also have their differences, owing to differences in vehicle type, technical systems involved, and the environment of operation. Additionally, pilots are generally professional operators, who undergo rigorous training and are generally required to update their licenses annually, which is not the case for road vehicle drivers.

Within the driving automation context, the earliest development of an automation feature was observed when General Motors presented the Firebird II in 1956, a concept car guided by electrical wires placed on the road, to cruise in highways of the future (General Motors, 1956). While the initial focus was on guided automated vehicles, the advancement of computers and computer vision in the 1980s saw the invention of a Mercedes Benz van, called VaMoRs, with a vision-based guidance system, that could drive in an automated manner on empty roads without traffic, for speeds of up to 60 kmph (Dickmanns & Zapp, 1987). In 1997, the California PATH consortium successfully demonstrated an 8-car platoon that was manoeuvred under fully automated control (Rajamani et al., 2000). In the early 2000s, research and development of automated vehicles gained further traction, when the Defence Advanced Research Projects Agency (DARPA) introduced the Grand Challenge of 2004, which required automated vehicles to compete against each other in an off-road course. DARPA has organised similar events since, that have resulted in autonomous vehicles and technologies being tested on public roads (Banks et al., 2019; Louw, 2017).

While the initial focus in automated vehicle (AV) research was on the engineering and technological developments required for automating some or all of the driving tasks and sub-tasks, more recently, research into the human factors issues surrounding automation is gaining momentum, with EU-funded projects such as the H2020 Automated Driving Applications & Technologies for Intelligent Vehicles (AdaptIVe; Etemad et al., 2017), L3Pilot project (Etemad, 2021), and the Hi-Drive project (Etemad, 2022) having more focused work on the human factors challenges of vehicle automation, along with the design, implementation and evaluation of different automation functions, for different levels of automation. Other national projects such as the Innovate UK/CCAV-funded HumanDrive project (Woolridge & Chan-Pensley, 2020) or the German Federal Ministry of Education and Research (BMBF) funded KomfoPilot project (Beggiato et al., 2018), have also investigated the human factors challenges surrounding vehicle automation. Today, driving automation systems can be seen in vehicles such as Tesla (Tesla, 2021), or General Motors Cadillac (General-Motors, 2021), where limited automation features are available, under certain operational design domains. The features of the automated system, its operational environment, and the responsibilities of the driver during automated driving, are

dependent on the level of automation, which is reviewed in greater detail in the next sub-section.

1.2.1 Levels of automation

Over the years, there have been several attempts to provide a concise taxonomy and definition for different levels or degrees of automation. It is important to have a rigorous and unanimously agreed upon classification for different levels of automation, so that the manufacturers and drivers alike are clear about the responsibilities and limitations of such features. The most widely adopted classification is that proposed by SAE International (SAE International, 2021), where automated driving is classified into 6 levels, from Level 0 to Level 5, in increasing order of automation features, as listed below. According to SAE International (2021), "*driving automation system*" refers to any automation system or feature, from Level 1 to Level 5, capable of performing part or all of the dynamic driving task (DDT) on a sustained basis, depending on its operational design domain (ODD). DDT constitutes "*all the real-time operational and tactical functions that are required to operate a vehicle in on-road traffic*", including lateral and longitudinal control, situational awareness, object and event response detection, execution, manoeuvre planning and enhancing conspicuity, via signalling and gestures (SAE International, 2021). ODD refers to a manufacturer-specified set of "*operating conditions under which a given driving automation system or feature thereof is specifically designed to function, including, but not limited to, environmental, geographical, and time-of-day restrictions, and/or the requisite presence or absence of certain traffic or roadway characteristics*" (SAE International, 2021). In contrast, an Automated Driving System (ADS), which generally refers to automation Levels 3 to 5, is defined as the software and hardware that is capable of performing a DDT on a sustained basis, irrespective of its ODD (SAE International, 2021).

The different levels of automation, as defined by SAE International (2021), is explained below:

- **Level 0 or No Driving Automation:** All driving tasks are done solely by the human driver without any assistance in acceleration, braking or lateral steering control. This could include warning systems. Here the role of automation is limited to warnings or support, such as momentary intervention.

- **Level 1 or Driver Assistance:** Either longitudinal or lateral control of the vehicle is continuously performed by the driver. The system assists in other tasks or either lateral or longitudinal tasks. Adaptive Cruise Control (ACC) and Lane Keeping Assist (LKA) are examples of Level 1 automation systems.
- **Level 2 (SAE L2) or Partial Driving Automation:** System takes over control of both lateral and longitudinal tasks (steering, acceleration and braking). But the driver is required to permanently monitor the vehicle. Volvo's Pilot Assist, Tesla's Autopilot and GM's SuperCruise are some examples of vehicles with SAE L2 automated system available in the market (Brooke, 2020).
- **Level 3 (SAE L3) or Conditional Driving Automation:** System takes over both lateral and longitudinal control of the vehicle under specific ODD. Generally, the drivers are not expected to be monitoring the system continuously if it falls under the ODD of the SAE L3 automation as specified by the manufacturer. But when the ODD limits are exceeded/about to be exceeded or there is a DDT performance relevant failure in the ADS system, a timely request is given by the system to the driver (DDT fall back-ready user) to take over the control of the vehicle. The system disengages at an appropriate time after issuing a request to intervene or does so immediately upon driver-initiated take over request. Honda Traffic Jam Assist is the first ADS to offer SAE L3 features (Honda EU, 2021). The driver is not required to continuously monitor the driver task and can focus on other activities supported by an on-board infotainment system, subjected to legal constraints within the respective country (SAE International, 2021).
- **Level 4 (SAE L4) or High Driving Automation**: The system performs all aspects of DDT and the DDT fall-back (transfer of control scenario) within the specified ODD, but without the expectation that a user will respond to a request to intervene or takeover. This means that the system can give a timely request for the driver to take over control if a DDT performance relevant system failure occurs or the ODD limits, as specified by the system, are exceeded. In case the driver does not respond to the takeover request, the ADS is will transition automatically into a minimal risk condition. Minimum risk condition is the "*condition to which a user or ADS may bring the vehicle after performing DDT fall-back in order to reduce the risk of a crash when a given trip cannot or should not be completed*" (SAE International, 2021). It is assessed independently by the ADS without any input

from the driver in SAE L4 systems. The driver is at the liberty of taking over control of DDT during takeover request or can also request the ADS to disengage and take over control. Currently, there are no SAE L4 enabled vehicles on the market.

- **Level 5 (SAE L5) for Full Driving Automation:** The system performs all aspects of DDT and DDT fall-back, in a sustained and unconditional manner (i.e. the ODD is unlimited) without any expectation that the driver will respond to a takeover request. Takeover requests happen only in case of a DDT performance-relevant system failure, but the ADS will transition automatically into minimal risk condition if the driver does not respond to the takeover request. The driver is at the liberty of taking over control of DDT during takeover request or can also request the ADS to disengage and take over control. It will be quite some time before an actual SAE L5 vehicle would enter the market realistically, although companies like Waymo are in the process of testing advanced SAE L4 and SAE L5 ADS equipped vehicles (Waymo, 2018).

In this thesis, I primarily focus on human factors challenges surrounding SAE L2 and SAE L3 automation systems, in which drivers are required to resume manual control of the vehicle, when prompted by a takeover request (SAE International, 2021). However, drivers can be affected by a host of factors, leading up to the transition of control, that can result in diminished performance and safety upon resumption of manual control, including, but not limited to, drivers' attention to the driving task (Louw, Kuo, et al., 2019), their workload levels (Müller et al., 2021), and discomfort and stress levels (Beggiato et al., 2019). In order to perform a safe transition of control, drivers are likely required to have good awareness of the immediate driving environment (Endsley, 1995) and of the automated system's state (Beggiato et al., 2015), to be better equipped at maintaining adequate safety margins from obstacles and other road users.

1.2.2 Transitions of control

Within the automated driving context, the terms *transitions* or *takeovers* have been used interchangeably to refer to the activation or deactivation of an automated driving function (Gold et al., 2013), a change in automation state (Merat et al., 2014), or, according to Louw (2017), the process and period where the transfer of responsibility of some or all aspects of a driving task, between a human driver and a

driving automation system, occurs. In this thesis, I use the definition for takeovers as that suggested by Louw (2017), mentioned above.

Martens et al. (2008) proposed an ontology that classifies transitions into four cases, with the three underlining factors of: 1) who has the control of the vehicle at the beginning of transition; 2) who gets the control of the vehicle during/after the transition and; 3) who initiates the transition, defining each transition case, as shown in **Table 1.1**.

Table 1.1. Different types of transitions (Martens et al., 2008)

	$\underline{O}_i \rightarrow A$	$\underline{O} \rightarrow A_i$	$O_i \leftarrow \underline{A}$	$O \leftarrow \underline{A}_i$
Who has "it"?	Operator	Operator	Automation	Automation
Who should get "it"?	Automation	Automation	Operator	Operator
Who initiates transition?	Operator	Automation	Operator	Automation

Table 1.1 shows the different directions of the transfer of control, with '*i*' being the initiator of the transition. It has been assumed that the driver has the 'required readiness state' whenever the transition of control request is initiated by the operator.

In this thesis, only automation-initiated transitions of control, from ADS to the human driver, was considered, as indicated in the last column of **Table 1.1**. For the safe operation of the vehicle after the transition, the driver is required to be "*ready*" to take over control of the driving task from the automated system. This "*driver readiness*" depends on their physical (such as hand position, head posture and foot position) and psychological states (such as stress, attention or workload levels), and requires good communication and collaboration between the human driver and the automated system (Carsten & Martens, 2019).

1.2.3 Human-automation interaction

Given that we are still a long way from achieving full autonomy in driving, for the foreseeable future, there will be a wide variety of vehicles offering different levels of automation features (Carsten & Martens, 2019). Therefore, rather than considering human driver and the automated system as independent and mutually exclusive in performing the DDT, they are more likely to communicate and collaborate in order to function as a joint cognitive system, to ensure the safe operation of the vehicle, as

depicted in **Figure 1.1** (Carsten & Martens, 2019; Flemisch et al., 2008). The term "*Human-automation interaction*" encompasses both implicit and explicit communication between the human driver and the automated system, and the communication interface used in such situations is called a Human-Machine Interface (HMI).

Successful human-automation interaction should include two-way communication between the automated system and the driver, and should not just be limited to the automated system being able to communicate its capabilities and limitations to the driver (Carsten & Martens, 2019), but also, for the system to be able to monitor and identify drivers' capabilities and limitations in performing the DDT, even if this communication is implicit, and if required, provide appropriate mitigation strategies to ensure safe transition of control (Mioch et al., 2017).

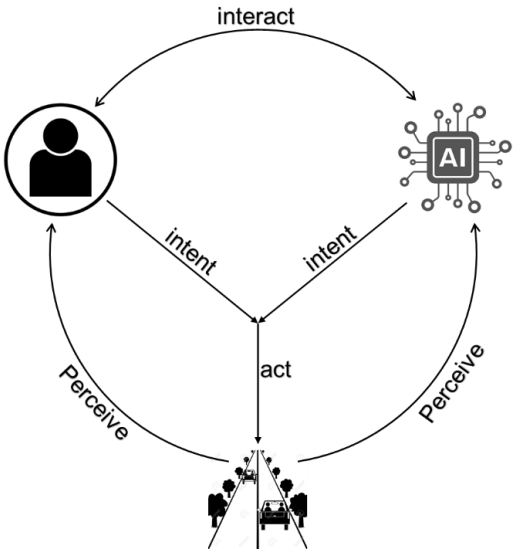

Figure 1.1. Model depicting collaborative control in automated driving (adapted from Flemisch et al., 2008)

As the automation level increases, it is likely that the automated system is involved in making decisions at strategic, tactical and control levels of the driving task (Carsten & Martens, 2019; Kircher et al., 2014; Michon, 1985). However, this can also result in the drivers being OOTL and engaging in other NDRTs (Carsten et al., 2012; Merat et al., 2014). Therefore, the automated system should be able to monitor drivers' state in an objective manner, and ensure drivers' physical and mental readiness, if and when the driver has to resume control, for safe operation of the vehicle (Mioch et

al., 2017). In case the driver is not ready, HMI-based responses/feedback can be useful in bringing the driver back into the loop (Beller et al., 2013; Lorenz et al., 2014). For example, aiding the driver to perform the takeover task using interfaces such as visual and augmented reality (AR)-based stimuli and messages, highlighting the path to follow, on the windscreen, while performing a lane change or takeover manoeuvre, has shown promising results (Lorenz et al., 2014). Haptic, auditory and speech-based human-machine interfaces have also been used to re-direct the driver's attention and assist them with the DDT (Mulder, Abbink, & Boer, 2012; Naujoks, Forster, Wiedemann, & Neukum, 2016; Pitts, Williams, Wellings, & Attridge, 2009). In order to identify the appropriate mitigation strategy to bring drivers back into the driving control loop, it is important that the automated system is able to monitor and correctly identify drivers' current state. For example, the automated system should provide different mitigation strategies to bring a fatigued or sleepy driver back into driving control loop, compared to an inattentive or distracted driver. A distracted driver would be required to refocus their attention to the key aspects of the driving task with the aid of auditory or visual HMI to refocus his attention, to successfully resume control of the driving task. However, in case the driver is sleepy, the vehicle should be safely brought to a halt, or direct the driver to pull over, as it would be harder to get back into the driving control look from a sleepy state.

1.3 Driver state

The driving task requires decision making at control (such as steering and longitudinal control), tactical (such as passing obstacles or negotiating roundabouts) and strategic (such as planning the route) levels. Automation substitutes some of these tasks (such as on control level when using ACC), and provides additional comfort to the driver. However, currently available systems require the driver to monitor the drive (SAE L2 vehicles) and/or resume control of the vehicle when prompted. To successfully monitor the drive or resume manual control of the vehicle, the drivers need to have appropriate readiness levels, and not be affected by driver state such as distraction, or high workload. The psychological state of the driver, which Gonçalves & Bengler (2015) define as a set of psychological "conditions" that affect the driver in a specific instant, affects their information processing ability, and in turn, performance (output) at the driving task.

Stanton & Young (2000) proposed a functional model, involving eight psychological constructs that could critically impact driver behaviour during automation: Situational Awareness, Mental Workload, Mental Model, Feedback, Locus of Control, Stress, Task Demands and Trust. However, Stanton & Young's model was a hypothetical model, and according to Heikoop et al. (2016), did not provide insights into the nature of the interrelationships between the different psychological constructs, that is whether the relationship was causal or co-relational, as well as whether the effect was positive or negative. It is important to know the interrelationships between the different psychological constructs, as drivers can be affected by multiple states, at any given instant. Therefore, Heikoop et al. (2016) adapted a descriptive model from Stanton & Young's model, based on past observations in scientific literature, describing the different interrelationships between different psychological constructs, as shown in **Figure 1.2**.

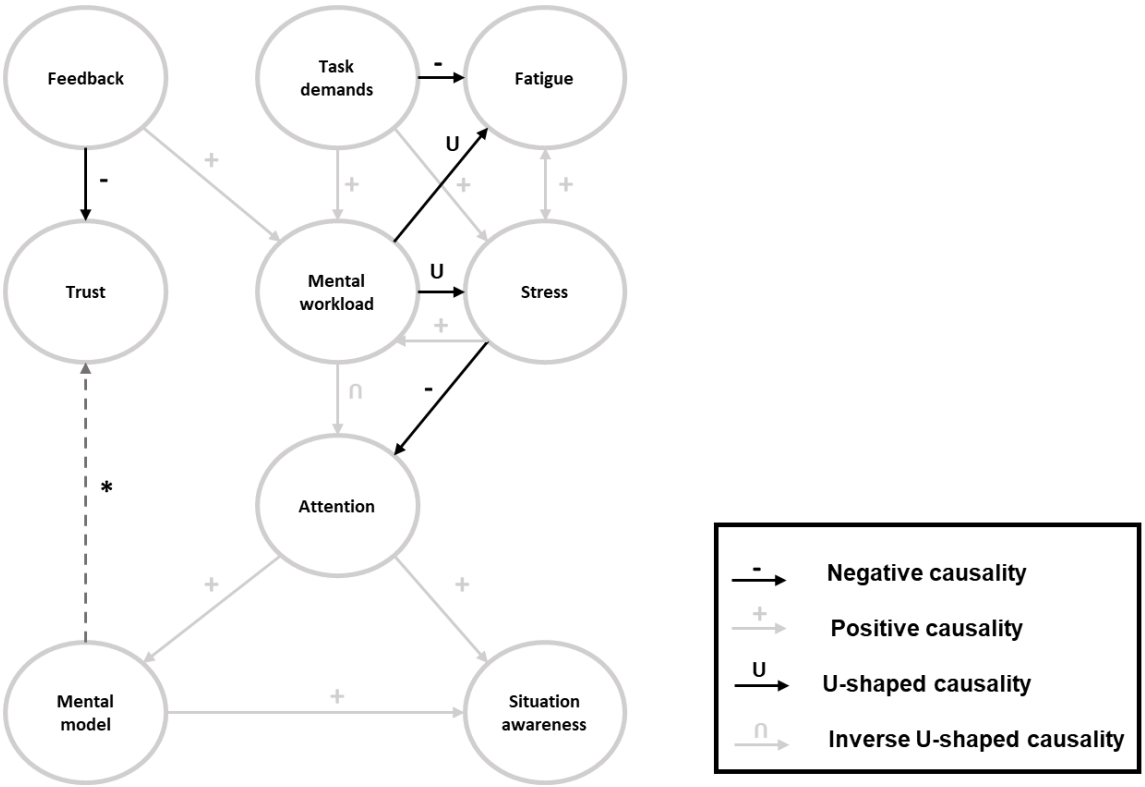

Figure 1.2. Relationship between different psychological constructs during automation (redrawn from Heikoop et al., 2016). * indicates how mental models can recalibrate drivers' trust.

Heikoop's model consisted of the following psychological constructs:

i. *Feedback*: Feedback in this model refers to automation-induced feedback that a driver receives, such as visual or auditory signals. In **Figure 1.2**, feedback is shown to have a negative causal relationship with trust. However, the studies that Heikoop et al. (2016) reviewed to derive this relationship between feedback and trust, are probably studies on improper feedback, resulting in decreasing drivers' trust in the system, with improper feedback. For example, these feedbacks could be around critical takeover events, where the automation initiated takeovers with not enough time budget for the driver to safely resume control of the vehicle, and thereby diminishing their trust in the system. Feedback is also indicated to have a positive causality to workload, which is likely due to drivers having to process additional information provided by the automated system, thereby increasing their workload levels.

ii. *Trust*: Within the driving automation context, Körber et al. (2018) defined trust as "the attitude of a user to be willing to be vulnerable to the actions of an automation based on the expectation that it will perform a particular action important to the user, irrespective of the ability to monitor or to intervene" (p. 19). While not explicitly indicated in **Figure 1.2**, Heikoop et al. (2016) observed in some of the studies they reviewed that trust has a negative causal relationship with stress, situational awareness and attention. That is, as stress, situational awareness and attention increases, drivers' trust decreases.

iii. *Task demand*: According to APA Dictionary of Psychology (2022), task demand is "the effect of a task's characteristics, including its divisibility and difficulty, on the procedures that an individual or group can use to complete the task." There are different functions and processes embedded in a task. The difficulty of the task could lie in the demand imposed by these set of functions and processes required to sucessfully perform the task. In **Figure 1.2**, task demand is shown to have a positive causality to both stress and workload, indicating that a highly demanding task can increase drivers' stress and workload levels. Task demand is also shown to have a negative causality to fatigue. However, this is not always the case, as increased task demand can result in task-related fatigue (May & Baldwin, 2009), which the authors seemed to have overlooked in their model. Task demand is further explored in section 1.3.3.3 of this thesis.

iv. *Fatigue*: Fatigue is a multidimensional construct that has been challenging to define, for researchers (Brown, 1994; Desmond & Hancock, 2001). According to Brown (1994), fatigue can be defined as the subjectively experienced disinclination to perform a task at hand, that can impair human performance and efficiency. Fatigue is shown to have a positive causality with stress, in **Figure 1.2**.

v. *Stress*: According to Lazarus (1966), stress is defined as "a relationship between the person and the environment that is appraised as personally significant and as taxing or exceeding resources for coping". Increase in stress can have a negative effect on driving performance (Matthews & Desmond, 2001). Heikoop et al.'s model suggests that stress has a positive causality to fatigue and workload, and a negative causality to attention.

vi. *Mental Workload*: In driving context, workload describes the relationship between and cognitive resources demanded by a task, and those resources available to be supplied for the task, by the driver (Parasuraman et al., 2008). Heikoop et al. (2016) suggested that mental workload has a U-shaped causality to fatigue and stress, and an inverse U-shaped causality to attention. That is, both high and low workload levels can result in increased fatigue and stress, as well as diminished attention. Optimal workload levels can result in diminished stress and fatigue, and increase attention to the driving task. Workload forms one of the central psychological constructs explored in this thesis, and is reviewed in detail in section 1.3.3.

vii. *Attention*: Drivers' attention to the driving task and driving environment can increase their situational awareness, as well as enable the drivers to create more accurate mental models of the driving task, as indicated by the positive causality to both situational awareness and mental models, in **Figure 1.2**.

viii. *Situational awareness*: According to Endsley (1995), situational awareness "is the perception of elements in the environment, with a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (p. 36). A high degree of situational awareness can result in drivers being less likely to be in an OOTL state during automation (Louw, 2017; Merat et al., 2018), resulting in better performance at the driving task, when required to resume manual control of the vehicle.

ix. *Mental models*: Heikoop et al. (2016) suggested that mental models, which is the dynamic representation of the world (Johnson-Laird, 1980; D. A. Norman, 1983),

has a positive causality to situational awareness as seen in **Figure 1.2**, and can recalibrate drivers' trust towards the automation over time.

An interesting observation, made by Heikoop et al. (2016), is the intermediary relationship existing between task demand, workload, attention and situational awareness, as well as that between feedback and situational awareness. Based on the scientific studies they reviewed, the authors noted that rather than having a direct relationship between constructs such as task demand and situational awareness, or workload and situational awareness, task demand has a positive causality to workload. That is, an increase in task demand likely increases drivers' mental workload, and mental workload is linked to drivers' attention levels with an inverse U-shaped causality, meaning both high and low workload levels result in diminished attention. Finally, attention to the road environment leads to an increase in situational awareness.

The complex nature of the interrelationship between different psychological constructs or states can be observed in Heikoop et al. (2016)'s model (see **Figure 1.2**). Additionally, Mental Workload appears to be a central construct in the model, and is directly affected by: Feedback, Task Demand, and Stress. In their model, mental workload also directly affects drivers' attention, which in turn affects mental models and situational awareness, and thereby, affecting driving performance and safety of the vehicle (Parasuraman et al., 2008). A notable omission in this model is driver comfort. It is likely that the conceptual similarities between lack of comfort, that is, discomfort, and related concepts of stress, and mental workload (Beggiato et al., 2019), might have resulted in its omission from Heikoop et al. (2016)'s model. However, driving comfort plays a key role in determining drivers' trust and wider acceptance of such features (Carsten & Martens, 2019; F. Walker, 2021), and therefore, its importance in driver state monitoring should not be undermined.

During automation, especially in currently available SAE L2 and SAE L3 automated driving systems (SAE International, 2021), the automation takes over the decision making at control-level tasks, such as lateral and longitudinal control of the vehicle. However, this can affect different drivers differently. For example, experienced drivers' are likely to use skill-based behaviour (that is drawn from past experience) to do control level tasks, requiring little or no decision making (Hale et al., 1990).

Automating such tasks can reduce boredom and monotony, and increase comfort to experienced drivers (Kircher et al., 2014). However, novice drivers are likely to use knowledge-based behaviour from their memory, to do control-level tasks. This requires information processing, and can result in increased workload. Automating control-level tasks can reduce the workload on a novice driver. This example illustrates how driving experience and subjective differences between drivers can result in drivers experiencing changes to different driver states such as comfort or workload, from similar external factors/stimuli.

This thesis focuses on human factors challenges surrounding drivers' workload and comfort levels, which is expanded in the next subsection, and whether they can be objectively captured, on a moment-to-moment basis, using physiological signals.

1.3.1 Automation and its effect on driver state

While technological advancements have resulted in developing higher-level automation features, they become redundant if drivers are reluctant to adapt or use such features. Such automation features are hypothesised to improve comfort, safety and driving experience (ERTRAC, 2017). Research has indicated that ensuring driving comfort can help improve the trust, wider acceptance and implementation of automation features (ERTRAC, 2017; Molnar et al., 2018). Siebert, Oehl, Höger, & Pfister (2013) argued that driving comfort determines the acceptance of automation features, with high comfort levels leading to better acceptance of such features, and thus, improving the safety, with the assumption that the automated system is infallible. However, within the automated driving domain, there is limited research that has focused on monitoring drivers' comfort levels objectively and in a continuous manner, and understanding which factors cause driver discomfort.

The second challenge is around ensuring optimal driving performance, and providing safety, especially when drivers have to resume manual control of the vehicle, after a transition. Until automation is able to retain full control of the DDT, without any form of human intervention, human drivers are an important part of the driver-automation eco-system. As such, drivers are required to maintain appropriate readiness levels, as dictated by the automation level and related ODD constraints. However, with an increase in automation level, drivers are more likely to engage in NDRTs (Carsten et al., 2012). Reduction in responsibilities for the driving task

provides the driver with additional time to do non-driving related activities, likely reduce workload from the driving task, and potentially increase driver comfort. Although, from a safety perspective, it can result in diminished situational awareness, and exacerbate the OOTL effect by taking the driver further away from the driving control loop (Merat et al., 2014). Additionally, it can potentially increase their overall workload levels, depending on the task demand posed by the NDRT, and thereby reducing their readiness levels, and negatively affecting driving performance, for when they have to resume manual control of the vehicle. This warrants further research into monitoring drivers' workload level objectively, to inform the system of dangerous underload or overload situations, in order to provide better support to the driver when they are required to resume manual control of the vehicle.

The next two sub-sections of this thesis expands on driver states of comfort and mental workload and how these affect driving performance and the safety of automated driving systems.

1.3.2 Comfort

Comfort is a key aspect of human life, and well-being, with people striving to attain higher levels of comfort throughout their lives (Slater, 1986). However, there is no universal and unanimously agreed upon definition of comfort. Given the subjective nature of comfort, it is extremely challenging to derive a quantitative definition of comfort (Slater, 1985). From a qualitative perspective, Slater (1986) defined comfort as *"a pleasant state of physiological, psychological and physical harmony between human being and the environment"* (p. 158). In his definition, Slater referred to physiological comfort as the human body's ability to maintain life, psychological comfort as the mind's ability to keep itself functioning satisfactorily without external help, and physical comfort as that which relates to the effect of the external environment on the human body (Slater, 1986).

In driving context, and especially automated driving, Beggiato et al. (2019) defined comfort as "as a subjective, pleasant state of relaxation resulting from confidence in safe vehicle operation, which is achieved by the removal or absence of uneasiness and distress" (p. 446). Beggiato et al. (2019) further suggested that lack of comfort, i.e. discomfort, shares similarities and overlaps with other related concepts such as stress, workload, anxiety or motion sickness. Given that comfort generally

refers to a relaxed, unaroused state of well-being, both Slater (1985) and Siebert et al. (2013) suggest that it is easier to measure discomfort rather than comfort, as signs of discomfort are more well-defined and pronounced. Therefore, this thesis will partly focus on understanding and measuring discomfort, as it is easier to define and quantify discomfort when compared to comfort. While traditionally, discomfort has been measured subjectively using comfort scales, physiological metrics have also been used as an objective indicator of discomfort in driving (Beggiato et al., 2019; Hartwich et al., 2018; Radhakrishnan et al., 2020), as seen in section 1.4 of this thesis. In the next two sub-sections, I will expand on different factors that affect discomfort in the manual driving, as well as in the automated driving context.

1.3.2.1 Discomfort and driving

Richards, Jacobson, & Kuhlthau (1978) suggested that comfort is a key variable in user acceptance of transportation systems, with comfort being one of the most direct psychological correlates of ride quality, and also, passenger satisfaction and willingness to use the transportation system again. In the mid-70s and 80s, engineers and researchers alike gave emphasis on physical and physiological elements of comfort, including reducing noise, vibration and harshness (NVH) in automobiles, to improve ride quality and passenger comfort (Bryan et al., 1978; Richards et al., 1978). This included reduction of engine noise and noise from other moving parts of the vehicle, isolating noise due to drag and wind, usage of advanced materials inside the passenger cabin to isolate unwanted noise, reducing the vibrational and jerk forces inside the cabin and, using improved shock absorber and suspension systems (Bein et al., 2012; Heißing & Ersoy, 2011). In addition to NVH, researchers have also suggested that thermal comfort can be improved by maintaining ambient temperature inside the passenger cabin, and driver/passenger comfort in vehicles can be enhanced by improving air quality and seat ergonomics (De Looze et al., 2003; Gameiro da Silva, 2002). From a human factors standpoint, Summala (2007) proposed a four factor classification for a driver to be in a “comfort zone”, based on safety margins (safety margins kept from road edges, potential hazards or other vehicles), vehicle-road system (road geometry, road environment, velocity and acceleration), rule-following (obeying traffic laws) and satisfactory progress of a trip (meeting one’s expectations from the pace or progress of travel).

Motion sickness is considered to be another key factor affecting discomfort. Motion sickness is generally caused by differences between actual motion as detected mechanically by the vestibular system, and motion as detected or perceived by the visual system or kinaesthetic input (Golding, 2016). For example, when a passenger is reading a book while travelling in a car, their vestibular system informs the brain that they are in a moving vehicle, whereas their eyes, which are focused on the stationary book in their hand, inform their brain otherwise. Treisman (1977) proposed the "toxin-detector hypothesis", where motion-sickness is considered to be an evolutionary mechanism, and that the brain has evolved to detect derangement of expected patterns of vestibular, visual and kinaesthetic information, as central nervous system malfunction. Accompanying symptoms such nausea experienced during motion-sickness is the body's apparent self-defence mechanism to get rid of such "neurotoxins". This has been later confirmed based on experimental evidence, as the most plausible underlying mechanism for motion sickness (Money & Cheung, 1983).

Within the driving context, discomfort and motion sickness is generally associated with longitudinal acceleration forces such as those experienced in multiple sharp braking manoeuvres (Vogel, Kohlhaas, & von Baumgarten, 1982), or high lateral or centripetal forces such as those associated with negotiating a sharp turn at high velocities (Golding, 2016; Wada et al., 2012). The different factors affecting driver discomfort during manual driving is shown in the **Figure 1.3** below.

During manual driving, the driver is less likely to get motion sickness, compared to a rear-seat passenger, as they have a full view of the road ahead, thereby resulting in lower disassociation between visual and vestibular cues. Additionally, the ability to predict the movement of the vehicle by being in control of the driving task, can lead to better coherence between visual and vestibular cues. However, as the role of the driver changes from that of an active participant in manual driving to a passive observer during HAD, it is likely that the driver engages in NDRTs (Carsten et al., 2012), which can increase the likelihood of motion-sickness (Diels & Bos, 2016). This role change in the driver and transfer of control authority to the automated system can constitute additional discomfort to the driver, warranting further investigation into discomfort-inducing scenarios within a similar experimental design framework, between automated and manual driving, which is explored in the next sub-section.

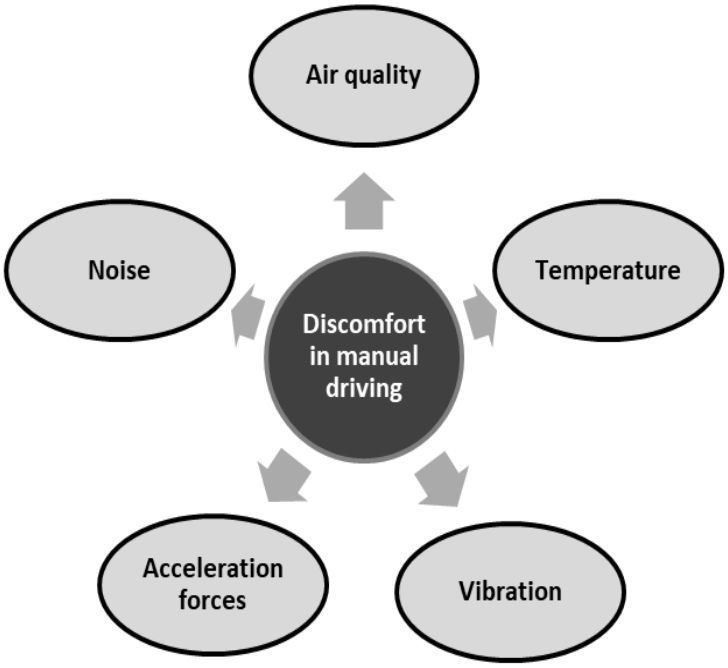

Figure 1.3. Main factors affecting driver discomfort during manual driving

1.3.2.2 Discomfort during automated driving

As noted by Summala (2007) and Siebert et al. (2013), safety margins, or rather, apparent safety margins, play a key role in reducing driver discomfort. In manual driving, drivers are likely to modulate their comfort levels, by having full control of the vehicle and its operation, and keeping adequate safety margins that they deem safe and comfortable. However, during automated driving, their comfort levels are modulated by how the driving automation system performs the driving task. Therefore, it is important to convey the safe operation of the vehicle to the driver, from a human perspective, even if the driving automation system is performing in a safe manner considering all static and dynamic risk elements, as it might not be apparent to the driver. If the driver judges the ADS to be unsafe, it can negatively affect their trust as well as acceptance of it. Maintaining a safe distance from hazards, other vehicles and road users, smooth execution of manoeuvres such as lane-changing, and minimising jerk can all help reduce driver discomfort during highly automated driving (HAD; Elbanhawi et al., 2015). Highly automated driving generally refers to automated systems where both lateral and longitudinal control of the vehicle is performed by the automated system, when engaged.

The acceleration and jerk forces experienced in the vehicle can affect driver discomfort. Higher acceleration and jerk forces (both magnitude and frequency) can increase driver discomfort, with studies showing that drivers tend to keep their lateral and longitudinal acceleration values under 2 m/s^2 (Bae et al., 2019; Bosetti et al., 2014). For HAD, it should be noted that the driver is not in control, and it is likely that they might have slightly lower safety and comfort thresholds for acceleration and jerk forces during HAD, compared to their own manual driving. Eriksson & Svensson (2015) suggested an acceleration and jerk threshold of 2 m/s^2 and 0.9 m/s^3, for minimising motion sickness and ensuring a comfortable driving experience during HAD.

Elbanhawi et al. (2015) suggested that naturalness of the drive could be another causal factor for driver discomfort in AVs. Since AVs are still in prototype/testing phase, most individuals are unaware of its usage and functions. As such, it is likely that drivers would base their expectations for a comfortable drive, based on their understanding of driver comfort in manual driving, or as a passenger in public transport systems. Therefore, it is of value to understand what particular features of an AV's driving manoeuvres are likely to constitute a comfortable riding experience for a human user. For example, it is to be seen whether the driver prefers a familiar set of driving manoeuvres akin to a human driver (naturalness of the drive), or rather, a more "machine-like" driving behaviour, which entails rigorously adhering to the road-centre, and not cutting corners on curve to minimise lateral forces and jerk. **Figure 1.4** below details factors affecting driver discomfort in HAD scenarios.

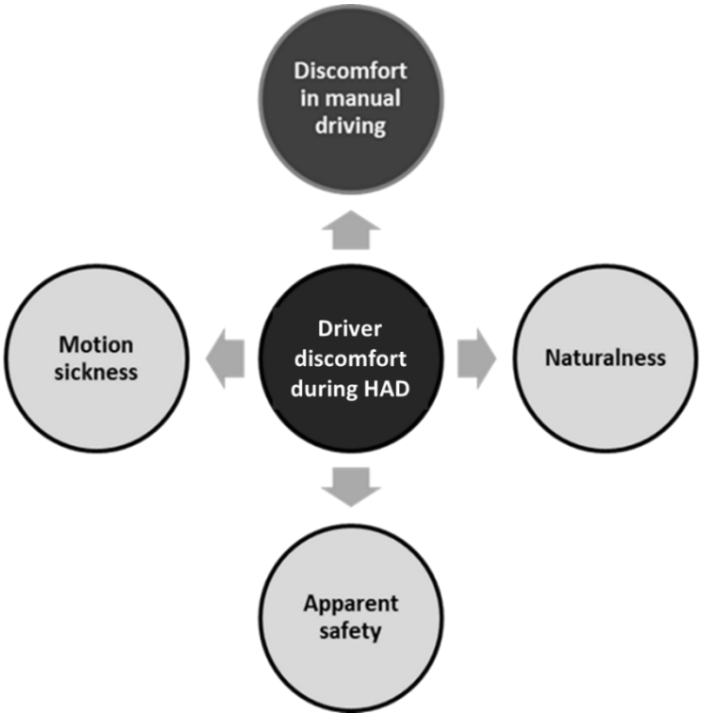

Figure 1.4. Factors affecting driver discomfort in HAD (adapted from Elbanhawi et al., 2015)

Given the subjective nature of comfort, most studies on comfort resorted to subjective measures to quantify comfort, such as subjective questionnaires (Thakurta et al., 1995) or comfort scales (Myers et al., 2008). However, given that driver discomfort can have a negative impact on drivers' trust as well as wider acceptance of such AV features, it is crucial that the system is able to incorporate a discomfort detection system, that is minimally invasive, to monitor driver discomfort objectively and on a moment to moment basis. This can help the AV to adapt its driving style to ensure that the driver is relaxed and at ease. Physiological measures such as ECG and EDA-based metrics have been used in the past to assess driver state, in manual driving (Foy & Chapman, 2018). However, there is limited research on using physiological metrics as a measure of discomfort, especially during highly automated driving. A detailed review of physiological measures is provided in Section 1.4. In this thesis, I investigate how driver discomfort is affected by the driving style of the automated vehicle, including acceleration and jerk forces, human-like driving profiles, safety-margins to obstacles and road boundaries, as well as different road geometries, and whether driver discomfort can be measured using physiological metrics. While lack of safety-margins has been observed to increase driver discomfort, it can also

affect drivers' workload levels. For example, when a vehicle maintains unsafe headway distance from a lead vehicle, it has been shown to increase drivers' vigilance, as well as workload levels (Liu et al., 2019). Additionally, as mentioned in section 1.3, depending on individual characteristics of the driver, such as driving experience, automating the driving task can affect both comfort and workload levels. The next sub-section introduces the concept of mental workload, and why it is crucial to understand and monitor drivers' workload levels during different stages of HAD.

1.3.3 Mental workload

Since the 1960s, researchers have used the concept of mental workload to explain an operator's task capabilities and limitations related to performance at one or more tasks. de Waard (1996) defined mental workload as the proportion of information processing capacity that is utilised for task performance. That is, mental workload describes the relationship between cognitive resources demanded by a task, and that which is available to the operator (Parasuraman et al., 2008). Workload has been used to describe human error and performance decrements seen in airline pilots in the aviation industry (S. G. Hart & Bortolussi, 1984; Roscoe, 1978). In the early days, pilots were required to monitor and process information from various sensors, during the flight, and make corrections if necessary, which could increase their workload levels. However, when part of pilots' tasks and responsibilities became automated with innovative technologies such as fly-by-wire systems, it was assumed that this would lead to a decrease in pilots' workload levels. However, automation posed several new human factors challenges, such as boredom due to monotony during periods of automation (Norman & Orlady, 1989). Conversely, automating tasks during periods of high pilot work rates, such as during take-off and landing, can contribute to an increase in cognitive load, due to an ever increasing amount of information processing required (Billings, 1991; Norman & Orlady, 1989).

Within the driving domain, workload is considered to be one of the contributing factors affecting the performance and safety of a vehicle (de Waard, 1996; Hancock & Caird, 1993; Parasuraman et al., 2008). The term workload in itself does not suggest to a particular driver state, but rather, a driver could have different workload levels (such as high workload, low workload, or optimal workload) at different points in time, and this can be collectively considered as a set of driver states, affecting the driver at

that particular instant. How a driver processes information, from the driving environment, affects their workload. Therefore, in order to better understand the concept of mental workload within the driving context, the following sub-section will focus on models of the driving task.

1.3.3.1 Models of the driving task

Rasmussens' (1979, 1983) model on human behaviour and performance classified human behaviour into three categories: skill-based, rule-based and knowledge-based behaviour.

Skill-based behaviour sits at the lowest level, and consists of well-learned procedures that are undertaken using sensory-motor skills, in an automated manner and without conscious control. At the intermediate level, rule-based behaviour is controlled by a *stored rule* that may be derived from past experiences. The boundary between skill-based and rule-based behaviour is not quite distinct. However, most rule-based behaviours involve conscious know-how of the situation, and of the rules being followed. At the highest level, knowledge-based behaviour involves conscious problem solving and decision making, based on analysis of the environment, to align with the overall aims and goals of the person, representing conscious control. However, Rasmussen’s taxonomy of human behaviour does not provide a dynamic insight into task hierarchy and corresponding performance.

Within the driving context, Michon (1985) proposed a three level hierarchy of cognitive control. The strategic/planning level consists of the overall planning of the trip, including route selection, based on the objectives/goals for the trip and evaluation of the costs and risks involved. The manoeuvring/tactical level consists of negotiating common driving situations such as curves, intersections, and overtaking, as well as maintaining safety margins from road edges and other road users, without violating the goals and objectives set in the strategic level, and is generally executed in a matter of seconds. The operational/control level involves immediate control of vehicle inputs, such as acceleration or braking, and is performed in an automatic action/response pattern in a matter of milliseconds. The strategic level of driving is largely based on past experiences and requires little or no new information, and is usually pre-planned, whereas tactical and control levels are data-driven and based on the immediate driving

environment and generally require real-time information and feedback (Ranney, 1994).

Hale et al. (1990) incorporated Michon's control hierarchy in driving with Rasmussen's taxonomy on human behaviour, as shown in **Table 1.2**. Driving experience plays an important role in this taxonomy. For experienced drivers, the majority of the driving tasks follow a diagonal pattern in the table, from the top-left to the bottom-right corner, whereas it is mostly on the top-right for novice drivers. That is, an experienced driver is likely to follow skill-based behaviour at control level, rule-based behaviour at manoeuvre/tactical level and knowledge-based behaviour at planning level. The exceptions reflect the differences in driving experience (experienced vs novice driver) and familiarity of the driving situations. For example, for a novice driver, a control-task, such as changing the gears, might require knowledge-based execution. As discussed earlier, in section 1.3.1, automating control level tasks can affect the driver states differently, for different drivers, as their driving behaviour (whether it is knowledge-based, skill-based or rule-based) could vary, for the same control level task, based on their driving experience. However, familiarity of the situation, such as travelling between home and work, or negotiating familiar intersections, can make use of skill-based behaviour for successful task execution, suggesting that automaticity (without conscious control) can be observed on all three levels of cognitive control. Hale et al. (1990) also suggest that drivers perform more homogenously in skill-based and rule-based levels, than in knowledge-based behaviour, and most human error-based theories have the greatest difficulty in predicting error if the behaviour is knowledge-based.

Table 1.2. Classification of driving task, based on Michon's control hierarchy and Rasmussen's skill-rule-knowledge based behaviour (adapted from Hale et al., 1990, p. 1383).

	Planning	Manoeuvre	Control
Knowledge	Navigating in a new area	Controlling skid	Novice on first lesson
Rule	Route selection between familiar routes	Passing other vehicles	Driving an unfamiliar vehicle

Skill	Travel between home and work	Negotiating familiar intersections	Cornering

Deconstructing driver tasks into sub-tasks helps to demonstrate how changes in driver-vehicle-environment can affect driving performance. Processing of information (sensory input) from the different agents in the driving task, such as the vehicle (speed), or the environment (surrounding traffic, road geometry, etc.) by the driver, is crucial across planning, manoeuvre and control levels, allowing the driver to take the correct action, in order to successfully perform the driving task and avoid errors. The processing of information requires mental resources, affecting drivers' workload levels. An understanding of information processing is required to better understand the concept of mental workload, and performance problems arising from it. Therefore, the next sub-section expands on information processing and its effect on human performance.

1.3.3.2 Information processing and performance

Information processing relates to changes in information that are detectable by an individual, including how individuals perceive, manipulate, analyse and retain information. Driving is a complex task, and the human driver can be considered as an input-output system of information processing (Kalsbeek, 1968).

At a basic level, human information processing consists of four stages: Sensory processing, Perception, Response Selection and Response Execution (Parasuraman et al., 2000; Wickens, 1984). Sensory Processing and Perception can be considered as input functions, whereas Response Selection and Response Execution can be considered as the output functions. Wickens' (1984) information processing model further expands on these four stages, detailing the intermediary stages such as working memory, cognition and attention resources, as shown in **Figure 1.5**.

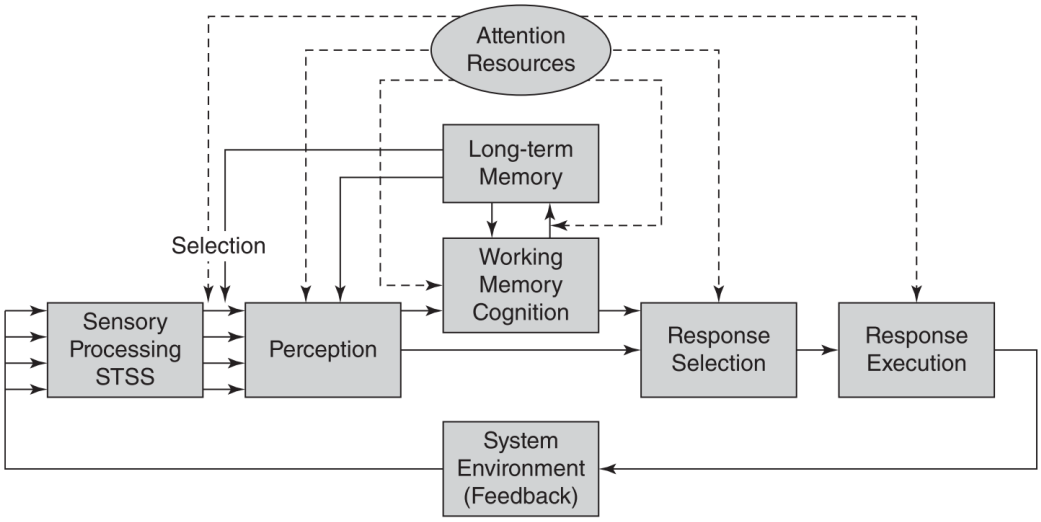

Figure 1.5. Model of human-information processing stages (redrawn from Wickens et al., 2013, p. 4)

One crucial aspect of Wickens' model of information processing is that, unlike a closed-loop system, it is a dynamic model, and involves constant interaction with the immediate environment (Wickens et al., 2013). This involves updating the sensory inputs constantly, based on the feedback from the environment, as well as any changes in system state, created due to previous response execution, which requires attentional/ mental resources at various stages of information processing, as shown in **Figure 1.5**.

Within information-processing theories is the concept of limited processing capacity (Kahneman, 1973; Wickens, 1984). Kahneman (1973) suggested a single undifferentiated *capacity*, from which mental *resources* are available for task performance. The words *capacity* and *resource* have been used interchangeably in the past, without differentiating each from one another (O'Donell & Eggemeier, 1986). However, the processing capacity of a human is limited (Kahneman, 1973; Wickens, 1984), and Wickens (1992) and de Waard (1996) defined this *capacity* as the maximum or upper limit of processing capability, and *resource* as the amount of processing facilities allocated, representing the mental effort supplied to improve processing efficiency.

However, resources are limited, and Wickens (1984)'s multiple resource theory postulates that people have a limited set of resources available for mental processes,

and these resources are shared across different tasks, modalities and processes. There is a central resource pool which is required for performing almost all tasks. Additionally, Wickens (1984) also suggested that different task modalities might require different sets of resource pools, for optimal task performance. When an overlap of resource requirement occurs, for example, doing two visual tasks simultaneously, requiring visual resources, it can affect the performance. Whereas for tasks that require different resources, for example, performing an auditory task along with a visual task, performance of both tasks can remain unchanged, as there is no overlap between visual and auditory resources, provided there is no performance decrement due to exhaustion of the central resource pool (de Waard, 1996). Resources are characterised by two general properties: they are generally deployed under voluntary control, and they are scarce/limited in quantity.

There is an implicit assumption in Wicken's multiple resource theory that the size of the resource pool is generally fixed. However, Young & Stanton (2002) introduced the Malleable Attentional Resources Theory (MART), which posits that the size of the resource pool is not fixed, and can change in size according to the changes in task demand. For example, in doing a low demand task, the operator's resources would shrink to accommodate and demand reduction. MART was able to explain why degradation in attention and performance was observed, even in low demand tasks. In order to better understand how human performance is affected by task demand, we look at the relationship between workload, task demand and performance within the driving context, in the next section.

1.3.3.3 Workload, task demand and performance

According to de Waard (1996), workload is the proportion of information processing capacity that is utilised for task performance. In relation to driving, Parasuraman et al. (2008) defined workload as the relationship between the mental resources demanded by a task, and those resources available to be supplied by the driver. In both these definitions, workload is dependent on the individual, as well as the nature of the task that is performed. Mulder (1986) distinguished mental workload into two components, one being the mental workload associated with processing the information presented in a controlled mode (or computational effort) and the other being the mental effort that is required to compensate when there is a dip in the driver's

energy resources (compensatory effort). Effort here reflects the driver's reaction to demand. The computational effort is used to maintain an adequate level of task performance, whereas compensatory effort is required when the performance decreases below a certain level, such as due to boredom or fatigue (G. Mulder, 1986; Silva, 2014).

One way to conceptualise mental workload is by considering its relationship to performance. Some researchers suggest that workload and performance follows an inverted U-shaped curve (Bruggen, 2015), similar to Yerkes-Dodson's Law on arousal and performance (Yerkes & Dodson, 1908). The relationship between workload and task performance is quite complex and not necessarily linear. Mental workload can vary between extremely low (underload) and extremely high levels (overload, or when the demand exceeds processing capacity). These two extremes between the optimal level (i.e. the appropriate mental workload where the driver is in a comfortable state, while catering to the task demands without any reduction in task performance) are not favourable and can lead to depreciation in task performance, and diminish information processing capabilities, for example, that which is required to successfully execute the driving task (Bruggen, 2015; Lenné et al., 1997; Rusnock & Borghetti, 2016; Silva, 2014). However, this model does not explicitly specify the interaction between task demand, with workload and performance.

Meister (1976) proposed a model to explain the relationship between task demand and performance. Within the driving context, de Waard (1996) proposed a modified version of Meister's model, to include workload as an additional dimension, in relation to driving, and is shown in **Figure 1.6**. He divided the workload into 6 regions as shown in the figure. Region D indicates the effects of monotonous tasks on workload and performance. Low task demand and under-stimulation (underload condition) will lead to decreased attention and diminished arousal (Grandjean, 1979), leading to low performance. The transition from region D to A1 is associated with monotony experienced by the driver where he needs to make a greater effort (compensatory effort) to maintain their performance level. In regions A1 (compensatory effort) and A3 (computational effort), the driver has to increase their effort to maintain their performance and match the task demands. A2 denotes the optimal performance region where the driver is able to meet all the task demands with

minimal or no extra effort at optimal workload levels, without any deterioration in their performance. However, when the task demand becomes too great and computational effort cannot be maintained, this leads to a dip in performance (region B) until the increasing task demand leads to mental overload, resulting in low performance efficiency (region C) (de Waard, 1996; Meister, 1976).

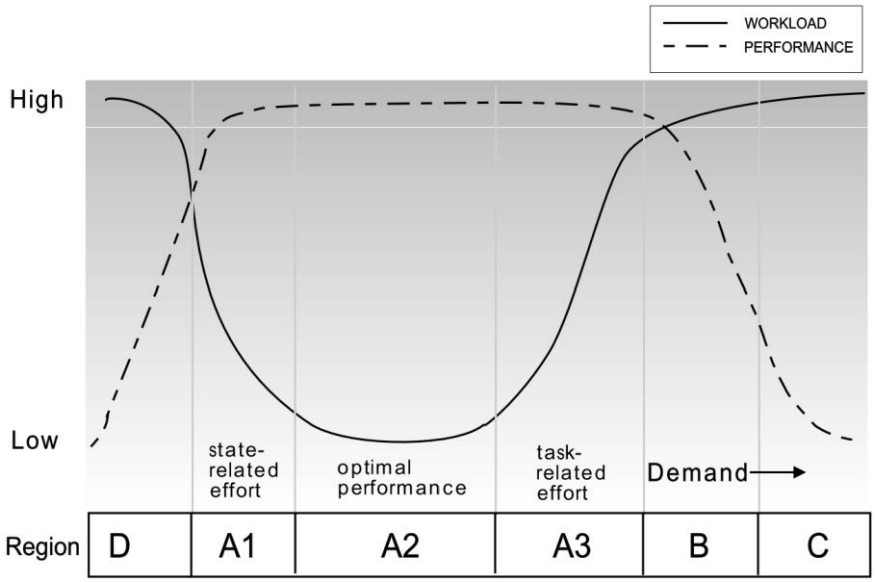

Figure 1.6. Workload, task demand and performance (de Waard, 1996, p. 24)

In this thesis, only task-related effort, and performance decrements associated with increased task demands (computational effort), will be considered. Drivers using different levels of automation are likely to experience workload differently. For example, drivers in SAE L2-enabled vehicles would have to monitor the drive constantly, which can result in similar levels of mental workload and attentional demand as manual driving (Lohani et al., 2021; Stapel et al., 2019). In SAE L3 automated driving, drivers are not required to monitor the drive, which can reduce their workload levels. However, they are more likely to engage in other non-driving related activities (Carsten et al., 2012), including visual NDRTs such as using their smartphone, which results in an increase in their workload levels. If and when they are required to resume manual control of the vehicle, they would have to switch back to the driving task, from an OOTL state, which is likely to negatively affect both performance and safety (Zeeb et al., 2016). The next sub-section discusses how mental workload can affect driving performance during HAD, and around takeovers.

1.3.3.4 Task switching, mental workload and HAD

A transition of control to manual driving would involve task switching, where the driver has to stop his current task, such as monitoring the drive (in SAE L2 enabled vehicles) or other non-driving related activities (in SAE L3 enabled vehicles), and switch their physical, mental and attentional resource to perform the manual driving task. However, this task switching comes at a switch cost, which affects both the driver's workload and task performance in resuming manual control of the vehicle (Dogan et al., 2017). Concurrent tasks, that use similar mental resources to perform two distinct tasks, such as talking on the phone and driving, both of which require mental resources for information processing, can result in mental overload if the driver is negotiating a complex driving environment and/or the phone conversation is mentally taxing, resulting in diminished performance (Wickens et al., 2015). Sequential task performance involves the driver performing one task or the other at a time, with the limited resources available to perform the tasks, as they are unable to perform both the tasks concurrently, without being in a mental overload condition (Wickens et al., 2015). There are many examples of task failures during sequential task performance in driving, such as texting on the phone while driving, where the driver has to choose between either texting or driving, at a time, which can take the driver's attention away from the road, resulting in accidents (Klauer et al., 2006). During automated driving, task switching from a demanding non-driving related activity, to the driving task during transition, can come with a switch cost, in terms of reaction time and higher error rates (Jersild, 1927; Monsell, 2003), as well as a decrement in performance in the follow on task, which is manual driving (Monsell, 2003).

Parasuraman et al. (2008) suggested that workload is a better predictor of drivers' future performance, than their current performance, and in order to ensure optimal driving performance when they have to resume manual control, their workload should be at an optimum level (neither underload nor overload condition, region A2 of **Figure 1.6**). For example, it can be argued that a driver's ability to safely resume control from automation is likely to be affected if they are engaged in a high workload task during HAD, such as a demanding NDRT, with worse performance observed in terms of higher reaction time, and poorer lateral control, indicated by larger deviations from the lane centre and higher lateral acceleration, compared to consequent driving performance after a no task period during HAD (Gold et al., 2015; Zeeb et al., 2016).

As drivers are more likely to engage in NDRTs as automation level increases (Carsten et al., 2012), it is imperative that drivers' workload levels are monitored at different stages of the HAD, to provide mitigation strategies, and reduce performance decrements and errors upon resumption of manual control during transitions (Merat et al., 2012).

Drivers' workload levels are not just affected by NDRTs, but also, the driving environment itself. Studies have shown that increased traffic (Radlmayr et al., 2014), complex driving environments such as driving through a busy city centre or suburban road (Foy & Chapman, 2018), and short time headways during car-following (Liu et al., 2019) increase drivers' workload levels. Therefore, in this thesis, we investigate the effect of different car-following situations and NDRTs, on drivers' workload levels, and whether driver workload can be objectively measured in a continuous manner.

Driver workload is highly subjective depending on the individual, and how they are affected by, or perceive it. Workload has been measured in most studies using subjective ratings, such as NASA-TLX (Sandra G. Hart & Staveland, 1988a; Stapel et al., 2019). However, Hart & Wickens (1990) suggested that such self-reported workload ratings reflect subjective impressions of workload, that may differ from the workload as reflected by task performance. The ability to objectively measure workload in real-time during different stages of HAD is crucial, as it can provide insights into drivers' capabilities and limitations, such as whether they are in an underload or overload state, ultimately helping to improve the safety of the automated system. Real-time, minimally intrusive, and continuous assessment of driver workload can be used to assist the driver, for example, to warn them of their dangerous overload or underload states (Merat et al., 2012). The next section of this thesis expands on how physiological metrics can be used as an objective indicator to assess driver states of discomfort and mental workload, during HAD.

1.4 Use of physiological metrics for driver state monitoring

Psychophysiology conceptualises how different psychological processes trigger physiological responses within the body, especially with a focus on higher cognitive processes (Cacioppo et al., 2007a). Physiological signals have been used to understand, and objectively quantify, different driver states such as fatigue, stress,

workload, attention or discomfort (Beggiato et al., 2018; Cho et al., 2017; Foy & Chapman, 2018; Mehler et al., 2009; Patel et al., 2011). The next sub-section provides a basic understanding of physiology and its relationship with a number of key psychological constructs, with a focus on autonomic (relating to autonomous nervous system or ANS) physiological functions.

1.4.1 Introduction to psychophysiology

Understanding how physiological signals can be used, as an objective measure of drivers' psychological state, requires a deeper understanding of the relationship between physiological signals and psychological constructs. The relationship between psychological constructs and physiological signals is quite complex, and varies across scenarios, environments and individuals. One issue surrounding psychophysiological relationship is its limited validity (such as validity in dynamic driving environments), because the clarity in relationship between psychological functions (e.g. cognitive functions) and physiological signals (e.g. skin conductance, heart rate) is observed only in certain well-prescribed assessment contexts (Cacioppo et al., 2007b).

The Autonomic Nervous System (ANS) controls the unconscious actions of the body, and helps in understanding how certain psychological states can elicit physiological responses. This consists of the sympathetic (SNS) and parasympathetic nervous system (PNS). PNS is generally associated with homeostasis or activities of the body when it is at a state of rest (called rest and digest functions), whereas (SNS) is generally associated with the 'fight or flight' response, which is the body's reaction to threat, a harmful event or survival. It also complements the PNS by being constantly active in the background at a very basic state, to maintain homeostasis of the body (Brodal, 2010; Cacioppo et al., 2007a). While historically, it was assumed that the status of the nervous system existed along a single continuum extending from the parasympathetic to sympathetic domain, more recent representations of ANS and its functioning suggests that the parasympathetic and sympathetic nervous systems can change reciprocally, coactively or independently (Berntson et al., 2007). Driver states such as high workload, or discomfort, results in increased SNS activity, as well as diminished PNS activity. **Table 1.3** below provides an overview of how PNS and SNS affect different organs of the human body.

Table 1.3. Effect of SNS and PNS on different organs of the human body (adapted from Brodal, 2010)

Organ	Sympathetic nervous system (SNS)	Parasympathetic nervous system (PNS)
Arteries	Vasoconstriction (increased blood pressure)	Vasodilation (decreased blood pressure)
Skin	Sweat secretion	No innervation
Heart	Increased heart rate	Reduced heart rate
Airways	Relaxation of bronchial muscles	Contraction of bronchial muscles, secretion from mucous glands
Eye	Pupillary dilation	Pupillary constriction, accommodation, secretion of tears

As can be observed in the table, different physiological responses, by different organs in the body, are controlled by either PNS or SNS, or in some cases, by both. Previous research has indicated promising results in using skin-based (such as Electrodermal activity or EDA), eye-based (pupillometry) and heart-based (such as Electrocardiogram or ECG) physiological signals, as indicators of drivers' discomfort or workload levels (Beggiato et al., 2019; Foy & Chapman, 2018; Mehler et al., 2009). Moreover, EDA and ECG measures are minimally intrusive compared to brain-based signals such as electroencephalogram (EEG), with technological advancements leading to even non-contact techniques to capture these signals (Kranjec et al., 2014). Therefore, this thesis focuses on minimally-intrusive or non-intrusive physiological measures of electrodermal activity (EDA), heart rate variability (HRV) along with pupillometry, as an indicator for driver state. The next sub-section of this thesis focuses on EDA and its relationship to psychological processes.

1.4.2 Electrodermal activity

Electrodermal activity (EDA) relates to changes in electrical potential of the skin, due to sweating. Research has indicated that the eccrine sweat glands located in the palmar and plantar surfaces are affected more by psychological factors, than thermoregulation (Edelberg, 1972; Shields et al., 1987). Recent research has shown

convincing evidence linking EDA to SNS activities (Beggiato et al., 2019; Boucsein, 2012).

The EDA signal is typically classified into tonic and phasic components. The tonic components record the slow changes in electrical conductivity of the skin and the most common measure for this is skin conductance level (SCL; see **Figure 1.7**). The fast or rapidly evolving component is the phasic component or skin conductance response (SCR; see **Figure 1.7**), which is usually triggered by an event (Braithwaite *et al.*, 2013). Instead of the conventional trough-to-peak method to separate the tonic and phasic components of an EDA signal, Benedek & Kaernbach (2010) proposed continuous decomposition analysis to extract the phasic component of the EDA signal, where the raw EDA data is deconvolved first using a general response shape of the EDA signal, following which, it is decomposed into the tonic and phasic components. This method has been shown to be more accurate, as well as computationally robust (Benedek & Kaernbach, 2010), and therefore, the processing of all EDA data included in this thesis is based on this method.

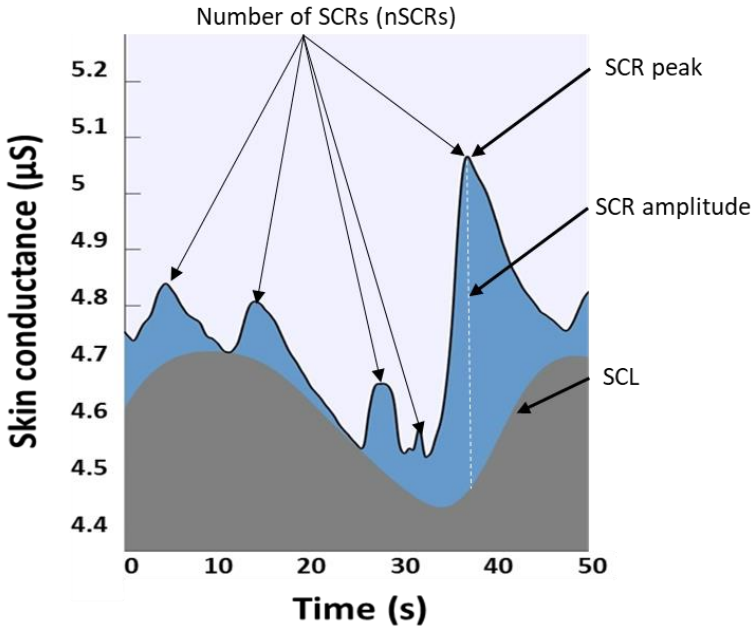

Figure 1.7. Tonic (grey) and Phasic (dark blue) components of an EDA signal after continuous decomposition analysis, done on MATLAB R2016a using Ledalab software (Benedek & Kaernbach, 2010).

It is extremely difficult to identify an isolated SCR event, with the corresponding exact stimuli, without controlled context. Although, this can be circumnavigated by having a strictly controlled experimental paradigm (e.g. having just one aspect of stimulus change across experimental conditions) or by using it in conjunction with other physiological metrics (Dawson et al., 2007). The various derived metrics from both the tonic and phasic components of an EDA signal, its definition and typical values, are shown in **Table 1.4**, some of which are visually represented in **Figure 1.7** .

Table 1.4. EDA metrics, definitions and typical values (Dawson et al., 2007)

Measure	Definition	Typical Values
Skin conductance level (SCL)	Tonic level of electrical conductivity of the skin	2-20 μS
Change in SCL	Gradual changes in SCL measured at two or more points in time	1-3 μS
Frequency of NS-SCRs	Number of SCRs in absence of identifiable eliciting stimulus	1-3 per min
SCR amplitude	Phasic increase in conductance shortly following stimulus onset	0.1-1.0 μS
SCR latency	Temporal interval between stimulus onset and SCR initiation	1-3 s
SCR rise time	Temporal interval between SCR initiation and SCR peak	1-3 s
SCR half recovery time	Temporal interval between SCR peak and point of 50% recovery of SCR amplitude	2-10 s
SCR habitation (trials to habituation)	Number of stimulus presentations before two or three trails with no response	2-8 stimulus presentations
SCR habituation (slope)	Rate of change of event related SCR amplitude	0.01-0.5 μS per trial
nSCRs/min	Number of significant SCRs above amplitude threshold (generally between 0.01 – 0.05 μS) in a minute	Stimulus dependent (10 – 45 nSCRs)

While electrodermal activation is sensitive to a wide range of stimuli, research has shown a strong relationship between EDA and SNS-related activities, such as arousal (Boucsein, 2012). Within the driving domain, EDA-based measures such as tonic SCL, and nSCR/min (see last row in **Table 1.4**) have been used as indicators of

driver workload or discomfort. For example, using a manual driving simulator study, Mehler et al. (2009) found that drivers' tonic SCL increased, when they performed a cognitive n-back (recall) task during driving. The n-back task required the driver to recall the number presented n-steps earlier in the sequence, at a given moment. The authors report a significant increase in SCL from baseline (no task) to 0-back and 1-back tasks, suggesting that SCL is sensitive to incremental increases in driver workload. Similarly, Foy & Chapman (2018) observed significant differences in driver workload, across different road environments such as dual-carriageways, suburban roads and city centre driving, with the authors finding a significant increase in drivers' nSCR/min values, as the driving environment became more complex and drivers' workload levels increased. The objective workload levels, as indicated by nSCR/min values, correlated to subjective NASA-TLX workload ratings collected in the study (Foy & Chapman, 2018; Sandra G. Hart & Staveland, 1988b). Research on the use of EDA signals in detecting discomfort during driving is quite limited. However, in a recent driving simulator study on HAD, Beggiato et al. (2019) observed that tonic SCL was sensitive to discomfort-inducing scenarios such as when the vehicle was negotiating a complex intersection without traffic lights, or when aggressively approaching a red traffic light. The authors observed significant increases in SCL levels around periods of discomfort (Beggiato et al., 2019).

However, the majority of the studies on EDA have been conducted in either laboratory environments, including fixed-base simulators for studies on driving. As such, there is limited validation for EDA signals in highly dynamic environments, such as real-world driving studies. EDA signals are known to be highly susceptible to motion-artefacts (Braithwaite et al., 2015; Taylor et al., 2015). Therefore, this thesis addresses this research gap by evaluating the validity of EDA signals, in highly dynamic driving environments, and developing novel methods to remove such artefacts. The next sub-section expands on the use of heart rate variability as an indicator of driver state.

1.4.3 Heart rate and heart rate variability

The chain of events that occur from one heartbeat to the next is referred to as a cardiac cycle, which begins with the depolarisation of the heart's sinoatrial (SA) node. This is the P wave in the electrical signal generated by the heart, also known as

an electrocardiogram (ECG), which is depicted in **Figure 1.8**. This is shortly followed by the contraction of the atrial and ventricular chambers, which appears as the QRS complex (**Figure 1.8**) on the ECG signal. The Q and R waves denote the depolarisation of the ventricular septum, followed by the bulk of the ventricular myocardium, respectively. At the end of a ventricular contraction, the atrioventricular valve closes, causing pressure to build up in the ventricular chamber, leading to blood filling up the ventricular chamber. Towards the end of ventricular contraction, the atrioventricular nodes repolarise, which is observed as the T wave in an ECG signal (**Figure 1.8**), initiating the relaxation of the ventricular chamber, and opening the aortic value to pump the blood across the body (Berntson et al., 2007).

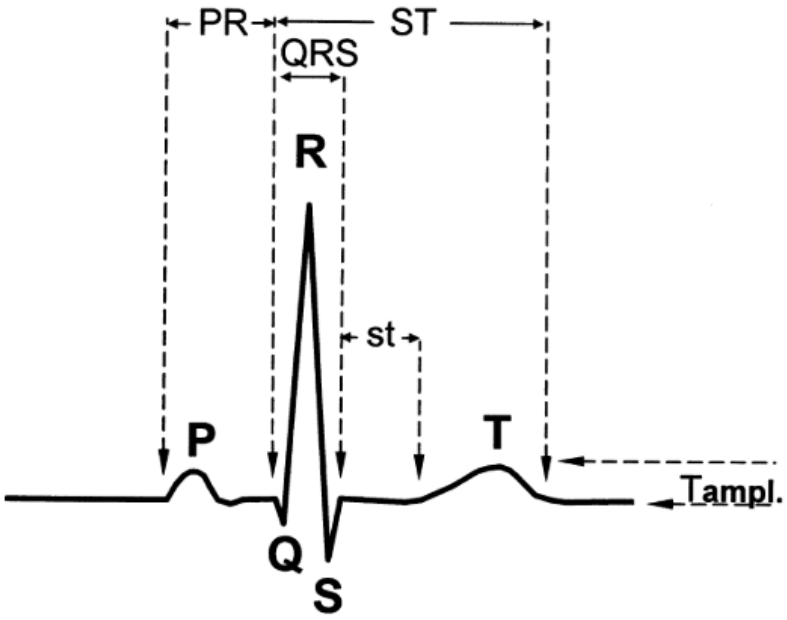

Figure 1.8. A typical ECG signal denoting a single cardiac cycle, including the QRS complex (adapted from Berntson et al., 2007, p.184).

Data acquisition, at high sampling rates of over 1000 Hz, minimises loss of information, and reduces noise in R-R waves, where the R peaks can be demarcated easily, even with motion (Berntson et al., 2007). Additionally, band-pass filtering could be implemented to further remove the noise from the ECG signal for R peak detection (An & Stylios, 2020). Patel et al. (2011) provides a well explained example of peak detection algorithm that can be used extract R-R peaks from an ECG signal. However, it should be noted that removing noise from the QRS complex is challenging and complicated, especially without distorting the shape of the QRS complex (van Gent et

al., 2018). Most autonomic cardiac responses can be derived from R-R intervals, and therefore, removal of noise from the QRS complex, while preserving its shape, is beyond the scope of this thesis. Although, zero-phase infinite impulse response (IIR) filters based on a high-pass Butterworth filter, with a cut-off frequency of 0.5 Hz, and filter order 2, has been shown to be an effective tool in removing motion artefacts from the QRS complex of the ECG signal (An & Stylios, 2020).

Heart rate (HR) refers to the number of heart beats that occur within a minute, where each heartbeat is generally measured by successive R wave peaks in the ECG signal, as R wave peaks are more pronounced and, hence, easier to identify. In contrast, photoplethysmogram (PPG), uses optical sensors which are sensitive to discolouration in the skin as blood perfuses through the blood vessels during each heartbeat, as opposed to ECG, which directly measures electrical signals from the heart. Additionally, PPG is highly susceptible to motion artefacts, which can completely distort the QRS complex in a highly dynamic environment (van Gent et al., 2018), and make peak detection almost impossible in certain cases of high movement. Therefore, this thesis focuses on using ECG signals as a measure for cardiac metrics.

Studies show that the heartbeat does not necessarily occur at regular intervals (Berntson et al., 2007), and this variation in the time interval between heartbeats is referred to as heart rate variability (HRV). The cardiovascular actions causing this variation in heart rate are generally associated with the activities of both SNS and PNS (Berntson et al., 2007). Of particular interest is the spectral functions of cardiac signals, in the high frequency (HF) region of an HRV signal (0.15 – 0.4 Hz), which is predominantly controlled by PNS activity and reflects vagal control (Berntson et al., 2007; Laborde et al., 2017). Recent research has indicated that SNS is constantly active, even during low arousal states, to maintain homeostasis, therefore, reduction of PNS activity could also be used as a better indicator of higher arousal states, and vagal or PNS indices offer a clearer interpretation of psychological constructs such as stress, workload or fatigue (Laborde et al., 2017; Mehler et al., 2009; Patel et al., 2011). Of particular interest amongst the vagal metrics, is the time domain metric of root mean squared of successive differences of R-R peak intervals (RMSSD), which, unlike spectral HF metrics, is not influenced by respiration rate (Hill et al., 2009), and

is thought to provide a better assessment of vagal tone (Otzenberger et al., 1998). A list of commonly used HRV metrics, and their description, is given in **Table 1.5** below.

Table 1.5. Summary of commonly used HRV parameters and their description (adapted from Laborde et al., 2017, p. 4)

	Variable	Description	Physiological origin
Time domain	SDNN	Standard deviation of all R-R intervals	Cyclic components responsible for HRV
	RMSSD	Root mean square of successive differences in R-R intervals	Vagal tone
	pNN50	Percentage of successive normal sinus R-R intervals more than 50 ms	Vagal tone
	Peak-valley	Time domain filter dynamically centred at the exact ongoing respiratory frequency	Vagal tone
Frequency domain	ULF	Ultra-low frequency	Circadian oscillations, core body temperature, metabolism
	VLF	Very-low frequency	Long-term regulation mechanisms, thermoregulation and hormonal mechanisms
	LF	Low frequency	Mix of sympathetic and vagal activity, baroreflex activity.
	HF	High frequency	Vagal tone or PNS
	LF/HF	Low frequency/high-frequency ratio	Mix of SNS and PNS activity
Non-linear indices	SD1	Standard deviation – Poincare plot crosswise	Depicts quick and high frequency changes in HRV
	SD2	Standard deviation – Poincare plot lengthwise	Depicts long term changes in HRV

Within the driving context, both HR and HRV-derived metrics have been used to indicate driver states such as fatigue (Patel et al., 2011), high workload (Mehler et al., 2009) or stress (Taelman et al., 2008). A general finding is that HR increases, while

HRV reduces. Studies show that RMSSD values reduce during high arousal states such as stress (Orsila et al., 2008), high workload (Mehler et al., 2009) or discomfort (Beggiato et al., 2019). While ECG signals are susceptible to motion-artefacts, these are less affected by motion, when compared to a PPG or EDA signal. One limitation of HRV-based metrics is that they require at least two minute (ideally five minute) windows, for accurate analysis (Bourdillon et al., 2017; Laborde et al., 2017). Since the signal has a high decay time, it might include carryover effects from previous stimuli, when presented with continuously fluctuating stimuli. However, there is limited research on the use of HRV-based metrics in the HAD context. Investigating the sensitivity of this metric to less pronounced stimuli, that induces discomfort or workload, warrants further research. However, as mentioned earlier, physiological changes can be caused by a variety of factors, and eye tracking helps us in understanding where the attention of the driver is, thereby providing more context as to why the physiological change is happening. In the next sub-section, eye tracking metrics and its usage in driver state detection is explained.

1.4.4 Eye-based metrics and pupillometry

In most modern eye trackers, the eye movement and the direction of gaze are obtained by optical tracking of corneal reflections, which helps assess visual attention. Eye trackers have become more and more compact and minimally intrusive over the years.

Most modern eye trackers utilize near-infrared technology along with a high-resolution camera (or other optical sensor) to track gaze direction, as well as demarcating the pupils from the rest of the eye. Gaze-based eye tracking metrics have been used in driving research as an indicator of visual attention (Chapman & Underwood, 1998; Louw, Kuo, et al., 2019; Louw, Madigan, et al., 2017; Merat et al., 2014; Wilkie & Wann, 2003). Eye tracking-based metrics are widely used for driver state monitoring, owing to the non-intrusive nature of measurement. **Table 1.6** summarises various commonly used eye-based indices, their definitions, corresponding psychological relationships and typical values.

Table 1.6. Eye-based metrics, definitions and typical values (Beggiato et al., 2019; Mathôt, 2018; Rauch et al., 2009)

Metric	Definition	Psychological relation	Typical values
Gaze (Pitch and Yaw)	The pitch and yaw angles of vectors indicating the direction of gaze	Visual attention	-
Fixations	Gaze cluster formed by series of gaze points close to each other in time and magnitude, indicating a period in which the eyes are locked onto an object	Visual attention	100 – 300 ms
Area of Interest (AOI)	User defined regions within the display stimulus	Visual attention	-
Time to first fixation	Amount of time it takes to fixate on a specific AOI, from stimulus onset	Indicative of detection time of hazards and objects	-
Pupil diameter	The diameter of the eye's pupil, which can increase or decrease due to stimuli	Arousal, workload and light intensity can result in an increase in pupil diameter.	2-4 mm in bright conditions and 4-8 mm in dark conditions
Blink frequency and duration	The closing and opening of eyelids, its frequency and duration	Indicative of cognitive load, discomfort and drowsiness	-
PERCLOS	Percentage eyelid closure	Drowsiness indicator	-

In addition to measuring drivers' visual attention, studies using pupillometry (measurement of pupil size and reactivity) have indicated a relationship between pupil dilation and constriction, and cognitive activity, with the pupil dilating when the cognitive load or task demand increases, even for non-visual tasks (Marquart et al., 2015; Mathôt, 2018).

For example, in their driving simulator-based study on manual driving, Palinko et al. (2010a) observed that drivers' pupils dilated significantly, with increases in cognitive load, presented in the form of two NDRTs: a twenty questions task, and a

last letter (memory recall) game. Using a driving simulator study to investigate driver discomfort in HAD, Beggiato et al. (2019) found that pupils dilated during discomfort-inducing situations, such as when the vehicle was negotiating a complex intersection without traffic lights, or when aggressively approaching a red traffic light. However, pupil diameter variations can be caused by other factors. These include variations due to brightness levels, also known as pupil light response, which results in larger pupil diameter during darker conditions (Mathôt, 2018; Spector, 1990). Shifting of focus between a nearer object to a farther object, can also result in pupil dilation, also known as accommodation reflux (Mathôt, 2018). Finally, orienting response, which occurs when there is a sudden change in the environment, for example as caused by a startle response to sounds, movement or touch, can also lead to pupil dilation. This is normally a small change in magnitude of pupil diameter, and of short duration, which occurs within 0.5 s to 1s of stimulus onset (Mathôt, 2018).

In addition to changes in pupil diameter, blink frequency and blink duration have also been associated with changes in cognitive load (de Waard, 1996; Marquart et al., 2015; Merat et al., 2012). However, results of studies using these metrics have been somewhat conflicting, with some reporting an increase (Recarte et al., 2008) and decrease in blink frequency and duration (Veltman & Gaillard, 1996), with an increase in cognitive load. Recarte et al. (2008) argued that visual attention can lead to blink inhibition, and thereby reducing blink frequency and duration, as blink suppression aids in decision making. However, blink rates and duration can increase due to performing a cognitive task, as the resources required to perform the cognitive task while driving can interfere with the resources required for blink inhibition (J. A. Stern et al., 1994). Therefore, in this thesis, blink metrics have been avoided in favour of pupil diameter, as the latter provides a stronger inference with the underlying psychological construct (in this case, cognitive activity or workload).

While eye tracking-based metrics can provide deeper insights into drivers' cognitive and attentional states, there are some limitations and drawbacks. For example, the sensitivity of eye trackers to psychological states, such as discomfort, is not well established, and warrants further research. Furthermore, for unobtrusive, dash-based, eye trackers, a loss of the eye or face features from the camera sensors can hinder real-time driver state detection, and potentially compromise driver safety.

For example, during HAD, if the drivers are engaged in an NDRT, and looking away from the camera's field of view, the eye tracking system would be unable to assess the driver's state. Therefore, combining eye tracking data with other physiological metrics, such as skin conductance and heart rate-based measures, can help improve driver state detection in future AVs.

1.5 Summary of key research gaps

We are still a long way off from achieving full autonomy, where all aspects of the driving task and sub-tasks are undertaken by the automated system, without requiring any intervention from the human driver. Therefore, driver state monitoring plays a critically important role in ensuring safety of the automated vehicle, especially in situations where the driver might have to resume control of the vehicle. Real-time detection of driver states, such as discomfort and high workload, can improve human-automation interaction by enabling the automated system to select appropriate mitigation strategies and assist the driver in safely and successfully resuming control of the vehicle, when required. This will eventually improve trust and wider acceptance of these systems by drivers (Carsten & Martens, 2019). However, there are still significant gaps in our knowledge about the effect of automation on certain driver states, such as discomfort or workload, and how such driver states can be objectively measured using physiological signals, as seen below:

- Physiological signals have been used in the past as an indicator of driver states such as discomfort (Beggiato et al., 2019) or high workload (Mehler et al., 2009). However, a vast majority of psychophysiological studies have been conducted in a laboratory environments. In case of the limited number of driving related psychophysiological studies, a fixed-base simulator environment was used (Beggiato et al., 2019; Foy & Chapman, 2018; Mehler et al., 2009). With regards to ECG signals, as mentioned in section 1.4.3, when data is recorded at frequencies above 500 Hz, the R-peaks are highly unlikely to be affected by motion artefacts. Simple zero phase infinite impulse response filters can be used to remove noise from the ECG signal. However, EDA signals have been shown to be highly sensitive to motion artefacts (Taylor et al., 2015). Therefore, further research is required to establish whether motion

artefacts can be removed from an EDA signal during a highly dynamic driving environment, in order to make useful interpretations of driver state from the EDA signal, in real-world driving environments.

- Ensuring driving comfort can help improve trust and encourage wider acceptance, as well as safety of automation features in driving (ERTRAC, 2017; Molnar et al., 2018; Siebert et al., 2013). However, as the driver changes from an active operator to a supervisory role, with an increase in automation level, the factors and threshold limits for driver discomfort changes. There has been limited research on how factors such as an AV's driving style (including acceleration and jerk forces, safety margins or "human-like driving") can induce driver discomfort, during automation. Additionally, the different measurement techniques that can be used to quantify driver discomfort warrant further investigation, including the usability of physiological metrics such as SCRs, mean HR and RMSSD, as an objective and continuous measure of discomfort, to enable real-time discomfort detection in future driving automation systems.
- Drivers' workload levels can affect both performance and safety of the vehicle during HAD, if and when the driver has to resume manual control of the vehicle (Merat et al., 2012; Parasuraman et al., 2008). Therefore, continuous monitoring of drivers' workload levels can be used in warning the driver of dangerous underload or overload conditions, and provide appropriate mitigation strategies (Merat et al., 2012). There has been limited research on understanding how different factors, such as presence of a lead vehicle, safety margins in terms of time headway maintained from a lead vehicle, monitoring the drive or engaging in NDRTs, can affect driver workload, during HAD and during transitions, and whether these more subtle changes (as opposed to changes due to more cognitively demanding memory recall tasks) in workload levels can be captured by physiological metrics.

1.6 Research questions and thesis overview

The main aim of this thesis is to investigate and validate the usage of physiological measures as an objective indicator of driver state in dynamic driving environments, and understand if such a methodology can be used to measure driver discomfort, and high workload, as well as how such states affect drivers' resumption

of control from automation. In particular, this thesis addresses the following research questions:

1. Can electrodermal activity be used to assess drivers' discomfort and workload levels, in a dynamic driving environment?
2. What are the primary factors contributing to driver discomfort during HAD, and are these reflected in drivers' physiological state?
3. Are changes in workload levels during different stages of HAD, including the transition of control, reflected in drivers' electrodermal activity and electrocardiogram-based physiological metrics?
4. How are drivers' attentional demand, and workload levels, affected at different stages of HAD, including transition of control, as reflected in their pupil diameter values?

Driver state manipulations were performed to induce driver states of discomfort and increased workload, in this research. Discomfort was manipulated by controlling the driving style of the automated vehicle, including acceleration and jerk forces, human-like driving profiles, safety-margins to obstacles and road boundaries as well as different road geometries. Driver workload was manipulated by using a non-driving related visual task, the presence of a lead vehicle during takeover scenarios and by manipulating the headway distance maintained by a lead vehicle, during car-following situations. Physiological indices derived from EDA, ECG and eye tracking measures were used as objective indicators of driver state, and validated against subjective ratings of either discomfort or workload.

The overall outline of this thesis, including the remaining five chapters, is shown in **Figure 1.9** below.

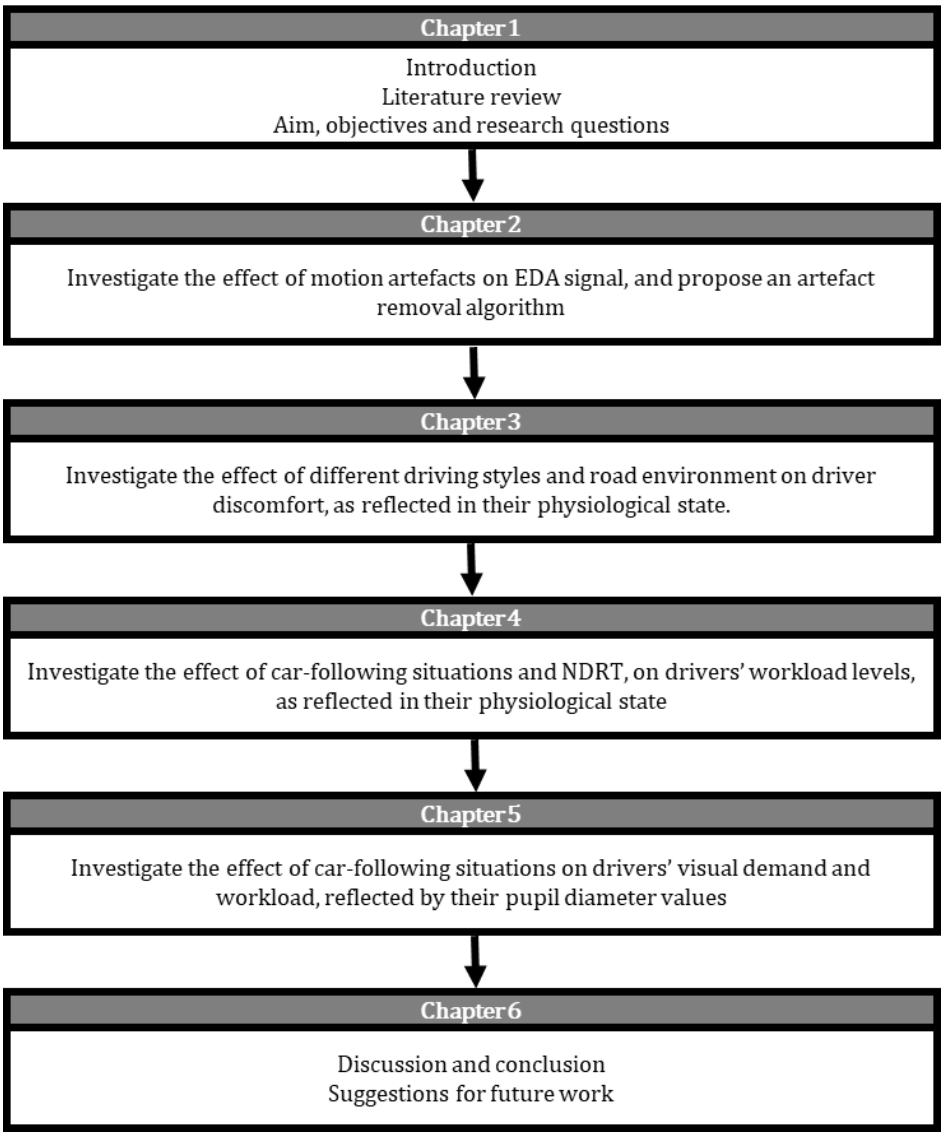

Figure 1.9. Thesis structure.

1.7 References

Alessandrini, A., Campagna, A., Site, P. D., Filippi, F., & Persia, L. (2015). Automated vehicles and the rethinking of mobility and cities. *Transportation Research Procedia*, *5*, 145–160. https://doi.org/10.1016/j.trpro.2015.01.002

An, X., & Stylios, G. K. (2020). Comparison of motion artefact reduction methods and the implementation of adaptive motion artefact reduction in wearable electrocardiogram monitoring. *Sensors*, *20*(5), 1468. https://doi.org/10.3390/s20051468

APA Dictionary of Psychology. (2022). *Task Demand*. American Pyschological

Association. https://dictionary.apa.org/task-demands

Bae, I., Moon, J., & Seo, J. (2019). Toward a comfortable driving experience for a self-driving shuttle bus. *Electronics (Switzerland)*, *8*(9), 943. https://doi.org/10.3390/electronics8090943

Banks, V. A., Plant, K. L., & Stanton, N. A. (2019). Driving aviation forward; contrasting driving automation and aviation automation. *Theoretical Issues in Ergonomics Science*, *20*(3), 250–264. https://doi.org/10.1080/1463922X.2018.1432716

Beggiato, M., Hartwich, F., & Krems, J. (2018). Using Smartbands, Pupillometry and Body Motion to Detect Discomfort in Automated Driving. *Frontiers in Human Neuroscience*, *12*, 338. https://doi.org/10.3389/fnhum.2018.00338

Beggiato, M., Hartwich, F., & Krems, J. (2019). Physiological correlates of discomfort in automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *66*, 445–458. https://doi.org/10.1016/j.trf.2019.09.018

Beggiato, M., Hartwich, F., Schleinitz, K., Krems, J. F., Othersen, I., & Petermann-Stock, I. (2015, November). What would drivers like to know during automated driving? Information needs at different levels of automation. *7th Conference on Driver Assistance*. https://doi.org/10.13140/RG.2.1.2462.6007

Bein, T., Bös, J., Mayer, D., & Melz, T. (2012). Advanced materials and technologies for reducing noise, vibration and harshness (NVH) in automobiles. In J. Rowe (Ed.), *Advanced Materials in Automotive Engineering* (pp. 254–298). Woodhead Publishing. https://doi.org/10.1533/9780857095466.254

Beller, J., Heesen, M., & Vollrath, M. (2013). Improving the driver-automation interaction: An approach using automation uncertainty. *Human Factors*. https://doi.org/10.1177/0018720813482327

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, *190*(1), 80–91. https://doi.org/10.1016/j.jneumeth.2010.04.028

Berntson, G., Quigley, K., & Lozano, D. (2007). Cardiovascular Psychophysiology. In G. Berntson, J. Cacioppo, & L. Tassinary (Eds.), *Handbook of Psychophysiology* (3rd ed., pp. 182–210). Cambridge University Press.

https://doi.org/https://doi.org/10.1017/CBO9780511546396

Billings, C. E. (1991). Human-centered automation: Principles and guidelines. In *International Journal of Aviation Psychology*. https://doi.org/10.1063/1.3663069

Bosetti, P., Da Lio, M., & Saroldi, A. (2014). On the human control of vehicles: An experimental study of acceleration. *European Transport Research Review*, *6*(2), 157–170. https://doi.org/10.1007/s12544-013-0120-2

Boucsein, W. (2012). *Electrodermal activity* (Second). Springer. https://doi.org/https://doi.org/10.1007/978-1-4614-1126-0

Bourdillon, N., Schmitt, L., Yazdani, S., Vesin, J. M., & Millet, G. P. (2017). Minimal window duration for accurate HRV recording in athletes. *Frontiers in Neuroscience*, *11*(AUG), 456. https://doi.org/10.3389/fnins.2017.00456

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2015). *A Guide for Analysing Electrodermal Activity & Skin Conductance Responses (SCRs) for Psychophysiological Experiments*. https://doi.org/10.1017.S0142716405050034

Braithwaite, J. J., Watson, D. G., Robert, J., & Mickey, R. (2013). A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments. …. https://doi.org/10.1017.S0142716405050034

Brodal, P. (2010). The central nervous system: Structure and function (4th ed.). In *The central nervous system: Structure and function (4th ed.).*

Brooke, L. (2020). *“Level 2+”: Making automated driving profitable, mainstream*. SAE International. https://www.sae.org/news/2020/12/rise-of-sae-level-2

Brown, I. D. (1994). Driver Fatigue. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *36*(2), 298–314. https://doi.org/10.1177/001872089403600210

Bruggen, A. (2015). An empirical investigation of the relationship between workload and performance. *Management Decision*, *53*(10), 2377–2389. https://doi.org/10.1108/MD-02-2015-0063

Bryan, M. E., Tempest, W., & Williams, D. (1978). Vehicle noise and the passenger. *Applied Ergonomics*, *9*(3), 151–154. https://doi.org/10.1016/0003-6870(78)90005-4

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (Eds.). (2007a). *The Handbook of Psychophysiology* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511546396

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2007b). Psychophysiological Science: Interdisciplinary Approaches to Classic Questions About the Mind. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (3rd ed., pp. 1–16). Cambridge University Press. https://doi.org/10.1017/CBO9780511546396.001

Cahour, B. (2008). Discomfort, affects and coping strategies in driving activity. *15th European Conference on Cognitive Ergonomics: The Ergonomics of Cool Interaction (ECCE '08)*, *369*, 1–7. https://doi.org/10.1145/1473018.1473046

Carsten, O., Lai, F. C. H., Barnard, Y., Jamson, A. H., & Merat, N. (2012). Control task substitution in semiautomated driving: Does it matter what aspects are automated? *Human Factors*, *54*(5), 747–761. https://doi.org/10.1177/0018720812460246

Carsten, O., & Martens, M. H. (2019). How can humans understand their automated cars? HMI principles, problems and solutions. *Cognition, Technology and Work*, *21*(1), 3–20. https://doi.org/10.1007/s10111-018-0484-0

Chapman, P. R., & Underwood, G. (1998). Visual search of driving situations: Danger and experience. *Perception*, *27*(8), 951–964. https://doi.org/10.1068/p270951

Cho, D., Ham, J., Oh, J., Park, J., Kim, S., Lee, N. K., & Lee, B. (2017). Detection of stress levels from biosignals measured in virtual reality environments using a kernel-based extreme learning machine. *Sensors*, *17*(10), 2435. https://doi.org/10.3390/s17102435

Dawson, M. E., Schell, A. M., & Filion, D. L. (2007). The Electrodermal System. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (3rd ed.). Cambridge University Press. https://doi.org/http://dx.doi.org/10.1017/CBO9780511546396.007

De Looze, M. P., Kuijt-Evers, L. F. M., & Van Dieën, J. (2003). Sitting comfort and discomfort and the relationships with objective measures. In *Ergonomics* (Vol. 46, Issue 10, pp. 985–997). https://doi.org/10.1080/0014013031000121977

de Waard, D. (1996). *The Measurement of Drivers' Mental Workload*. s.n.

Desmond, P. A., & Hancock, P. A. (2001). Active and passive fatigue states. In *Stress, workload, and fatigue.* (pp. 455–465). Lawrence Erlbaum Associates Publishers.

Dickmanns, E. D., & Zapp, A. (1987). Autonomous High Speed Road Vehicle Guidance by Computer Vision. *IFAC Proceedings Volumes*, *20*(5), 221–226. https://doi.org/10.1016/s1474-6670(17)55320-3

Diels, C., & Bos, J. E. (2016). Self-driving carsickness. *Applied Ergonomics*, *53*, 374–382. https://doi.org/10.1016/j.apergo.2015.09.009

Edelberg, R. (1972). Electrodermal Recovery Rate, Goal-Orientation, and Aversion. *Psychophysiology*, *9*(5), 512–520. https://doi.org/10.1111/j.1469-8986.1972.tb01805.x

Elbanhawi, M., Simic, M., & Jazar, R. (2015). In the Passenger Seat: Investigating Ride Comfort Measures in Autonomous Cars. *IEEE Intelligent Transportation Systems Magazine*, *7*(3), 4–17. https://doi.org/10.1109/MITS.2015.2405571

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 32–64. https://doi.org/10.1518/001872095779049543

Endsley, M. R., & Kiris, E. O. (1995). The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(2), 381–394. https://doi.org/10.1518/001872095779064555

Eriksson, J., & Svensson, L. (2015). *Tuning for Ride Quality in Autonomous Vehicle Application to Linear Quadratic Path Planning Algorithm* [Uppsala University]. http://www.teknat.uu.se/student

ERTRAC. (2017). *Automated Driving Roadmap*. European Road Transport Research Advisory Council. http://www.ertrac.org/uploads/documentsearch/id48/ERTRAC_Automated_Driving_2017.pdf

Etemad, A. (2021). *L3 Pilot Final Event*. https://l3pilot.eu/fileadmin/user_upload/Downloads/Final_Event/13102021/L3Pil

ot_Final_Event_presentation_01_Aria_Etemad_20211013.pdf

Etemad, A. (2022). *Hi-Drive: Addressing challenges towards the deployment of higher automation*. https://www.hi-drive.eu/

Etemad, A., Metzner, S., Inhülsen, H., Hildebrandt, B., Wark, T., Alessandretti, G., Knapp, A., Johansson, E., Söderman, M., Kessler, C., Andreone, L., Langenberg, J., Zlocki, A., & Rösener, C. (2017). *AdaptIVe Project Final Report*. AdaptIVe Consortium. http://adaptive-ip.eu/

Flemisch, F., Kelsch, J., Löper, C., Schieben, A., & Schindler, J. (2008). Automation spectrum , inner/outer compatibility and other potentially useful human factors concepts for assistance and automation. In D. de Waard, F. O. Flemisch, B. Lorenz, H. Oberheid, & K. A. Brookhuis (Eds.), *Human Factors for Assistance and Automation* (pp. 1–16). Shaker.

Ford-Motors. (2021). *Ford's 'Mother of All Road Trips' Tests BlueCruise Hands-Free Driving Ahead of Over-the-Air Push to F-150, Mustang Mach-E | Ford Media Center*. https://media.ford.com/content/fordmedia/fna/us/en/news/2021/04/14/ford-mother-of-all-road-trips-bluecruise-hands-free-driving.html

Foy, H. J., & Chapman, P. (2018). Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation. *Applied Ergonomics*, *73*, 90–99. https://doi.org/10.1016/j.apergo.2018.06.006

Gameiro da Silva, M. C. (2002). Measurements of comfort in vehicles. *Measurement Science and Technology*, *13*(6). https://doi.org/10.1088/0957-0233/13/6/201

General-Motors. (2021). *Super Cruise - Hands Free Driving | Cadillac Ownership*. https://www.cadillac.com/world-of-cadillac/innovation/super-cruise

General Motors. (1956). *The Story of Firebird II - "Three-Zero-Four."* General Motors Heritage Centre. https://www.gmheritagecenter.com/docs/gm-heritage-archive/historical-brochures/1956-firebird-II/1956_Firebird_II_Brochure.pdf

Gold, C., Berisha, I., & Bengler, K. (2015). Utilization of drivetime - Performing non-driving related tasks while driving highly automated. *Proceedings of the Human Factors and Ergonomics Society*. https://doi.org/10.1177/1541931215591360

Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). Take over! How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society*, *57*(1), 1938–1942. https://doi.org/10.1177/1541931213571433

Gold, C., Happee, R., & Bengler, K. (2018). Modeling take-over performance in level 3 conditionally automated vehicles. *Accident Analysis and Prevention*, *116*(October 2017), 3–13. https://doi.org/10.1016/j.aap.2017.11.009

Golding, J. F. (2016). Motion sickness. In J. M. Furman & T. Lempert (Eds.), *Handbook of Clinical Neurology* (Vol. 137, pp. 371–390). Elsevier. https://doi.org/10.1016/B978-0-444-63437-5.00027-3

Gonçalves, J., & Bengler, K. (2015). Driver State Monitoring Systems– Transferable Knowledge Manual Driving to HAD. *Procedia Manufacturing*, *3*, 3011–3016. https://doi.org/10.1016/j.promfg.2015.07.845

Grandjean, E. (1979). Fatigue in industry. *Occupational and Environmental Medicine*, *36*(3), 175–186. https://doi.org/10.1136/oem.36.3.175

Hale, A. R., Stoop, J., & Hommels, J. (1990). Human error models as predictors of accident scenarios for designers in road transport systems. *Ergonomics*, *33*(10–11), 1377–1387. https://doi.org/10.1080/00140139008925339

Hancock, P. A., & Caird, J. K. (1993). Experimental evaluation of a model of mental workload. *Human Factors*, *35*(3), 413–429. https://doi.org/10.1177/001872089303500303

Hart, S. G., & Bortolussi, M. R. (1984). Pilot errors as a source of workload. *Human Factors*, *26*(5), 545–556. https://doi.org/10.1177/001872088402600506

Hart, Sandra G., & Staveland, L. E. (1988a). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology: Measuring Mental Workload* (1st ed., Vol. 52, Issue C, pp. 139–183). Elsevier Ltd. https://doi.org/10.1016/S0166-4115(08)62386-9

Hart, Sandra G., & Staveland, L. E. (1988b). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*,

52(C), 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Hart, Sandra G., & Wickens, C. D. (1990). Workload Assessment and Prediction. In H. R. Booher (Ed.), *Manprint* (pp. 257–296). Springer Netherlands. https://doi.org/10.1007/978-94-009-0437-8_9

Hartwich, F., Beggiato, M., & Krems, J. F. (2018). Driving comfort, enjoyment and acceptance of automated driving–effects of drivers' age and driving style familiarity. *Ergonomics*, *61*(8), 1017–1032. https://doi.org/10.1080/00140139.2018.1441448

Heikoop, D. D., de Winter, J. C. F., van Arem, B., & Stanton, N. A. (2016). Psychological constructs in driving automation: a consensus model and critical comment on construct proliferation. *Theoretical Issues in Ergonomics Science*, *17*(3), 284–303. https://doi.org/10.1080/1463922X.2015.1101507

Heißing, B., & Ersoy, M. (2011). Ride Comfort and NVH. In B. Heißing & M. Ersoy (Eds.), *Chassis Handbook* (pp. 421–448). Vieweg+Teubner. https://doi.org/10.1007/978-3-8348-9789-3_5

Hill, L. B. K., Siebenbrock, A., Sollers, J. J., & Thayer, J. F. (2009). Are all measures created equal? Heart rate variability and respiration. *Biomedical Sciences Instrumentation*, *45*, 71–76.

Hjortskov, N., Rissén, D., Blangsted, A. K., Fallentin, N., Lundberg, U., & Søgaard, K. (2004). The effect of mental stress on heart rate variability and blood pressure during computer work. *European Journal of Applied Physiology*. https://doi.org/10.1007/s00421-004-1055-z

Honda EU. (2021). *Honda launches next generation Honda SENSING Elite safety system with Level 3 automated driving features*. https://hondanews.eu/eu/en/cars/media/pressreleases/329456/honda-launches-next-generation-honda-sensing-elite-safety-system-with-level-3-automated-driving-feat

Jersild, A. T. (1927). Mental set and shift. *Archives of Psychology*, *14*, *89*, 81.

Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science*, *4*(1), 71–115. https://doi.org/10.1016/S0364-0213(81)80005-5

Kahneman, D. (1973). Attention and Effort. In *Prentice-Hall series in experimental psychology*. Prentice-Hall. https://doi.org/10.2307/1421603

Kalsbeek, J. W. H. (1968). Measurement of mental work load and of acceptable load: Possible applications in industry. *International Journal of Production Research*, *7*(1), 33–45. https://doi.org/10.1080/00207546808929795

Kircher, K., Larsson, A., & Hultgren, J. A. (2014). Tactical driving behavior with different levels of automation. *IEEE Transactions on Intelligent Transportation Systems*, *15*(1), 158–167. https://doi.org/10.1109/TITS.2013.2277725

Klauer, S. G., Klauer, S. G., Dingus, T. a., Dingus, T. a., Neale, V. L., Neale, V. L., Sudweeks, J. D., Sudweeks, J. D., Ramsey, D. J., & Ramsey, D. J. (2006). The Impact of Driver Inattention On Near Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data. *Analysis*. https://doi.org/DOT HS 810 594

Körber, M., Baseler, E., & Bengler, K. (2018). Introduction matters: Manipulating trust in automation and reliance in automated driving. *Applied Ergonomics*, *66*, 18–31. https://doi.org/10.1016/j.apergo.2017.07.006

Kranjec, J., Beguš, S., Geršak, G., & Drnovšek, J. (2014). Non-contact heart rate and heart rate variability measurements: A review. In *Biomedical Signal Processing and Control* (Vol. 13, Issue 1, pp. 102–112). Elsevier. https://doi.org/10.1016/j.bspc.2014.03.004

Laborde, S., Mosley, E., & Thayer, J. F. (2017). Heart rate variability and cardiac vagal tone in psychophysiological research - Recommendations for experiment planning, data analysis, and data reporting. In *Frontiers in Psychology* (Vol. 8, p. 213). Frontiers. https://doi.org/10.3389/fpsyg.2017.00213

Lazarus, R. S. (1966). Psychological stress and the coping process. In *Psychological stress and the coping process.* McGraw-Hill.

Lenné, M. G., Triggs, T. J., & Redman, J. R. (1997). Time of day variations in driving performance. *Accident Analysis and Prevention*, *29*(4 SPEC. ISS.), 431–437. https://doi.org/10.1016/S0001-4575(97)00022-5

Liu, K., Green, P., & Liu, Y. (2019). Traffic and Ratings of Driver Workload: The Effect

of the Number of Vehicles and Their Distance Headways. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 2134–2138. https://doi.org/10.1177/1071181319631051

Lohani, M., Cooper, J. M., Erickson, G. G., Simmons, T. G., McDonnell, A. S., Carriero, A. E., Crabtree, K. W., & Strayer, D. L. (2021). No Difference in Arousal or Cognitive Demands Between Manual and Partially Automated Driving: A Multi-Method On-Road Study. *Frontiers in Neuroscience*, *15*. https://doi.org/10.3389/fnins.2021.577418

Lorenz, L., Kerschbaum, P., & Schumann, J. (2014). Designing take over scenarios for automated driving: How does augmented reality support the driver to get back into the loop? *Proceedings of the Human Factors and Ergonomics Society*. https://doi.org/10.1177/1541931214581351

Louw, T. (2017). *THE HUMAN FACTORS OF TRANSITIONS IN HIGHLY AUTOMATED DRIVING* [University of Leeds]. https://doi.org/10.13140/RG.2.1.2788.9760

Louw, T., Kuo, J., Romano, R., Radhakrishnan, V., Lenné, M. G., & Merat, N. (2019). Engaging in NDRTs affects drivers' responses and glance patterns after silent automation failures. *Transportation Research Part F: Traffic Psychology and Behaviour*, *62*, 870–882. https://doi.org/10.1016/j.trf.2019.03.020

Louw, T., Madigan, R., Carsten, O., & Merat, N. (2017). Were they in the loop during automated driving? Links between visual attention and crash potential. *Injury Prevention*, *23*, 281–286. https://doi.org/10.1136/injuryprev-2016-042155

Marquart, G., Cabrall, C., & de Winter, J. (2015). Review of Eye-related Measures of Drivers' Mental Workload. *Procedia Manufacturing*, *3*, 2854–2861. https://doi.org/10.1016/j.promfg.2015.07.783

Martens, M. H., Pauwelussen, J., Schieben, A., Merat, N., Jamson, A. H., Caci, P., Paulwelussen, Schieben, A., Merat, N., Jamson, A. H., & Caci. (2008). Human Factors' aspects in automated and semi-automatic transport systems: State of the art. In *Deliverable 3.2.1 of the Citymobil project* (Issue May 2006). http://www.citymobil-project.eu/downloadables/Deliverables/D3.2.1-PU-Human Factors aspects-CityMobil.pdf

Mathôt, S. (2018). Pupillometry: Psychology, Physiology, and Function. *Journal of Cognition*, *1*(1), 1–23. https://doi.org/10.5334/joc.18

Matthews, G., & Desmond, P. A. (2001). Stress and driving performance: Implications for design and training. In *Stress, workload, and fatigue.* (pp. 211–231). Lawrence Erlbaum Associates Publishers.

May, J. F., & Baldwin, C. L. (2009). Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. *Transportation Research Part F: Traffic Psychology and Behaviour*, *12*(3), 218–224. https://doi.org/10.1016/j.trf.2008.11.005

Mehler, B., Reimer, B., Coughlin, J., & Dusek, J. (2009). Impact of Incremental Increases in Cognitive Workload on Physiological Arousal and Performance in Young Adult Drivers. *Transportation Research Record: Journal of the Transportation Research Board*, *2138*, 6–12. https://doi.org/10.3141/2138-02

Mehler, B., Reimer, B., & 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. In *Human Factors* (Vol. 54, Issue 3, pp. 396–412). SAGE PublicationsSage CA: Los Angeles, CA. https://doi.org/10.1177/0018720812442086

Meister, D. (1976). *Behavioral foundations of system development*. Wiley. https://books.google.co.uk/books/about/Behavioral_foundations_of_system_develop.html?id=AXBbAAAAMAAJ&redir_esc=y

Merat, N., Jamson, A. H., Lai, F. C. H., & Carsten, O. (2012). Highly automated driving, secondary task performance, and driver state. *Human Factors*, *54*(5), 762–771. https://doi.org/10.1177/0018720812442087

Merat, N., Jamson, A. H., Lai, F. C. H., Daly, M., & Carsten, O. M. J. (2014). Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*(PB), 274–282. https://doi.org/10.1016/j.trf.2014.09.005

Merat, N., Seppelt, B., Louw, T., Engström, J., Lee, J. D., Johansson, E., Green, C. A., Katazaki, S., Monk, C., Itoh, M., McGehee, D., Sunda, T., Unoura, K., Victor, T., Schieben, A., & Keinath, A. (2018). The “Out-of-the-Loop” concept in

automated driving: proposed definition, measures and implications. *Cognition, Technology & Work*, 1–12. https://doi.org/10.1007/s10111-018-0525-8

Michon, J. (1985). A critical view of driver behavior models: what do we know, what should we do? In L. Evans & R. C. Schwing (Eds.), *Human behavior and traffic safety* (pp. 485–520). Plenum Press. https://doi.org/10.1007/978-1-4613-2173-6

Mioch, T., Kroon, L., & Neerincx, M. A. (2017). Driver Readiness Model for Regulating the Transfer from Automation to Human Control. *Proceedings of the 22nd International Conference on Intelligent User Interfaces - IUI '17*, *March*, 205–213. https://doi.org/10.1145/3025171.3025199

Molnar, L. J., Ryan, L. H., Pradhan, A. K., Eby, D. W., St. Louis, R. M., & Zakrajsek, J. S. (2018). Understanding trust and acceptance of automated vehicles: An exploratory simulator study of transfer of control between automated and manual driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *58*, 319–328. https://doi.org/10.1016/j.trf.2018.06.004

Money, K. E., & Cheung, B. S. (1983). Another function of the inner ear: Facilitation of the emetic response to poisons. *Aviation Space and Environmental Medicine*, *54*(3), 208–211.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134–140. https://doi.org/10.1016/S1364-6613(03)00028-7

Mulder, G. (1986). The Concept and Measurement of Mental Effort. In *Energetics and Human Information Processing* (pp. 175–198). https://doi.org/10.1007/978-94-009-4448-0_12

Mulder, M., Abbink, D. A., & Boer, E. R. (2012). Sharing control with haptics: Seamless driver support from manual to automatic control. *Human Factors*, *54*(5), 786–798. https://doi.org/10.1177/0018720812443984

Müller, A. L., Fernandes-Estrela, N., Hetfleisch, R., Zecha, L., & Abendroth, B. (2021). Effects of non-driving related tasks on mental workload and take-over times during conditional automated driving. *European Transport Research Review*, *13*(1), 1–15. https://doi.org/10.1186/s12544-021-00475-5

Myers, A. M., Paradis, J. A., & Blanchard, R. A. (2008). Conceptualizing and

Measuring Confidence in Older Drivers: Development of the Day and Night Driving Comfort Scales. *Archives of Physical Medicine and Rehabilitation*, *89*(4), 630–640. https://doi.org/10.1016/j.apmr.2007.09.037

Naujoks, F., Forster, Y., Wiedemann, K., & Neukum, A. (2016). *Speech improves human-automation cooperation in automated driving*. https://doi.org/10.18420/muc2016-ws08-0007

Nof, S. Y. (2009). Automation: What It Means to Us Around the World. In *Springer Handbook of Automation* (pp. 13–52). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-78831-7_3

Norman, D. A. (1983). Some Observations on Mental Models. In D. Gentner & A. L. Stevens (Eds.), *Mental Models* (1st ed., pp. 7–14). Psychology Press. https://doi.org/10.4324/9781315802725-5

Norman, S., & Orlady, H. (1989). Flight Deck Automation: Promises and Realities. *NASA/FAA/Industry Workshop*. https://ntrs.nasa.gov/api/citations/19900004068/downloads/19900004068.pdf

O'Donell, R., & Eggemeier, T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance* (Vol. 2, pp. 1–49). John Wiley & Sons.

Orsila, R., Virtanen, M., Luukkaala, T., Tarvainen, M., Karjalainen, P., Viik, J., & Savinainen, M. (2008). Perceived mental stress and reactions in heart rate variability—a pilot study among employees of an electronics company. *International Journal of Occupational Safety and Ergonomics*, *14*(3), 275–283. https://doi.org/10.1080/10803548.2008.11076767

Otzenberger, H., Gronfier, C., Simon, C., Charloux, A., Ehrhart, J., Piquard, F., & Brandenberger, G. (1998). Dynamic heart rate variability: A tool for exploring sympathovagal balance continuously during sleep in men. *American Journal of Physiology - Heart and Circulatory Physiology*, *275*(3 44-3). https://doi.org/10.1152/ajpheart.1998.275.3.h946

Palinko, O., Kun, A. L., Shyrokov, A., & Heeman, P. (2010). Estimating cognitive load using remote eye tracking in a driving simulator. *Eye Tracking Research and Applications Symposium (ETRA)*, 141–144.

https://doi.org/10.1145/1743666.1743701

Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse and Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(2), 230–253.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans.*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs. *Journal of Cognitive Engineering and Decision Making*, *2*(2), 140–160. https://doi.org/10.1518/155534308X284417

Patel, M., Lal, S. K. L., Kavanagh, D., & Rossiter, P. (2011). Applying neural network analysis on heart rate variability data to assess driver fatigue. *Expert Systems with Applications*, *38*(6), 7235–7242. https://doi.org/10.1016/j.eswa.2010.12.028

Pitts, M. J., Williams, M. A., Wellings, T., & Attridge, A. (2009). Assessing subjective response to haptic feedback in automotive touchscreens. *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '09*, 11. https://doi.org/10.1145/1620509.1620512

Radhakrishnan, V., Merat, N., Louw, T., Lenné, M. G., Romano, R., Paschalidis, E., Hajiseyedjavadi, F., Wei, C., & Boer, E. R. (2020). Measuring drivers' physiological response to different vehicle controllers in highly automated driving (HAD): Opportunities for establishing real-time values of driver discomfort. *Information (Switzerland)*, *11*(8), 390. https://doi.org/10.3390/INFO11080390

Radlmayr, J., Gold, C., Lorenz, L., Farid, M., & Bengler, K. (2014). How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. *Proceedings of the Human Factors and Ergonomics Society*, *2014-Janua*, 2063–2067. https://doi.org/10.1177/1541931214581434

Rajamani, R., Tan, H. S., Law, B. K., & Zhang, W. Bin. (2000). Demonstration of integrated longitudinal and lateral control for the operation of automated vehicles in platoons. *IEEE Transactions on Control Systems Technology*, *8*(4), 695–708.

https://doi.org/10.1109/87.852914

Ranney, T. A. (1994). Models of driving behavior: A review of their evolution. *Accident Analysis and Prevention*, *26*(6), 733–750. https://doi.org/10.1016/0001-4575(94)90051-5

Rasmussen, J. (1979). On the structure of knowledge-a morphology of metal models in a man-machine system context. In *Risø-M No. 2192*. Risø National Laboratory.

Rasmussen, J. (1983). Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models. *IEEE Transactions on Systems, Man and Cybernetics*, *SMC-13*(3), 257–266. https://doi.org/10.1109/TSMC.1983.6313160

Rauch, N., Kaussner, A., Boverie, S., & Giralt, A. (2009). HAVEit: Report on Driver Assessment Methodology. *(Deliverable D32.1)*. http://haveit-eu.org/LH2Uploads/ItemsContent/24/HAVEIt_212154_D32.1_public_version.pdf

Recarte, M. Á., Ṕerez, E., Conchillo, Á., & Nunes, L. M. (2008). Mental workload and visual impairment: Differences between pupil, blink, and subjective rating. *Spanish Journal of Psychology*, *11*(2), 374–385. https://doi.org/10.1017/s1138741600004406

Richards, L. G., Jacobson, I. D., & Kuhlthau, A. R. (1978). What the passenger contributes to passenger comfort. *Applied Ergonomics*, *9*(3), 137–142. https://doi.org/10.1016/0003-6870(78)90003-0

Roscoe, A. H. (1978). Stress and workload in pilots. *Aviation, Space, and Environmental Medicine*, *49*, 630–636.

Rusnock, C. F., & Borghetti, B. J. (2016). Workload profiles: A continuous measure of mental workload. *International Journal of Industrial Ergonomics*. https://doi.org/10.1016/j.ergon.2016.09.003

SAE International. (2021). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. https://www.sae.org/standards/content/j3016_202104/

Sarter, N. B., & Woods, D. D. (1992). Pilot Interaction With Cockpit Automation: Operational Experiences With the Flight Management System. *The International*

Journal of Aviation Psychology, *2*(4), 303–321. https://doi.org/10.1207/s15327108ijap0204_5

Seeing Machines Inc. (2022). *Guardian Backup-driver Monitoring System (BdMS)*. https://www.seeingmachines.com/wp-content/uploads/2018/12/SeeingMachines-Backup-Driver-Monitoring-Sales-Sheet-08-web-revised.pdf

Sheridan, T. B., & Parasuraman, R. (2005). Human-Automation Interaction. In *Reviews of Human Factors and Ergonomics* (Vol. 1, Issue 1). SAGE PublicationsSage CA: Los Angeles, CA. https://doi.org/10.1518/155723405783703082

Shields, S. A., MacDowell, K. A., Fairchild, S. B., & Campbell, M. L. (1987). Is Mediation of Sweating Cholinergic, Adrenergic, or Both? A Comment on the Literature. *Psychophysiology*, *24*(3), 312–319. https://doi.org/10.1111/j.1469-8986.1987.tb00301.x

Shimomura, Y., Yoda, T., Sugiura, K., Horiguchi, A., Iwanaga, K., & Katsuura, T. (2008). Use of Frequency Domain Analysis of Skin Conductance for Evaluation of Mental Workload. *Journal of PHYSIOLOGICAL ANTHROPOLOGY*, *27*(4), 173–177. https://doi.org/10.2114/jpa2.27.173

Siebert, F. W., Oehl, M., Höger, R., & Pfister, H. R. (2013). Discomfort in Automated Driving - The Disco-Scale. *Communications in Computer and Information Science*, *374*(PART II), 337–341. https://doi.org/10.1007/978-3-642-39476-8_69

Silva, F. P. da. (2014). Mental Workload, Task Demand and Driving Performance: What Relation? *Procedia - Social and Behavioral Sciences*, *162*, 310–319. https://doi.org/10.1016/j.sbspro.2014.12.212

Slater, K. (1985). *Human Comfort*. Charles C Thomas Publishers.

Slater, K. (1986). The assessment of comfort. *Journal of the Textile Institute*, *77*(3), 157–171. https://doi.org/10.1080/00405008608658406

Spector, R. H. (1990). The pupils. In H. K. Walker, W. D. Hall, & J. W. Hurst (Eds.), *Clinical Methods: The History, Physical, and Laboratory Examinations* (3rd ed.). Butterworths. https://www.ncbi.nlm.nih.gov/books/NBK381/

Stanton, N. A., & Marsden, P. (1996). From fly-by-wire to drive-by-wire: Safety implications of automation in vehicles. *Safety Science*, *24*(1), 35–49. https://doi.org/10.1016/S0925-7535(96)00067-7

Stanton, N. A., & Young, M. S. (2000). A proposed psychological model of driving automation. *Theoretical Issues in Ergonomics Science*, *1*(4), 315–331. https://doi.org/10.1080/14639220052399131

Stapel, J., Mullakkal-Babu, F. A., & Happee, R. (2019). Automated driving reduces perceived workload, but monitoring causes higher cognitive load than manual driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *60*, 590–605. https://doi.org/10.1016/j.trf.2018.11.006

Steck, F., Kolarova, V., Bahamonde-Birke, F., Trommer, S., & Lenz, B. (2018). How Autonomous Driving May Affect the Value of Travel Time Savings for Commuting. *Transportation Research Record*, *2672*(46), 11–20. https://doi.org/10.1177/0361198118757980

Stern, J. A., Boyer, D., & Schroeder, D. (1994). Blink rate: A possible measure of fatigue. *Human Factors*, *36*(2), 285–297. https://doi.org/10.1177/001872089403600209

Summala, H. (2007). Modelling driver behaviour in automotive environments. In P. C. Cacciabue (Ed.), *Modelling Driver Behaviour in Automotive Environments: Critical Issues in Driver Interactions with Intelligent Transport Systems* (pp. 189–207). Springer London. https://doi.org/https://doi.org/10.1007/978-1-84628-618-6_11

Taelman, J., Vandeput, S., Spaepen, A., & Van Huffel, S. (2008). Influence of mental stress on heart rate and heart rate variability. *IFMBE Proceedings*, *22*, 1366–1369. https://doi.org/10.1007/978-3-540-89208-3_324

Taylor, S., Jaques, N., Chen, W., Fedor, S., Sano, A., & Picard, R. (2015). Automatic identification of artifacts in electrodermal activity data. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, *2015-Novem*, 1934–1937. https://doi.org/10.1109/EMBC.2015.7318762

Tesla. (2021). *Autopilot* | *Tesla*. https://www.tesla.com/autopilot

Thakurta, K., Koester, D., Bush, N., & Bachle, S. (1995, February 1). Evaluating short and long term seating comfort. *SAE Technical Papers*. https://doi.org/10.4271/950144

Treisman, M. (1977). Motion sickness: An evolutionary hypothesis. *Science*, *197*(4302), 493–495. https://doi.org/10.1126/science.301659

Trommer, S., Kolarova, V., Fraedrich, E., Kröger, L., Kickhöfer, B., Kuhnimhof, T., Lenz, B., & Phleps, P. (2016). Autonomous Driving - The Impact of Vehicle Automation on Mobility Behaviour. In *Ifmo- Institute for Mobility Research* (Issue December). www.ifmo.de

van Gent, P., Farah, H., van Nes, N., & van Arem, B. (2018). Heart Rate Analysis for Human Factors: Development and Validation of an Open Source Toolkit for Noisy Naturalistic Heart Rate Data. *Humanist 2018 Conference*, *June*.

Veltman, J. A., & Gaillard, A. W. K. (1996). Physiological indices of workload in a simulated flight task. *Biological Psychology*, *42*(3), 323–342. https://doi.org/10.1016/0301-0511(95)05165-1

Vogel, H., Kohlhaas, R., & von Baumgarten, R. J. (1982). Dependence of motion sickness in automobiles on the direction of linear acceleration. *European Journal of Applied Physiology and Occupational Physiology*, *48*(3), 399–405. https://doi.org/10.1007/BF00430230

Wada, T., Konno, H., Fujisawa, S., & Doi, S. (2012). Can passengers' active head tilt decrease the severity of carsickness?: Effect of head tilt on severity of motion sickness in a lateral acceleration environment. *Human Factors*, *54*(2), 226–234. https://doi.org/10.1177/0018720812436584

Walker, F. (2021). *To trust or not to trust? : assessment and calibration of driver trust in automated vehicles* [University of Twente]. https://doi.org/10.3990/1.9789055842766

Wertheim, A. H., & Hogema, J. H. (1997). *Thresholds, comfort and maximum acceptability of horizontal accelerations associated with car driving*.

Wickens, C. D. (1984). Processing Resources in Attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 63–102). Academic Press.

https://doi.org/R. Parasuraman, R. Davies

Wickens, C. D., Gutzwiller, R. S., & Santamaria, A. (2015). Discrete task switching in overload: A meta-analyses and a model. *International Journal of Human Computer Studies*, *79*, 79–84. https://doi.org/10.1016/j.ijhcs.2015.01.002

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2013). *Engineering Psychology and Human Performance* (4th ed.). Psychology Press. https://doi.org/10.4324/9781315665177

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2016). Engineering psychology and human performance. In *Engineering Psychology and Human Performance* (4th ed.). HarperCollins Publishers. https://doi.org/10.4324/9781003177616

Wilkie, R. M., & Wann, J. P. (2003). Eye-movements aid the control of locomotion. *Journal of Vision*, *3*(11), 677–684. https://doi.org/10.1167/3.11.3

Williams, T. J. (2009). Advances in Industrial Automation: Historical Perspectives. In *Springer Handbook of Automation* (pp. 5–11). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-78831-7_2

Wilson, G. F. (2002). An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures. In *International Journal of Aviation Psychology* (Vol. 12, Issue 1 SPEC, pp. 3–18). Lawrence Erlbaum Associates, Inc. https://doi.org/10.1207/S15327108IJAP1201_2

Woolridge, E., & Chan-Pensley, J. (2020). *Measuring User's Comfort in Autonomous Vehicles*. HumanDrive Consortium. https://humandrive.co.uk/wp-content/uploads/2020/07/HumanDrive-HF-Comfort-White-Paper-v1.1.pdf

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*, 459–482. https://doi.org/10.1037/h0073415

Young, M. S., & Stanton, N. A. (2002a). Attention and automation: New perspectives on mental underload and performance. *Theoretical Issues in Ergonomics Science*, *3*(2), 178–194. https://doi.org/10.1080/14639220210123789

Young, M. S., & Stanton, N. A. (2002b). Malleable Attentional Resources Theory: A

New Explanation for the Effects of Mental Underload on Performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *44*(3), 365–375. https://doi.org/10.1518/0018720024497709

Young, M. S., Stanton, N. A., & Harris, D. (2007). Driving automation: Learning from aviation about design philosophies. *International Journal of Vehicle Design*, *45*(3), 323–338. https://doi.org/10.1504/IJVD.2007.014908

Zeeb, K., Buchner, A., & Schrauf, M. (2016). Is take-over time all that matters? the impact of visual-cognitive load on driver take-over quality after conditionally automated driving. *Accident Analysis and Prevention*, *92*, 230–239. https://doi.org/10.1016/j.aap.2016.04.002

2 ANALYSING ELECTRODERMAL ACTIVITY IN DYNAMIC DRIVING ENVIRONMENTS

ABSTRACT: This methodology proposes a novel and rapid automated motion artefact detection algorithm for electrodermal activity (EDA) data in dynamic environments. Building on a well validated shape-based artefact detection method, we improved the detection algorithm by adding an additional slope constraint. We also proposed a novel approach to account for missing or removed data, using modified-Akima (Makima) interpolation technique. Upon visual inspection, our shape and slope-based algorithms, along with the Makima interpolation techniques, detects and corrects from motion artefacts in ambulatory environments, while preserving the shape of the EDA signal.

2.1 Introduction

Electrodermal activity (EDA) refers to the autonomic changes in the electrical properties of the skin, which is observed by studying the current flow between two points of contact on the skin's surface, when an electrical potential is applied (Braithwaite et al., 2015). While the primary function of eccrine sweat glands is thermoregulation, the sweat glands located on the palmar and plantar regions of the body is more responsive to psychologically significant stimuli than thermal stimuli (Cacioppo et al., 2007a). EDA is directly related to sympathetic nervous system (SNS) activity, with strong correlation to psychological stimuli such as arousal, stress, anxiety or attentional demand (Cacioppo et al., 2007a; Kübler et al., 2014). EDA-based measures have been increasingly used to understand driver state in driving studies, given its sensitivity in detecting psychological constructs such as stress, workload, arousal and discomfort (Foy & Chapman, 2018; Kübler et al., 2014; Radhakrishnan et al., 2020).

However, a major concern with the usage of EDA signals in driving-related studies is its susceptibility to motion artefacts and noise (Taylor et al., 2015), especially when conducted in a dynamic environment such as a real-world vehicle or a motion-based driving simulator. Artefacts can be attributed to electronic noise, or due to movement or, adjustments that results in variation in contact between the skin and electrodes (Taylor et al., 2015).

Traditionally, most studies as well as dedicated software packages for analysing EDA signals were designed for fixed laboratory environments, and as such, robust motion artefact detection were not a necessity owing to low prevalence of noise and artefacts within such an environment. However, in a more dynamic environment, neglecting or not accurately accounting for noise/artefacts in the signal can result in incorrect results, such as when noise is mistaken/misinterpreted as skin conductance response (SCR) events. This paper aims to address this issue, by proposing a robust and real-time artefact removal algorithm that can be used to pre-process the EDA signal, especially when used in dynamic environments such as road vehicles or motion-based driving simulators.

2.1.1 Related work

An EDA signal typically consists of two components, a slowly evolving tonic component called skin conductance level (SCL, see **Figure 2.1**) and a rapidly evolving phasic component called Skin conductance response (SCR; Braithwaite et al., 2015). Boucsein (2012) provided a comprehensive and detailed overview on the characteristics and shape of an SCR signal. A typical SCR event (see **Figure 2.1**) lasts for 1-5 s, has an amplitude of at least 0.01 µS, a latency of 1-3 s from stimulus onset, and is characterised by a steep onset followed by exponential decay (Boucsein, 2012; Braithwaite et al., 2015). However, if two SCR events happen closer in time, it is not necessary that the first SCR event decays before the onset of the second SCR event (Taylor et al., 2015). The above-mentioned characteristics of SCRs act as a guideline for distinguishing SCR events from noise.

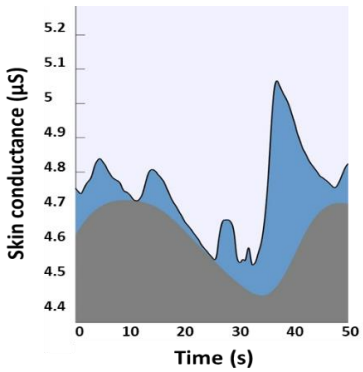

Figure 2.1. A typical EDA signal, with the phasic component or SCRs (dark blue) and the tonic component or SCL (grey), plotted on MATLAB R2016a using ledalab software (Benedek & Kaernbach, 2010)

While some studies implemented manual inspection of EDA signal, with the help of expert labellers, to differentiate noise from the signal, this is not feasible for large scale studies, or for real-time analysis of EDA signals, especially within the driving context. Most studies on EDA have tried to address the issue of artefact removal by using exponential smoothing of the EDA signal (Hernandez et al., 2011) or by using lowpass or bandpass filters (Benedek & Kaernbach, 2010; Setz et al., 2010). However, using a lowpass filter with an acceptable cut-off frequency of around 1 Hz, only smoothens the signal for smaller variations, and does not completely remove all the artefacts, especially the large magnitude artefacts that can be caused by movement. Such artefacts can get mistaken as genuine physiological responses, as observed by Taylor et al. (2015) in their paper.

Another approach proposed by researchers is the usage of accelerometer data to identify and remove periods with relatively high movement (Braithwaite et al., 2015; Taylor et al., 2015). However, the driving task involves hand and foot movements, and generally, periods of high movements such as a takeover scenario in highly automated driving (HAD), also coincides with high physiological activation and arousal. The EDA data during such periods, can inform us of drivers' psychological state, and therefore, warrants a deeper understanding. Taylor et al. (2015) used accelerometer data along with machine learning based approach, where the EDA features such as those based on amplitude and mean SCR values, along with wavelet coefficients were extracted and analysed using a binary classifier (valid or invalid) for 5 s segments in the signal. However, their paper lacked clarity in terms of what governing criteria was used to determine if a segment was valid or invalid. As observed by Kleckner et al. (2018), this lack of transparency makes it hard for researchers to understand why the data segments are classified as valid or invalid, and inhibits them from giving different instructions to the participants to improve data collection quality. Moreover, such machine learning and wavelet transformation-based approaches are complex and computationally intensive, and therefore, not ideal for real-time usage.

Kocielnik et al. (2013) and Kikhia et al. (2016) have used recommendations from Boucsein (2012) regarding the shape and characteristics of a typical SCR signal, to identify noise. They discarded data segments where the signal increased by more than 20% or decreased by more than 10%, within a 1 s time window, and their

approach was verified using visual inspection by experts (Kikhia et al., 2016; Kocielnik et al., 2013). Kikhia et al. (2016) suggested using linear interpolation for filling signal gaps due to noise. However, given that an SCR follows a steep onset followed by an exponential decay, we believe the linear interpolation approach is inadequate for this purpose. In this paper, we propose a simple and improved shape-based motion artefact removal algorithm for EDA signal. We further propose a novel approach for filling up signal gaps due to motion artefacts, which best captures the characteristics of an SCR signal.

The remainder of this paper is structured as follows: Section 2 introduces our novel methodological proposal for motion artefact removal in EDA signals. In Section 3, we conducted a case study using physiological data collected from a previous study in dynamic driving simulator environment, as well as discussions around it. Section 4 is the conclusion section.

2.2 Methodological proposal

The cleaning of EDA signals to remove motion artefacts is a two-step process. The first step involves an artefact detection algorithm, that correctly detects and removes artefacts from the EDA signal. In the second step, we fill in the missing data using appropriate interpolation technique, given the total missing data is within an acceptable range of the entire dataset (usually under 5%; Kikhia et al., 2016).

2.2.1 Artefact detection

As mentioned in the introduction, the three main sources of artefacts in EDA signals are: electronic noise, loss of contact between the skin and electrode surface, or hand and body movements. The electronic noise generated by AC power systems is generally between 50/60 Hz and can be filtered using a lowpass filter such as a Butterworth filter, with a cut-off frequency of 1-3 Hz (Boucsein, 2012). In case of loss in contact between the skin surface and the electrodes, we used a threshold value of 0.05 µS, which is well below realistic values for an EDA signal (Kleckner et al., 2018). When 90% of the values within a 5 s window does not exceed this threshold value, the data for that 5 s window is removed (Kikhia et al., 2016; Kleckner et al., 2018).

However, the EDA device should be carefully checked for zero error, and the threshold value for loss of contact between the electrodes and skin surface.

Removing motion and motion artefacts due to hand and body movements is more complex than the other two sources of artefacts. Based on the shape of an typical SCR as described by Boucsein (2012), both Kocielnik et al. (2013) and Kikhia et al. (2016) suggested that an EDA signal cannot increase by more than 20%, or decrease by more than 10%, within a 1 s time window. However, there could be a possibility that 1 s time window is too short to accurately reveal motion artefacts, or there could be residual noise in the preceding or succeeding windows. Therefore, we propose using 3 different time windows of 1 s, 2 s and 3 s. The method used in finding maximum increase or decrease in EDA values in each time window is explained in detail below.

Let y be the EDA value at the start of the time window (t = 0). Maximum and minimum possible values of y after 1 s (t = 1), as proposed by Kocielnik et al. (2013) and Kikhia et al. (2016):

$$y_{max}(1) = 1.2 * y(0) \quad (1)$$

$$y_{min}(1) = 0.9 * y(0) \quad (2)$$

Given the maximum and minimum values after 1 s, maximum and minimum possible values of y after 2 s (t = 2) would be given by:

$$y_{max}(2) = 1.2 * y_{max}(1) \quad (3)$$

$$y_{min}(2) = 0.9 * y_{min}(1) \quad (4)$$

Combining equation (1) and equation (3), and equation (2) and equation (4), we get:

$$y_{max}(2) = 1.44 * y(0) \quad (5)$$

$$y_{min}(2) = 0.81 * y(0) \quad (6)$$

Similarly, the maximum and minimum possible values of y after 3 s (t =3) is given by:

$$y_{max}(3) = 1.728 * y(0) \quad (7)$$

$$y_{min}(3) = 0.729 * y(0) \quad (8)$$

The above equations were used as limiting criteria with 3 different time windows of length 1 s, 2 s and 3 s. A data point is initially compared to each of the data points, succeeding it within a 1 s time window, followed by each of the data points succeeding it within a 2 s time window, and finally, each of the data points succeeding it in a 3 s time window, and the time windows move continuously across the signal. If any of the succeeding data points failed to meet the limiting criteria in any of the 3 time windows, it was replaced with not a number (NaN) values. To ensure the starting data point is not noisy, we suggest trimming the data to a few seconds after the recording is started, ideally from a time when the participant is stationary. However, when the EDA signal is recorded at higher sampling rates, motion artefacts of very short time duration can corrupt the signal. The shape based algorithm mentioned above might not be sufficient to remove such artefacts, given its short time duration, and the time windows of shape-based algorithms being a minimum of 1 s. Therefore, in the next section, we explore how an additional slope based constraint can be used to remove such artefacts from the EDA signal.

2.2.1.1 Filtering using maximum and minimum slope for the signal

Initially, the EDA signal was filtered for noise using the shape-based algorithm described by equations (1) and (2). In addition to the shape constraint mentioned above, we use a slope constraint, derived from the recommendations based on Boucsein (2012) and Kikhia et al. (2016). Accordingly, the EDA signal does not increase by more than 20% in 1 s window or decrease less than 10% in 1 s window. While a direct value for maximum (0.2 µS/s) and minimum slope (-0.1 µS/s) within that 1 s window, as derived from Boucsein (2012) could be used as the threshold values, EDA samples are recorded at frequencies far greater than 1 Hz. As such, there would be multiple data points (n) within a 1 s time window, depending on the frequency of recording. Therefore, the slope of the line of best fit for any given 1 s time window is given by:

$$m = \frac{\sum_{i=1}^{n}\left(\left(x_i - \frac{\sum_{i=1}^{n} x_i}{n}\right)\left(y_i - \frac{\sum_{i=1}^{n} y_i}{n}\right)\right)}{\sum_{i=1}^{n}\left(x_i - \frac{\sum_{i=1}^{n} x_i}{n}\right)^2} \quad (9)$$

This results in generating a set of slopes, from which the 95th and 5th percentile was computed. The 95th and 5th percentile values were used as the maximum and minimum slopes respectively, for any two data points in the signal, within a 1 s time window. However, if the 95th and 5th percentile of the slope is lower than two times the maximum slope (< 0.4 µS/s) or greater than minimum slope (> -0.2 µS/s) derived from Boucsein (2012) and Kikhia et al. (2016)'s recommendations, then these values ([0.4 -0.2]) are used as the maximum and minimum slopes. In the next section, we investigate different interpolation techniques that can be used to interpolate the gaps due to the removed noisy data, in the EDA signal.

2.2.2 Missing data treatment and interpolation

An EDA signal is generally processed as a continuous signal over time, and therefore, any missing data can lead to errors. As a general rule, interpolation of missing data is performed if the percentage of missing data is less than 10**%** of the entire data segment (Dong & Peng, 2013).

As mentioned earlier, previous research have recommended using linear or moving average interpolation techniques to fill missing data in an EDA signal (Kikhia et al., 2016; Kleckner et al., 2018). A linear interpolant is a straight line between the two end nodes and is a special case of polynomial interpolation. Moving average filters generally incorporate a moving window of length 'n', where the missing values are replaced with n-point means within such a window. However, neither of these interpolation techniques accurately capture the steep rise and exponential decay observed in an SCR.

More effective interpolation techniques are based on piecewise cubic polynomials, rather that linear polynomials (Moler, 2004). Cubic spline is a commonly used piecewise cubic interpolation technique. While it retains the shape characteristics of an SCR signal, one of the unwanted consequences is that it can create unwanted undulations between two points, resulting in larger errors (Dan et al., 2020). A piecewise cubic Hermite interpolation polynomial (PCHIP) addresses this issue by aggressively reducing any undulations between two datapoints. However, it can aggressively flatten the signal, especially around troughs and peaks of an SCR event. Akima (1970) proposed a piecewise cubic Hermite interpolation technique, that that avoids excessive local undulations, while not as aggressively as PCHIP.

Let each frame or node in the signal be represented by x, and the corresponding value of the EDA signal at that frame be y. For each interval $[x_n\ x_{n+1})$, cubic Hermite interpolation finds a cubic polynomial that finds a corresponding data value of y_n, y_{n+1}. Additionally, it also provides derivatives $\Delta y_n, \Delta y_{n+1}$ for the nodes x_n, x_{n+1} respectively.

Let $\Delta y_n = (y_{n+1} - y_n)/(x_{n+1} - x_n)$ be the slope at interval $[x_n\ x_{n+1})$. Then, Akima's derivative at x_n is defined as:

$$\partial_n = \frac{|\Delta y_{n+1} - \Delta y_n|\Delta y_{n-1} + |\Delta y_{n-1} - \Delta y_{n-2}|\Delta y_n}{|\Delta y_{n+1} - \Delta y_n| + |\Delta y_{n-1} - \Delta y_{n-2}|} \quad (10)$$

Where ∂_n represents a weighted average between the slopes Δy_{n-1} and Δy_n of the intervals $[x_{n-1}\ x_n)$ and $[x_n\ x_{n+1})$.

Let $\omega_1 = |\Delta y_{n+1} - \Delta y_n|$ and $\omega_2 = |\Delta y_{n-1} - \Delta y_{n-2}|$ be the weights. Therefore, equation (10) becomes:

$$\partial_n = \frac{\omega_1}{\omega_1 + \omega_2}\Delta y_{n-1} + \frac{\omega_2}{\omega_1 + \omega_2}\Delta y_n \quad (11)$$

The Akima's derivative ∂_n is computed using the 5 points x_{n-2}, x_{n-1}, x_n, x_{n+1}, x_{n+2}. For the edge points x_1, x_n, it requires the slopes of previous/next data points, which does not exist. Therefore, Akima (1970) proposed using quadratic extrapolation to interpolate the end nodes.

However, the interpolant to switches to a different formula for edge cases, and also, can result in overshoots or undershoots when data is constant for more than 2 consecutive nodes. In order to overcome this problem, MATLAB R2019a software (Mathworks, 2019) uses a modified Akima or Makima interpolation technique with slightly modified weights, where:

$$\omega_1 = |\Delta y_{n+1} - \Delta y_n| + \frac{|\Delta y_{n+1} + \Delta y_n|}{2} \quad (12)$$

$$\omega_2 = |\Delta y_{n-1} - \Delta y_{n-2}| + \frac{|\Delta y_{n-1} + \Delta y_{n-2}|}{2} \quad (13)$$

To best preserve the shape of an SCR event, the Makima interpolation technique was used to fill any consecutive data gaps created for up to 5 s. **Figure 2.2** depicts the flow chart of the proposed algorithm.

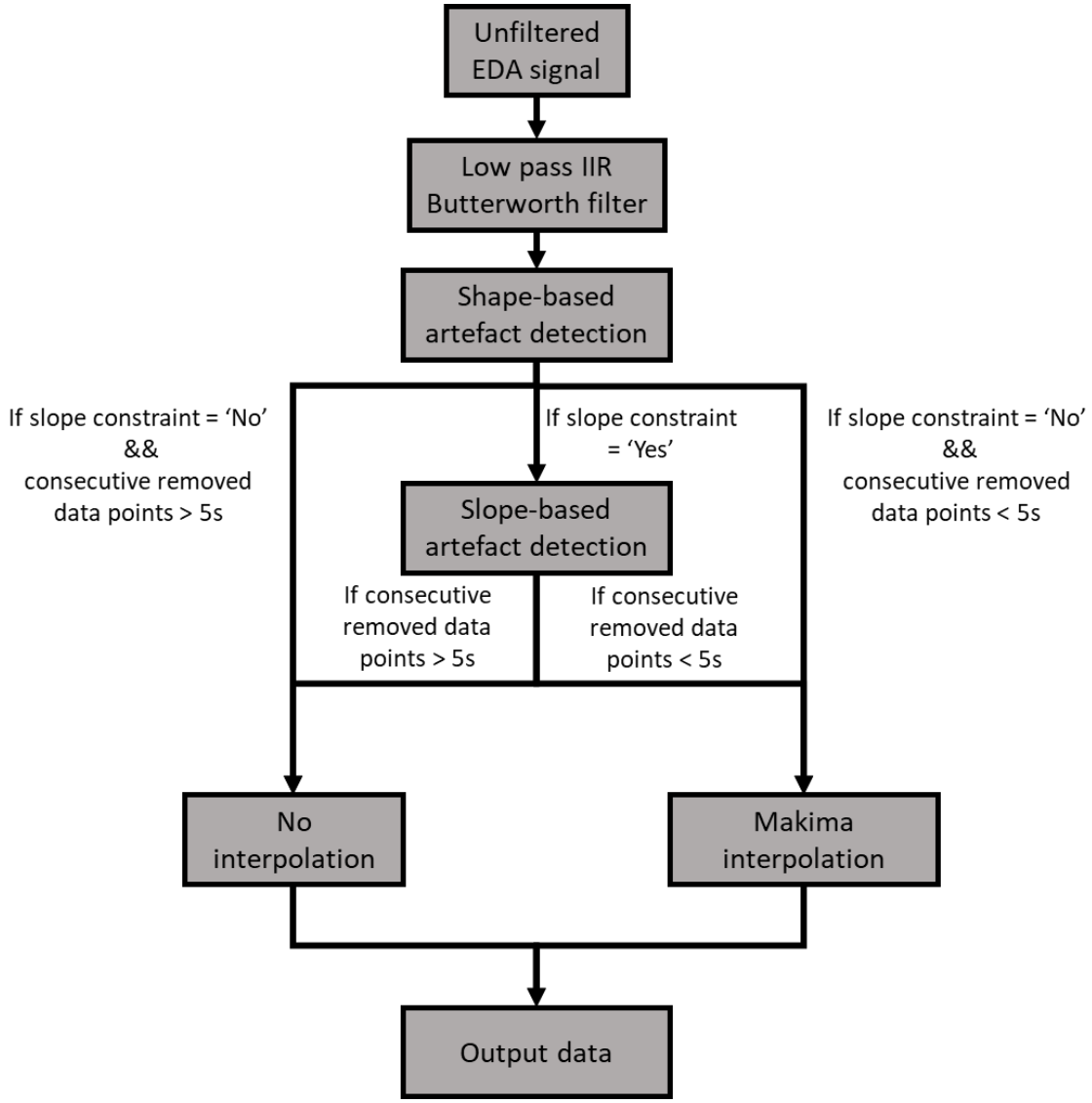

Figure 2.2. Flow chart of the artefact detection algorithm and filling signal gaps after artefact removal.

In the next section, we do a case study of a sample EDA data, and apply the above proposed algorithm to remove motion artefacts and fill in missing data using the proposed interpolation technique.

2.3 Case Study and Discussion

The proposed algorithm in section 2.2 was applied to a sample dataset that included ~ 5 minutes of automated driving a takeover (~ 15 s) and ~ 5 minutes of manual driving. The data was recorded in the University of Leeds Driving Simulator

(UoLDS). UoLDS is a full motion-based driving simulator, consisting of a 4 m diameter spherical projection dome, with a 300º field of view projection system, and a Jaguar S-type cab, housed within the dome. Its electric powered motion system has 8 degrees of freedom and consists of a 500 mm stroke-length hexapod motion platform, carrying the 2.5 t payload of the dome and vehicle cab combination, and allowing movement in all six orthogonal degrees of freedom of the Cartesian inertial frame. The motion platform is mounted on a railed gantry, which further allows 5m of effective motion in surge and sway.

The sample EDA data was collected using the BIOPAC MP35 data acquisition system, with an acquisition rate (frequency) of 512 Hz. The results of the noise removal algorithm mentioned in Section 2.2 are discussed below.

2.3.1 Artefact detection

After low-pass filtering, the shape-based algorithm, as presented in section 2.2 of this paper, have been used to detect motion artefacts, as shown in the **Figure 2.3** below.

The unfiltered data, with noise highlighted in pink bands, can be observed in **Figure 2.3(a)**. The shape-based artefact detection has been validated in Kocielnik et al. (2013) and Kikhia et al. (2016). However, since this data was recorded at high sampling rates, some of the motion artefacts, as indicated by the green arrows and confirmed by visual inspection, are kept in the filtered signal. This could affect the interpolation technique implemented to treat the missing data, as observed later on in this section.

Therefore, the data was treated to additional slope constraint as mentioned in section 2.2.1.1. As observed in **Figure 2.4** below, the residual motion artefacts (indicated using green arrows), have been removed from the signal. Additionally, sharp spikes in the signal are also removed. We believe the addition of a slope constraint, along with the shape-based constraints, improves the accuracy of data, by filtering out noisy spikes due to motion, in an EDA signal.

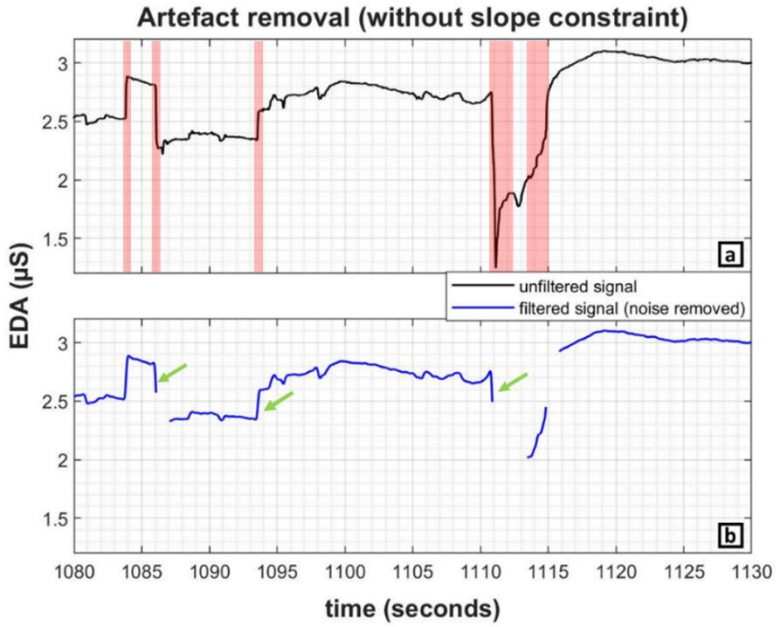

Figure 2.3. **(a)** Unfiltered signal (motion artefacts in pink bands) and **(b)** filtered signal (low-pass IIR filter) with motion artefacts removed, without slope constraint mentioned in section 2.2.1.1. Green arrows indicate points where artefacts were not detected by the algorithm.

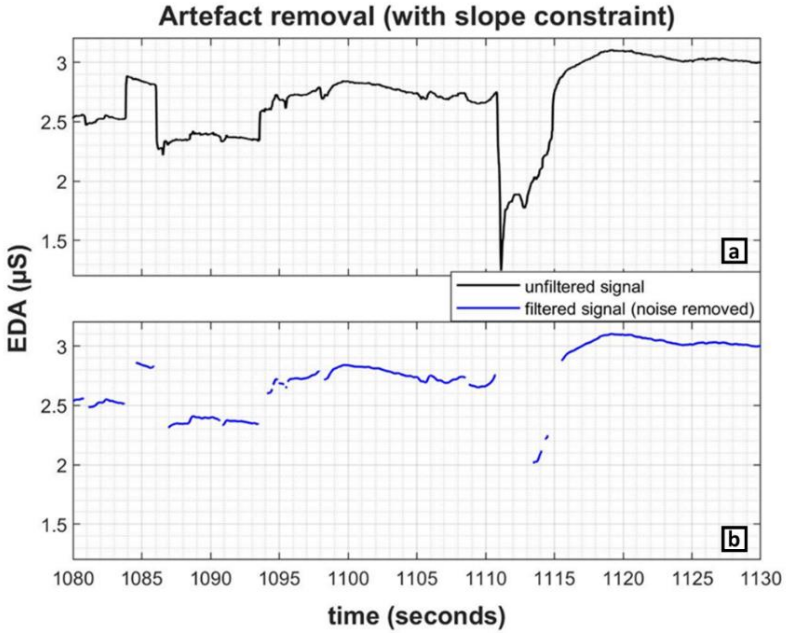

Figure 2.4. (a) Unfiltered signal and **(b)** filtered signal (low-pass IIR filter) with motion artefacts removed, with slope constraint mentioned in section 2.2.1.1.

2.3.2 Interpolating the removed noisy data

As discussed in section 2.2.2, a Makima interpolation (labelled as custom filter in **Figure 2.5**) technique was used to interpolate missing and removed data. By visual inspection, it can be observed that the spline and custom (makima) interpolation techniques, closely preserve the shape of the EDA signal, compared to linear and moving average interpolation techniques. Additionally, the effect of using slope constraint in removing noisy data, on different interpolation techniques is discussed. As seen in **Figure 2.5(b)**, the spline interpolation technique was prone to overshooting, when slope constraint was not implemented. This was also observed in the custom (Makima) filter, although to a much smaller extend. This can be likely attributed to the dependency of spline interpolation techniques on the slope characteristics at the edges of the data being interpolated.

In order to overcome this problem, as well as to discount the noisy data at the end nodes, we implemented the slope constraint discussed in section 2.2.1.1. As observed visually in **Figure 2.6(b)**, the custom (Makima) and spline interpolation techniques performed better with slope constraints and avoided overshoots.

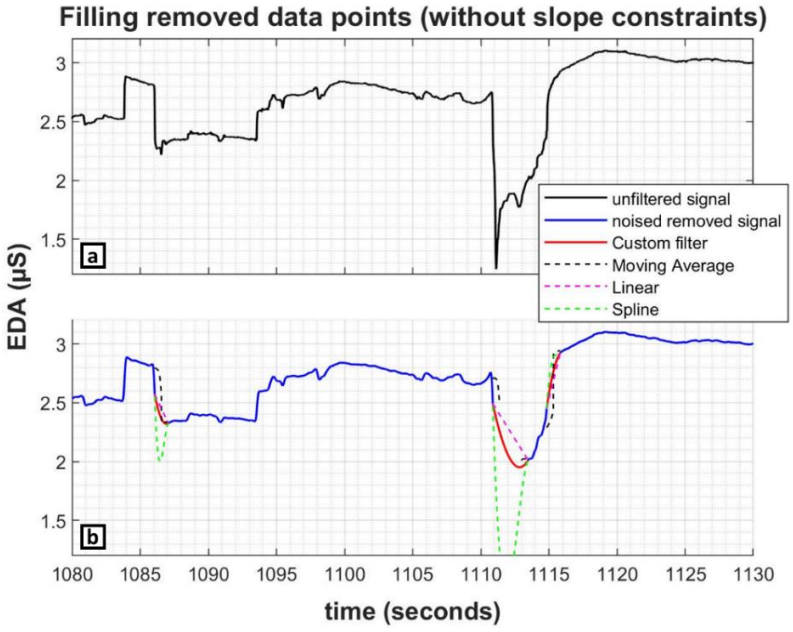

Figure 2.5. (a) Unfiltered signal and **(b)** filtered signal (low-pass IIR filter) with motion artefacts removed (without slope constraints), and data interpolated using custom filter (Makima), moving average, Linear and Spline interpolation techniques.

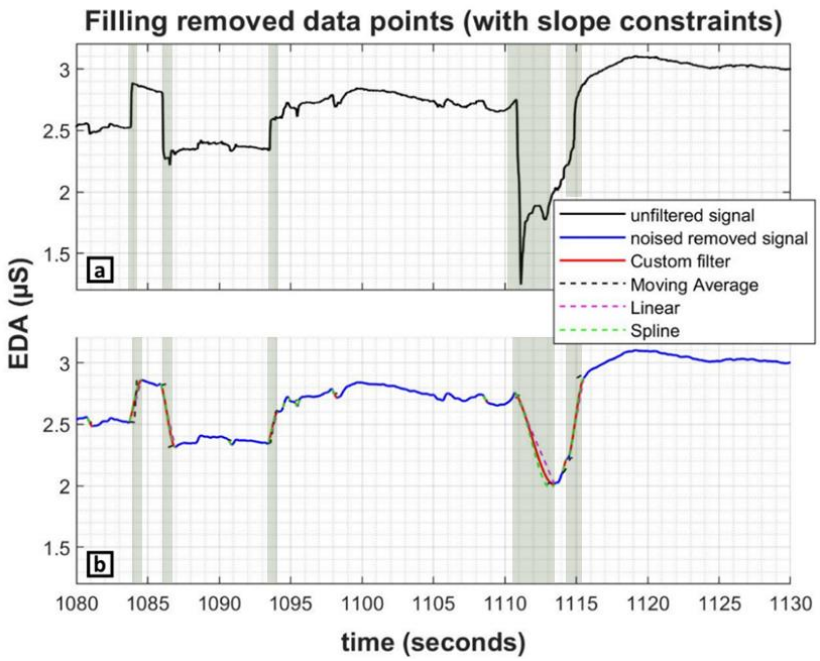

Figure 2.6. **(a)** Unfiltered signal (motion artefacts in light green bands) and **(b)** filtered signal (low-pass IIR filter) with motion artefacts removed (with slope constraints), and data interpolated using custom filter (Makima), moving average, Linear and Spline interpolation techniques.

Overall, our algorithm was robust in detecting and correcting motion artefacts,. In **Figure 2.7**, the noisy parts in the signal can be observed in pink and yellow bands. It can be observed visually that the combination of artefact detection and interpolation techniques mentioned in section 2.2 were able to automatically remove motion artefacts, and interpolate the missing data points without overshoots. The band highlighted in yellow in **Figure 2.7(a)** has been further expanded in **Figure 2.8**.

As observed in **Figure 2.8**, even with slope corrections, normal spline interpolation techniques tend to overshoot. Therefore, the custom (Makima) filter was able to best preserve the shape of the SCR. Additionally, both moving average and linear interpolation techniques tend to distort the shape characteristics of a typical SCR waveform, potentially affecting derived metrics such as SCR amplitudes. Overall, our proposed algorithm was successful in identifying and removing motion artefacts from the EDA signal, and fill missing data using a novel interpolation technique.

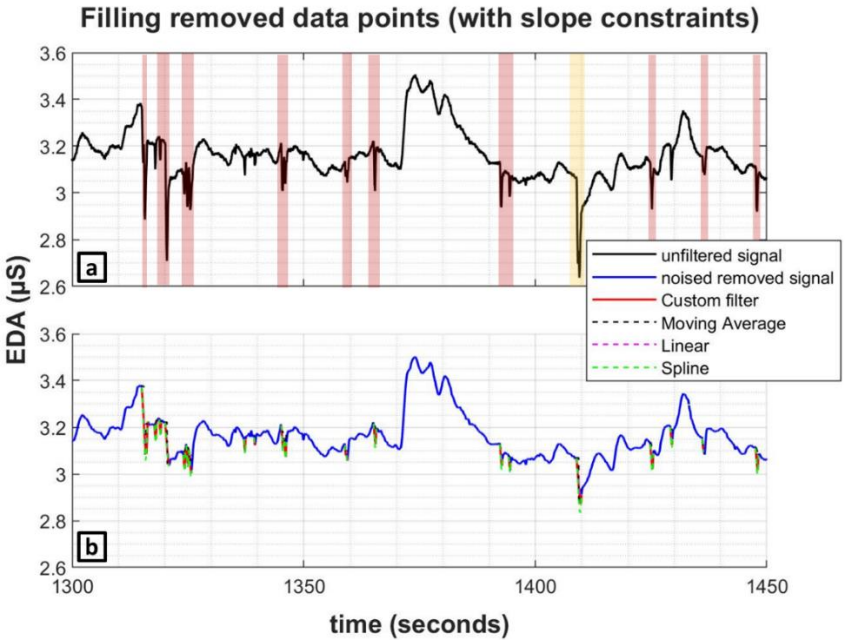

Figure 2.7 (a) Unfiltered signal, with noisy parts highlighted in pink and yellow bands; **(b)** filtered signal (low-pass IIR filter) with motion artefacts removed (with slope constraints), and data interpolated using custom filter (Makima), moving average, Linear and Spline interpolation techniques.

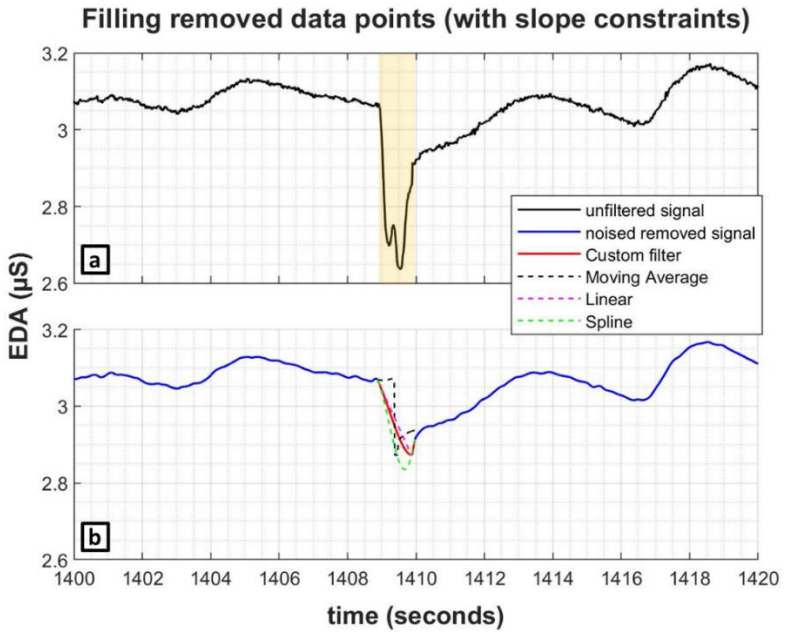

Figure 2.8. (a) Unfiltered signal, with noisy part highlighted in yellow band; **(b)** filtered signal (low-pass IIR filter) with motion artefacts removed (with slope constraints), and data interpolated using custom filter (Makima), moving average, Linear and Spline interpolation techniques

2.3.3 Considerations and Limitations

Generally, an artefact detection algorithm's performance is determined by comparison with artefacts detected manually using expert labellers (Kleckner et al., 2018). However, the shape-based constraint for the EDA signal, described in section 2.2.1, has been well validated in studies such as Kocielnik et al. (2013) and Kikhia et al. (2016). Future studies could further validate our proposed slope constraint, and whether it improves artefact detection accuracy in EDA signals.

For our slope constraint, we suggested using 95th percentile and 5th percentile of the slopes for the entire dataset, as the maximum and minimum slope. However, if the dataset is relatively clean and without motion artefacts, then these values would be well below the threshold values for slope (maximum slope of 0.2 and minimum slope of -0.1). Therefore, depending on the dataset, we would suggest adjusting this value between 95th and 99th percentile for maximum slope, and between 1st and 5th percentile for minimum slope. Additionally, to ensure good data does not get incorrectly identified as motion artefact, we incorporated a buffer for the slope thresholds. That is, if the 95th percentile of the set of slopes were less than 0.4 (two times the derived maximum slope values derived from Boucsein (2012) and Kocielnik et al. (2013) 's recommendation) or the 5th percentile of the set of slopes were greater than -0.2 (two times the derived minimum slope value), then the maximum and minimum slope thresholds were considered as 0.4 and -0.2. This conservative approached helped us in ensuring no good data was wrongly classified as motion artefacts.

In terms of the interpolation techniques used to fill the missing data points, our custom Makima filter strongly resembled the characteristics and shape of an SCR, as described by Boucsein (2012). It also prevented overshoots and undulations that are normally observed when using other spline interpolation techniques. Future studies could further validate this approach. Additionally, the dataset used in this methodology paper was collected using BIOPAC MP35 data acquisition system. Future studies could investigate the performance of the artefact detection algorithm when the data is collected using different hardware.

2.4 Conclusion

To conclude, we were able demonstrate a novel and rapid technique that can be used to automatically remove motion artefacts from an EDA signal, in dynamic driving environments, with potential for real-time application. This can help assist future driver state monitoring systems (DSM), in accurately classifying driver states such as discomfort, high workload or stress (Beggiato et al., 2019; Foy & Chapman, 2018), using EDA-based metrics.

2.5 References

Akima, H. (1970). A New Method of Interpolation and Smooth Curve Fitting Based on Local Procedures. *Journal of the ACM (JACM)*, *17*(4), 589–602. https://doi.org/10.1145/321607.321609

Beggiato, M., Hartwich, F., & Krems, J. (2019). Physiological correlates of discomfort in automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *66*, 445–458. https://doi.org/10.1016/j.trf.2019.09.018

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, *190*(1), 80–91. https://doi.org/10.1016/j.jneumeth.2010.04.028

Boucsein, W. (2012). *Electrodermal activity* (Second). Springer. https://doi.org/https://doi.org/10.1007/978-1-4614-1126-0

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2015). *A Guide for Analysing Electrodermal Activity & Skin Conductance Responses (SCRs) for Psychophysiological Experiments*. https://doi.org/10.1017.S0142716405050034

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (Eds.). (2007). *The Handbook of Psychophysiology* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511546396

Dan, E. L., Dinsoreanu, M., & Muresan, R. C. (2020, May 1). Accuracy of Six Interpolation Methods Applied on Pupil Diameter Data. *22nd IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR)*. https://doi.org/10.1109/AQTR49680.2020.9129915

Dong, Y., & Peng, C. Y. J. (2013). Principled missing data methods for researchers. In *SpringerPlus* (Vol. 2, Issue 1, pp. 1–17). Springer. https://doi.org/10.1186/2193-1801-2-222

Foy, H. J., & Chapman, P. (2018). Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation. *Applied Ergonomics*, *73*, 90–99. https://doi.org/10.1016/j.apergo.2018.06.006

Hernandez, J., Morris, R. R., & Picard, R. W. (2011). Call Center Stress Recognition with Person-Specific Models. In S. D'Mello, A. Graesser, B. Schuller, & J.-C. Martin (Eds.), *Affective Computing and Intelligent Interaction: Vol. 6974 LNCS* (Issue PART 1, pp. 125–134). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-24600-5_16

Kikhia, B., Stavropoulos, T. G., Andreadis, S., Karvonen, N., Kompatsiaris, I., Sävenstedt, S., Pijl, M., & Melander, C. (2016). Utilizing a wristband sensor to measure the stress level for people with dementia. *Sensors (Switzerland)*, *16*(12). https://doi.org/10.3390/s16121989

Kleckner, I. R., Jones, R. M., Wilder-Smith, O., Wormwood, J. B., Akcakaya, M., Quigley, K. S., Lord, C., & Goodwin, M. S. (2018). Simple, Transparent, and Flexible Automated Quality Assessment Procedures for Ambulatory Electrodermal Activity Data. *IEEE Transactions on Biomedical Engineering*, *65*(7), 1460–1467. https://doi.org/10.1109/TBME.2017.2758643

Kocielnik, R., Sidorova, N., Maggi, F. M., Ouwerkerk, M., Westerink, J. H. D. M., Kocielnik Eindhoven, R. T., Sidorova Eindhoven, N. T., Maria Maggi, F., Ouwerkerk, M., & HDM Westerink, J. (2013). Smart technologies for long-term stress monitoring at work. *Proceedings of CBMS 2013 - 26th IEEE International Symposium on Computer-Based Medical Systems*, 53–58. https://doi.org/10.1109/CBMS.2013.6627764

Kübler, T. C., Kasneci, E., Rosenstiel, W., Schiefer, U., Nagel, K., & Papageorgiou, E. (2014). Stress-indicators and exploratory gaze for the analysis of hazard perception in patients with visual field loss. *Transportation Research Part F: Traffic Psychology and Behaviour*, *24*, 231–243. https://doi.org/10.1016/j.trf.2014.04.016

Mathworks. (2019). *MATLAB R2019a*.

Moler, C. B. (2004). Interpolation. In *Numerical Computing with Matlab* (pp. 93–116). Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9780898717952.ch3

Radhakrishnan, V., Merat, N., Louw, T., Lenné, M. G., Romano, R., Paschalidis, E., Hajiseyedjavadi, F., Wei, C., & Boer, E. R. (2020). Measuring drivers' physiological response to different vehicle controllers in highly automated driving (HAD):

Opportunities for establishing real-time values of driver discomfort. *Information (Switzerland)*, *11*(8), 390. https://doi.org/10.3390/INFO11080390

Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G., & Ehlert, U. (2010). Discriminating stress from cognitive load using a wearable eda device. *IEEE Transactions on Information Technology in Biomedicine*, *14*(2), 410–417. https://doi.org/10.1109/TITB.2009.2036164

Taylor, S., Jaques, N., Chen, W., Fedor, S., Sano, A., & Picard, R. (2015). Automatic identification of artifacts in electrodermal activity data. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, *2015-Novem*, 1934–1937. https://doi.org/10.1109/EMBC.2015.7318762

3 DRIVER DISCOMFORT DURING HIGHLY AUTOMATED DRIVING

ABSTRACT: This study investigated how driver discomfort was influenced by different types of automated vehicle (AV) controllers, compared to manual driving, and whether this response changed in different road environments, using heart-rate variability (HRV) and electrodermal activity (EDA). A total of 24 drivers were subjected to manual driving and four AV controllers: two modelled to depict “human-like” driving behaviour, one conventional lane-keeping assist controller, and a replay of their own manual drive. Each drive lasted for ~15 min and consisted of rural and urban environments, which differed in terms of average speed, road geometry and road-based furniture. Drivers showed higher skin conductance response (SCR) and lower HRV during manual driving, compared to the automated drives. There were no significant differences in discomfort between the AV controllers. SCRs and subjective discomfort ratings showed significantly higher discomfort in the faster rural environments, when compared to the urban environments. Our results suggest that SCR values are more sensitive than HRV-based measures to continuously evolving situations that induce discomfort. Further research may be warranted in investigating the value of this metric in assessing real-time driver discomfort levels, which may help improve acceptance of AV controllers.

3.1 Introduction

In the recent past, there has been an increasing interest in implementing vehicles with a range of advanced driver assistant systems (ADAS), fuelled by manufacturers’ desire to introduce higher levels of vehicle automation capability (SAE International, 2018). The primary motivation for these implementations is their hypothesised provision of increased road safety, and enhanced mobility, accessibility, efficiency and comfort (ERTRAC, 2017). According to Carsten & Martens (2019), manufacturers have been using comfort as one of the main selling points for ADAS. Additionally, the comfort of the driver is considered to be a determining factor for the broader acceptance of the automated system (Siebert et al., 2013). Therefore, it can be argued that, if an automated system can measure driver comfort in real-time, it can adapt its driving style/behaviour to match the drivers’ expectations accordingly, and thereby potentially increase acceptance. This could have the additional benefit of

reducing unnecessary driver initiated takeovers, which can otherwise jeopardise the safety of the vehicle and its occupants (Beggiato et al., 2018). This study, conducted as part of the HumanDrive project, considered the effect of a number of road and vehicle-based factors on driver comfort, investigating whether physiological metrics can be used to provide an objective measure of comfort, to help inform the design process when investigating the acceptance of future automated vehicles.

Currently, there is no unanimously agreed on definition of comfort. In a general context, Slater (1986) described comfort as “a pleasant state of physiological, psychological and physical harmony between human being and the environment” (p. 158). In the context of driving, and especially highly automated driving (HAD), Beggiato, Hartwich, & Krems (2019), defined comfort as “a subjective, pleasant state of relaxation resulting from confidence in safe vehicle operation which is achieved by the absence of uneasiness and distress” (p. 446). Beggiato et al. (2019) further suggested this is still a rather broad definition of comfort, and is associated with other concepts, such as stress, mental workload, fear, motion sickness or anger, with stress and mental workload having the closest link to discomfort (i.e., lack of comfort). Siebert et al. (2013) argued that it is easier to measure discomfort rather than comfort, since signs of discomfort tend to be more well-defined and pronounced, compared to the un-aroused relaxed state of comfort. Summala (2007) proposed four factors that need to be maintained above a certain threshold to keep drivers within their “comfort zone” during manual driving. These are safety margins (to road edges, obstacles or other vehicles), vehicle-road system (accelerations, road geometry), rule-following (obeying traffic laws, maintaining speed limits) and good progress of the trip (meeting one’s expectations for the pace or progress of the travel). However, assuming 100% performance of the automated system, Siebert et al. (2013) noted that the rule-following factor for comfort is redundant in HAD, as the automated vehicle (AV) will almost certainly follow the rules, and that good progress of the trip is dependent on traffic conditions, rather than automation state in itself, assuming the route selected by the AV is similar to that in manual driving, where the navigation system decides/recommends the optimal route to be followed. Therefore, in this paper, we focus specifically on how factors that affect the safety margins, and vehicle-road system, affect driver discomfort, for manual and automated driving.

Summala (2007), suggested that sufficient safety margins from potential hazards are required for a driver to feel safe and comfortable. Factors influencing these safety margins, and likely to increase driver discomfort, include situations which increase drivers' stress levels, such as navigating in crowded cities, interactions with other road users, or when passing another car/obstacle (Cahour, 2008; Healey & Picard, 2005).

Comfort is affected by jerk and acceleration forces of the vehicle, with higher accelerations and jerks (in terms of both magnitude and frequency) associated with an increase in discomfort (Beard & Griffin, 2013; Martin & Litwhiler, 2008; Wertheim & Hogema, 1997), and an increase in motion sickness (H. Vogel et al., 1982). Drivers tend to keep their lateral and longitudinal acceleration under 2 m/s^2 for a comfortable driving experience (Bae et al., 2019; Bosetti et al., 2014; Moon & Yi, 2008). However, it should be noted that drivers' comfort threshold for lateral acceleration varies with respect to their velocity, with an increase in velocity resulting in lower threshold values for lateral acceleration (Bosetti et al., 2014; Levison et al., 2007). Within the public transport domain, especially in railway systems, standard acceleration values are limited to under 1.47 m/s^2, and jerk values are kept under 0.6 m/s^3, to ensure passenger comfort (Bae et al., 2019; Martin & Litwhiler, 2008; Powell & Palacín, 2015). However, the acceleration and jerk thresholds used in public transport systems consider both seated and standing passengers. Therefore, it may be permissible to have slightly higher thresholds in HAD, where passengers are typically seated. For instance, Eriksson & Svensson (2015) suggested an acceleration and jerk threshold of under 2 m/s^2 and 0.9 m/s^3 respectively, to ensure a comfortable ride in HAD.

Because AVs are still in the prototype and testing phase, most individuals have not had a real-world experience of HAD. Therefore, our expectations of what constitutes a 'comfortable' experience during HAD can only be based on our current understanding of users' comfort in either manual driving, or in other surface transport modes. However, there are considerable differences between these modes, in terms of Summala's (2007) proposed four factors, described above, making them difficult to compare to HAD. Thus, to assist with the development of more acceptable AVs, and to ensure user uptake of these systems in the future, it is of value to understand what particular features of an AV's manoeuvres are likely to enhance or diminish user

discomfort. For example, humans try to minimise the jerk during manual driving, whereas most current ADAS features tend to have a relatively higher jerk, due to their preference to stay closer to the lane centre and unwillingness to cut corners, unlike human drivers. Thus, it is important to know if users would prefer, and feel more comfortable with, a more “human-like” AV controller, which favours manoeuvres that result in lower acceleration and jerk, over a more conventional AV controller, with very strict margins for optimal and accurate lane-keeping and vehicle velocities.

Studies on comfort in manual driving have used subjective measures, such as comfort questionnaires (Thakurta et al., 1995) and comfort scales (Myers et al., 2008). Since comfort is highly subjective, it can be challenging to measure it accurately and reliably on a moment-to-moment basis. In a real-world HAD scenario, the driver may become annoyed if they are asked to rate their comfort levels time and again during the drive, especially when they have the option to engage in more appealing non-driving related activities. Thus, in HAD, there is a need for a non-intrusive, objective, discomfort detection system, which can ultimately be used to adapt the automated system’s driving style, to ensure the driver is relaxed and at ease (Beggiato et al., 2019). Physiological techniques are one example of such objective methods, which have been used in the past to assess driver state both in HAD (Beggiato et al., 2019) and manual driving (Lal & Craig, 2002; Mehler et al., 2009). Recent technological advancements have led to the development of non-intrusive physiological devices that measure heart rate variability (HRV) and electrodermal activity (EDA), such as wearable smart-band sensors like Empatica E4 (McCarthy et al., 2016a) or Microsoft band 2 (Beggiato et al., 2019), and non-contact methods, such as those listed in (Kranjec et al., 2014). Previously, studies have shown strong correlations between stress and workload, and users’ HRV, and EDA. A general finding is that heart rate (HR) increases, and HRV (including the time-domain based metric of root mean square of successive differences in R-R intervals (RMSSD)) decreases, during periods of high stress or workload (Cinaz et al., 2010; Healey & Picard, 2005; Mehler et al., 2011; Orsila et al., 2008).

An EDA signal consists of the slow-changing tonic component called skin conductance level (SCL) and the rapidly changing phasic component, known as skin conductance response (SCR, Braithwaite et al. 2015). SCRs are generally used to

understand short-term fluctuations in the EDA signal, due to a short-term stimulus (for example, being startled or passing an obstacle), whereas SCL is used to understand the overall change in a person's skin conductance when the stimulus is spread over a longer period (for example, fatigue induced by driving for a long time). SCRs have a much shorter decay time than SCLs, and, hence, can more accurately capture differences in manipulations, without the need for recovery/resting periods in between (Braithwaite et al., 2015; Cacioppo et al., 2007a). In the context of driving, both SCL and SCRs have been shown to increase with an increase in stress and workload for a driver (Foy & Chapman, 2018; Healey & Picard, 2005; Mehler et al., 2009), and, thus, are associated with increases in discomfort (Beggiato et al., 2019). Based on these findings, we analysed RMSSD, HR and SCR responses per minute (nSCR/min) in this study, as the objective physiological metrics of drivers' comfort.

3.1.1 Current study

This study was undertaken as part of a 10-member consortium of the HumanDrive project, part-funded by the UK's Centre for Connected and Autonomous Vehicles (CCAV), via Innovate UK. The main aim of the project was to develop an advanced vehicle controller, which allowed the vehicle to perform a 'natural', human-like, driving style, using artificial intelligence (AI), and deep learning techniques. As outlined above, developing a human-like controller could potentially help with the broader acceptance of AVs, driven by a more natural driving style, which is familiar to the driver. Using manual driving data collected from 44 drivers in an earlier HumanDrive study, an aggregated model for human-like controllers, focusing on both vehicle safety and comfort, was developed for the present study (see also Hajiseyedjavadi et al., in prep), for more details of the controllers). An environment-specific risk model was developed to guide the design of the experiments. The simulated drives were constructed to include risk elements present in the drive, based on road width and curvature, as well as on the presence of road-based furniture and obstacles, such as hedges of different heights, grass/asphalt verges, pedestrian refuges and parked-cars or roadworks (see Louw et al., 2019, for more details). The development of this risk model was based on satisficing risk corridors, proposed by Boer (2016), where a set of vehicle states are within acceptable bounds. The vehicle state includes velocity and lateral offset. The trajectory of the vehicle is always within this risk corridor and adopts a comfortable smoothness for the ride. The model holds

that drivers' perceived risk level is based on minimum time to lane crossing, wherein the lateral position for the vehicle stays within the road boundaries (Boer, 2016). Based on this model, two human-like AV controllers (SLOW and FAST, with the FAST controller having higher velocities than the SLOW controller) were developed, and compared to a conventional controller (LKAS), and drivers' replay of their own drive (see Section 3.2.3, for more details). To understand how the different physical characteristics of a drive can affect drivers' discomfort, our study exposed participants to a range of accelerations, induced by the four different AV controllers and manual driving. Participants experienced these controllers in two different road environments (rural and urban), which included a variety of road geometries, such as roads of different curvatures/width/speed limit, containing a range of road furniture/obstacles (parked cars, roadworks and pedestrian refuges). Previous studies on driver discomfort during HAD, such as Beggiato et al. (2019), have focused on discrete situations causing discomfort, such as negotiating an intersection, exit ramp or an obstacle. In our study, we considered the effects of longer, repeated exposure to different road environment, human-like AV controllers and interactions with road furniture and obstacles, on drivers' discomfort. Drivers' HR and EDA data were compared to drivers' self-reported level of perceived discomfort for each road environment, which was measured in real-time, using a button pressing technique (see Section 3.2.4 for more details). We addressed the following research questions:

i. How is driver discomfort, as measured by changes in physiological state (i.e., HRV and EDA), affected by the various controllers, and manual driving?

ii. Do drivers' discomfort levels change, based on the behaviour of the different controllers, in the different road environments (rural and urban)?

iii. Does the change in drivers' physiological state reflect their self-reported level of perceived discomfort during HAD?

3.2 Materials and Methods

3.2.1 Participants

In total, 24 participants (10 Female), each with a valid UK driving licence, took part in this driving simulator-based study. Their mean age was 43 ± 17 years, with a mean driving experience of 23 ± 18 years. All participants gave consent to take part in the study, in accordance with the rules and regulations of the University of Leeds ethics committee (LTTRAN-086) and were compensated with £50 for taking part in the study. Participants were pre-screened for physiological data collection and those with pre-existing heart conditions were not included in the study, as per recommendations in Braithwaite et al. (2015) and Laborde, Mosley, & Thayer (2017). In addition, participants were requested to avoid consuming food and beverages that had cardiac stimulants such as caffeine or alcohol for 24 h before they took part in the study.

3.2.2 Apparatus

The experiment was conducted in the full motion-based University of Leeds Driving Simulator (UoLDS), which consists of a Jaguar S-type cab housed in a 4 m diameter spherical projection dome with a 300-degree field-of-view projection system. The simulator also incorporates an 8 degree-of-freedom electrical motion system. This consists of a 500 mm stroke-length hexapod motion platform, carrying the 2.5 t payload of the dome and vehicle cab combination, and allowing movement in all six orthogonal degrees-of-freedom of the Cartesian inertial frame. Additionally, the platform is mounted on a railed gantry that allows a further 5 m of effective travel in surge and sway. Drivers' physiological data were collected using a Biopac MP35 data acquisition system at 1000 Hz, which consisted of ECG electrodes and an EDA sensor.

3.2.3 Study Design

The study used a within-participant design and included a short familiarisation drive for ~10 min. Each participant experienced five drives: a MANUAL drive, two with human-like AV controllers (SLOW and FAST), a replay of their manual drive (REPLAY) and one conventional lane-keeping assist-based AV controller (LKAS) which did not adapt its behaviour to road furniture, such as kerbs or hedges. Each drive consisted

of two different road environments (rural and urban). The design of the drives and the road environments are discussed below.

3.2.3.1 Road Design

Each drive was 15.8 km long, and incorporated several situations that demanded greater attention and a shift in lateral position and speed, which could be deemed uncomfortable by the driver based on how it was negotiated, presented across two different road environments (rural and urban, see **Figure 3.1**). The speed limits, geometries, and obstacle locations, for each road are listed in **Table 3.1** and **Figure 3.2**. The road design was similar across all drives except for LKAS, which did not include any obstacles, which were partly within the lane, such as roadworks or parked cars.

Figure 3.1. **(a)** Rural environment with roadworks; **(b)** urban environment

Table 3.1. Road geometry and furniture across different segments (in the order they were experienced).

Segment	Obstacles	Environment	Speed Limit (mph)	Road Width (m)	Radius and Number of Curves			
					100 m	170 m	252 m	750 m
Segment 1	-	rural	60	7.3	-	2	3	-
Segment 2	4	rural	60	5.8	1	4	-	-
Segment 3	4	urban	40	7.3	-	-	-	5
Segment 4	-	rural	60	5.8	1	4	-	-
Segment 5	6	urban	40	7.3	-	-	-	5

Roads in the rural environments were narrower than those in the urban environments, except in the first segment, which was wider than the other two rural segments (see **Table 3.1**). We did this to assess whether a decrease in road-width

increased discomfort within the same road environment. Overall, rural environments were designed to have narrower roads, tighter curves, and higher speed limits (and therefore, higher resultant acceleration), along with the presence of obstacles (parked-cars and roadworks, see **Figure 3.1**). These factors were designed to increase the attentional demand of the driver at varying degrees, which could possibly induce discomfort depending on how they were negotiated by the controllers, or drivers' individual manual driving style. There were more obstacles (parked-cars, roadworks, or pedestrian refuge, see **Figure 3.2**) in the urban environments (10), when compared to the rural environments (4), to investigate whether participants' discomfort increased with the number of obstacles.

3.2.3.2 Experimental Design

The five drives were counterbalanced, with the exception of the MANUAL drive, which was always the first drive for every participant, so that data could be collected for their REPLAY drive, although participants were not explicitly informed about this. As discussed in the Introduction, the SLOW and FAST controllers were modelled, based on data collected during manual driving across similar road segments in a previous HumanDrive study (see Louw et al., 2019). They were designed to mimic human-like driving, based on a risk model, which defined a range of acceptable vehicle states, such as velocity and lateral offset, depending on drivers' perceived risk levels in response to different road furniture and features present in the drive, such as parked-cars or sharp curves. The FAST controller had higher velocities, compared to the SLOW controller, with a maximum difference of 4 m/s, and a minimum difference of 0.15 m/s. The driving data used to create the models (see Hajiseyedjavadi et al., in press) showed that when driving at higher velocities, drivers' time to lane crossing (TLC) decreased, and, in order to maintain their preferred safety boundary, they moved further away from the road edge. Taking this knowledge into account, we increased the lateral offset of the FAST controller from the left edge of the road, at a rate of 5 cm for every 1 m/s increase in relative speed, compared to the SLOW controller. The LKAS controller was a simple lane-keeping assist controller, which had a constant velocity for most parts of the drive (at the speed limit for that section), except for when the vehicle had to negotiate a curve, or when it moved from an urban to rural environment (or vice-versa). The LKAS controller mostly kept to the lane centre (even when on curves). The objective of the design of the different drives with these

controllers was to understand how discomfort was affected by factors such as manual and automated driving, the behaviour of the human-like AV controllers, a conventional lane-keeping controller and the controller based on one's own driving style. The different drives and their properties are shown in **Figure 3.2**, **Table 3.2** and **Table 3.3**, which show that the LKAS controller had the highest resultant acceleration (combined lateral and longitudinal accelerations) in rural environments, whereas the SLOW controller had the lowest resultant acceleration in rural environments. The 95th percentile of resultant acceleration and lateral jerk values across all the drives in rural environments was higher than the suggested comfort threshold value for acceleration and jerk (2 m/s^2 and 0.9 m/s^3, respectively, according to Eriksson & Svensson (2015), whereas it was well below this threshold across all drives in the urban environments. The resultant acceleration values were mainly governed by the lateral accelerations, as the longitudinal accelerations were minimal, and within the suggested comfort threshold for longitudinal acceleration, across both environments, for all controllers.

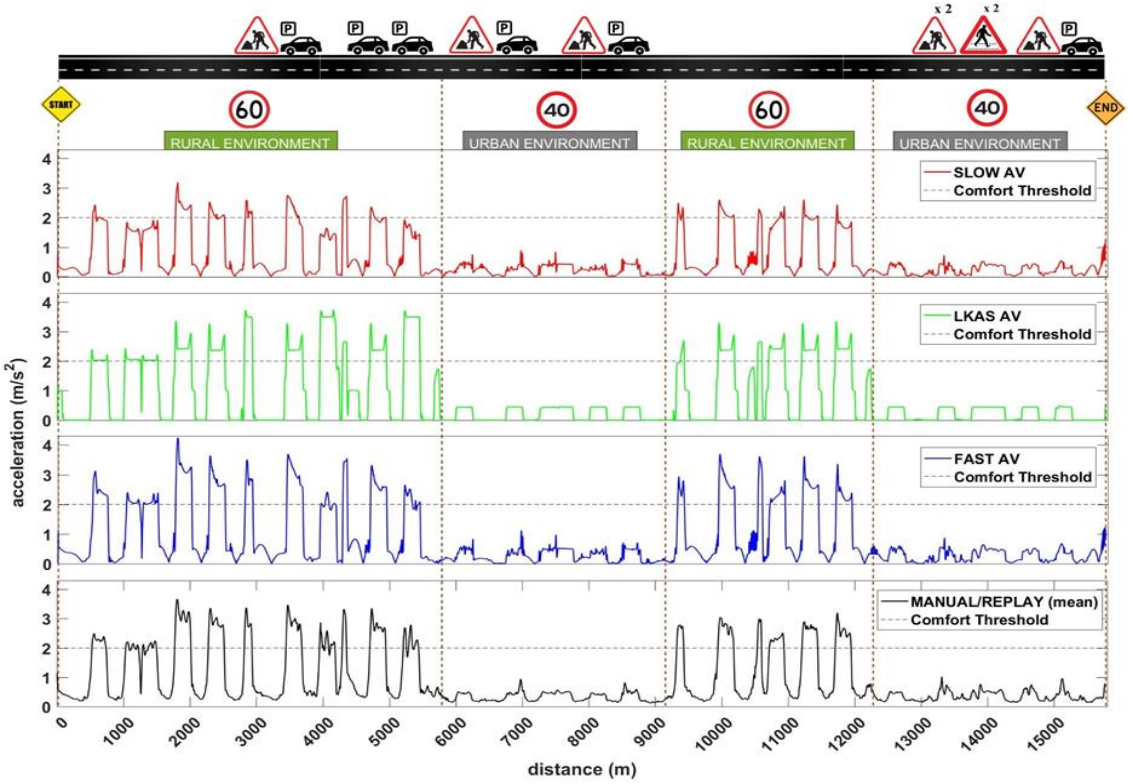

Figure 3.2. Resultant acceleration of the different controllers and manual driving, along with the location of obstacles across all drives, except LKAS.

Table 3.2. The 95th percentile of resultant acceleration (in m/s^2) for different drives across different road environments.

	MANUAL	*SLOW*	*LKAS*	*FAST*	*REPLAY*
Rural	3.42	2.34	3.48	3.20	3.42
Urban	0.74	0.47	0.45	0.57	0.74

Table 3.3. The 95th percentile of absolute values of lateral jerk (m/s^3) for different drives across different road environments.

	MANUAL	*SLOW*	*LKAS*	*FAST*	*REPLAY*
Rural	2.27	1.38	1.71	2.13	2.27
Urban	0.66	0.83	0.19	0.83	0.66

3.2.4 Subjective Discomfort Rating (Button Presses)

For each of the automated drives, the participants heard 41 auditory beep triggers. These beeps were played immediately after the participants were exposed to any obstacles, changes in road furniture, changes in road curvature or changes in road environment. In response to these triggers, they were required to press one of two buttons on an Xbox handset, to state: *"Yes, I found the behaviour to be safe/natural/comfortable"* (right button) or *"No, I did not find the behaviour to be safe/natural/comfortable"* (left button). This response explicitly pertained to the behaviour of the car within a couple of seconds around the moment of the beep's occurrence. Additionally, participants were encouraged to give this binary input whenever they felt necessary, across each drive.

3.2.5 Procedure

Upon arrival, the participants were briefed with the description of the study, after which they were invited to sign a consent form, with an opportunity to ask questions. Three ECG electrodes were then attached to the participant's chest, and 2 EDA electrode bands were attached on the index and middle finger of their non-dominant hand. They then performed a manual familiarisation drive, where they could become accustomed to the simulator environment and vehicle controls. Participants were

instructed to adhere to the posted speed limit and to obey the normal rules of the road. After each drive, the participants were given a 10-min break, during which they were asked to complete a set of subjective questionnaires relating to that drive and the controllers. The results of the subjective questionnaires are not within the scope of this paper and will not be reported here.

3.2.6 Data Analysis Tools

The ECG data was processed on Kubios HRV premium software (Tarvainen et al., 2014). EDA signals were pre-processed, and artefacts were removed using custom algorithms[1] based on recommendations in (Braithwaite et al., 2015) and (Kikhia et al., 2016), on MATLAB R2016a. The data were analysed using Ledalab v3.9 (Benedek & Kaernbach, 2010), a MATLAB-based software package.

3.2.7 Statistical Analysis

Statistical analysis was conducted on IBM SPSS Statistics 26. Shapiro Wilk's test, which showed that not all estimates across the independent variables were normally distributed, but, in general, the majority of the estimates (>75%) were normally distributed for each of the dependent variables used. We judged the repeated measures ANOVA to be sufficiently robust to these issues, with only a small effect on Type I error rate (Blanca et al., 2017). For statistical significance, an α-value of 0.05 was used, and partial eta-squared was computed as an effect size statistic. Degrees of freedom were Greenhouse-Geisser corrected when Mauchly's test showed a violation of sphericity. Pair-wise comparisons with Bonferroni corrections were used to determine the differences in different drives and road segments. Pearson's correlation coefficient was used for any correlation analyses. Data from participants 24 and 14 were classified as outliers, and the data recorded from participants 10 and 15 were of poor quality, and, hence, these were discarded for RMSSD and HR analysis. Participant 12 did not respond to the instructions given for button presses, and participant 13 had an abnormally high rate of button presses. Therefore, these participants were not considered in the subjective button press analysis.

[1] Described in detail in Chapter 2 of this thesis

3.3 Results

Initially, the data were analysed for five separate segments (three in rural and two in urban environments) for each of the five drives, but results for physiological metrics, and the button presses, were not statistically different between the different segments, within the same environment. Therefore, the physiological and button press data across the three rural and 2 urban segments were aggregated for analysis, with the two independent variables being drive (MANUAL, SLOW, LKAS, FAST, REPLAY) and environment (rural and urban). The dependent variables were RMSSD, mean HR and nSCR/min.

3.3.1 Physiological Metrics

To understand how the behaviour of the AV controllers and manual driving affected drivers' physiological response, and discomfort, across the different road environments, we conducted a 5 (Drive: SLOW, LKAS, FAST, MANUAL and REPLAY) × 2 (Environment: rural*,* urban) repeated-measures ANOVA on all three physiological metrics (RMSSD, mean HR, nSCR/min). As discussed in the Introduction, previous research has shown that RMSSD values tend to decrease with an increase in discomfort, whereas mean HR and nSCR/min values tend to increase with an increase in discomfort (Beggiato et al., 2019; Foy & Chapman, 2018).

There was a main effect of drive on RMSSD values, $F(2.4, 45.2) = 5.27$, $p = 0.006$, $\eta_p^2 = 0.22$, **Figure 3.3a**, with post-hoc tests showing significantly lower RMSSD values in the MANUAL drive, compared to the *LKAS* ($p = 0.007$) and FAST ($p = 0.008$) drives. No other significant differences were found between the drives. There was no effect of environment on RMSSD, or any interactions between drive and environment.

There was a main effect of drive on drivers' mean HR, $F(4, 76) = 6.81$, $p < 0.001$, $\eta_p^2 = 0.23$, **Figure 3.3b**, with post-hoc tests showing that drivers had significantly higher mean HR values in the MANUAL drive, compared with the FAST *drive (p = 0.001).* There were no significant differences between the other drives. There was no main effect of environment and no interactions between drive and environment.

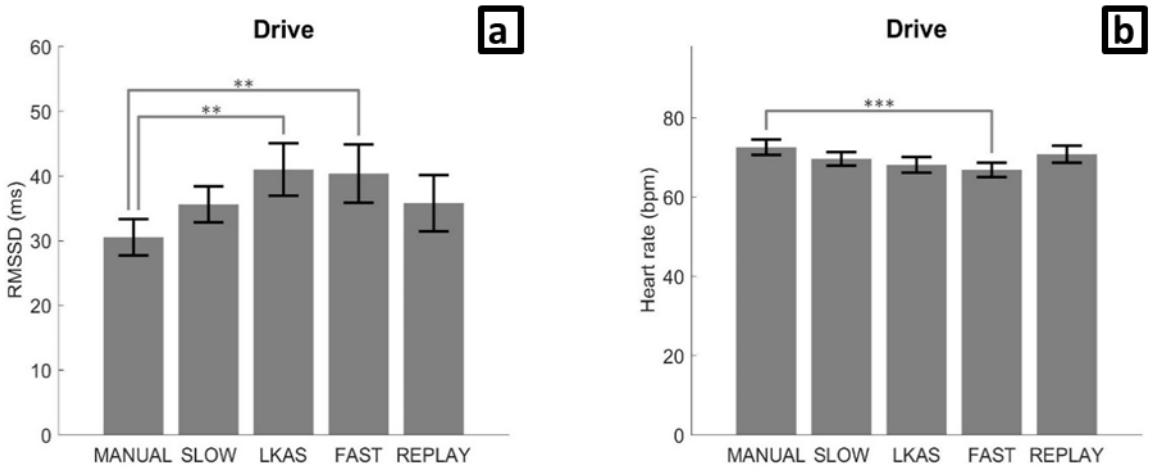

Figure 3.3. **(a)** Root mean square of successive differences (RMSSD) and **(b)** heart rate (HR) plots for drive. ** p ≤ 0.01, *** p ≤ 0.001. Error bars denote s.e.

There was a main effect of drive on nSCR/min, $F(4, 92) = 4.70$, $p = 0.002$, $\eta_p^2 = 0.17$, **Figure 3.4a**, with post-hoc tests showing that there were significantly higher nSCRs/min in the MANUAL drive, compared to the SLOW ($p = 0.006$) and REPLAY drives ($p = 0.005$). There were no other significant differences. There was also a main effect of environment on drivers' nSCR/min, $F(1, 23) = 40.54$, $p < 0.001$, $\eta_p^2 = 0.64$, **Figure 3.4b**, with higher values seen in the rural environments, than the urban environments ($p < 0.001$). An interaction between drive and environment, $F (4, 92) = 3.37$, $p = 0.013$, $\eta_p^2 = 0.13$, **Figure 3.4c**, was also observed. Pairwise comparisons with Bonferroni corrections ($\alpha = 0.002$) revealed that, in the MANUAL drive, drivers had a significantly higher nSCR/min while driving in rural environments, compared to the urban environments ($p < 0.001$). Additionally, within the rural environments, drivers showed significantly higher nSCR/min values in the MANUAL drive, when compared to the SLOW ($p < 0.001$), FAST ($p < 0.001$) and REPLAY ($p = 0.001$) drives. Amongst the AV controllers, LKAS showed the largest reduction in nSCR/min values between rural and urban environments (20.3% reduction in mean nSCR/min from rural to urban).

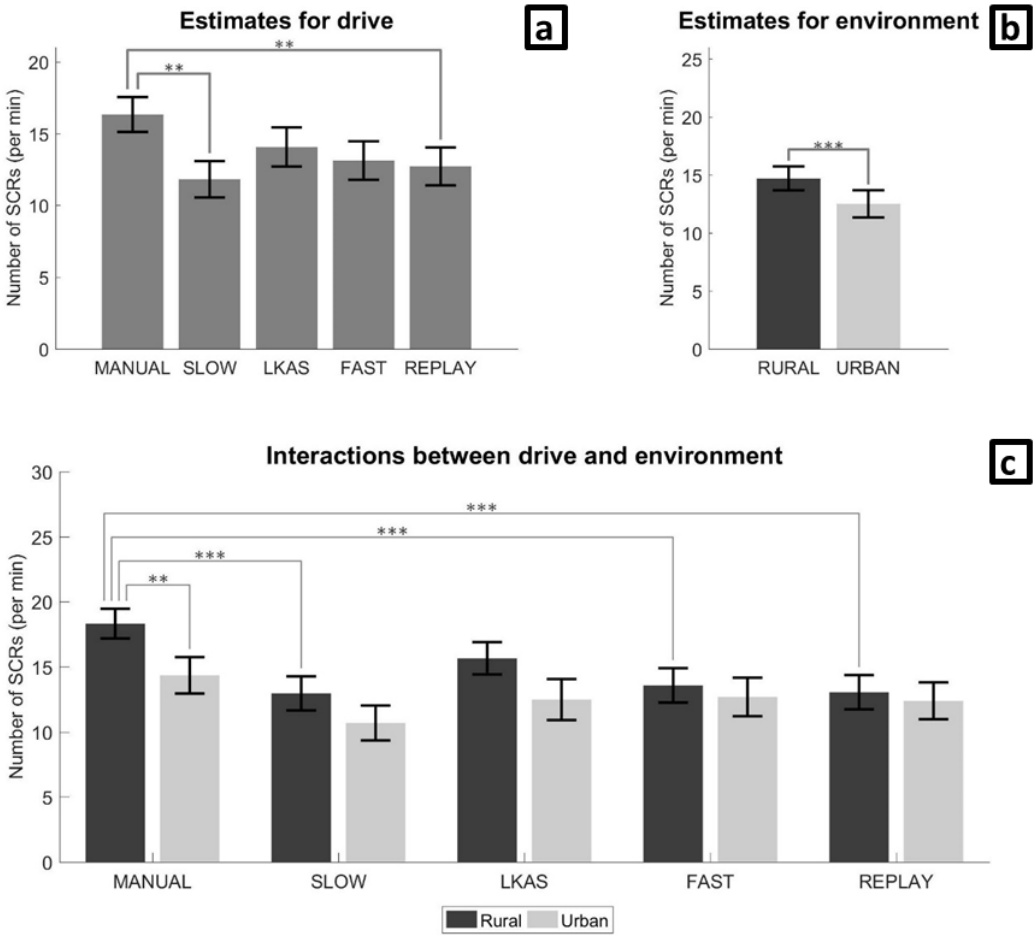

Figure 3.4. Number of skin conductance responses (SCRs) per minute (nSCR/min) for: **(a)** each drive; **(b)** across different environments; **(c)** and interaction effects. **p ≤ 0.01, ***p ≤ 0.001. Error bars denote s.e.

3.3.2 Subjective Discomfort Ratings (Button Presses)

In the previous section, we reported a comparison of drivers' physiological state during each drive. However, physiological signals are sensitive to a wide range of stimuli, and are prone to individual differences. Therefore, care must be taken when interpreting a psychological construct, such as discomfort, using physiological measures only (Beggiato et al., 2019). Hence, we used data from the button presses (see section 3.2.4, in the Methods section) to establish whether the changes in physiological state correlated with the participants' overall subjective discomfort rating. Correlation analysis showed that button presses and nSCR/min were significantly positively correlated ($r(20) = 0.46$, $p = 0.04$).

To normalise the button press data across all participants, the percentage of NO presses was calculated in relation to the total number of presses, for each road environment, in each drive. A 4 × 2 repeated measures ANOVA was performed on the percentage of NO presses to assess discomfort, comparing the values across the four drives (SLOW, LKAS, FAST, and REPLAY) at two different road environments (rural and urban).

ANOVA results showed no main effect of drive on participants' button presses, but there was a main effect of environment, where drivers reported a significantly higher percentage of discomfort ratings in the rural, compared to the urban environment, $F(1, 21) = 9.83$, $p = 0.005$, $\eta_p^2 = 0.32$, **Figure 3.5a**. This pattern is similar to that observed for drivers' nSCR/min values, above.

There was also an interaction effect, $F(3, 63) = 3.16$, $p = 0.031$, $\eta_p^2 = 0.13$, **Figure 3.5b**. Pair-wise comparisons with Bonferroni corrections (α = 0.003125) did not show any significant differences between any of the drives, in each environment. Discomfort ratings were similar across all the drives in the rural environment. However, there was a 43.8% and 52.3% reduction in mean discomfort ratings for LKAS and REPLAY drives, respectively, in the urban environment, compared to their respective values in the rural environment.

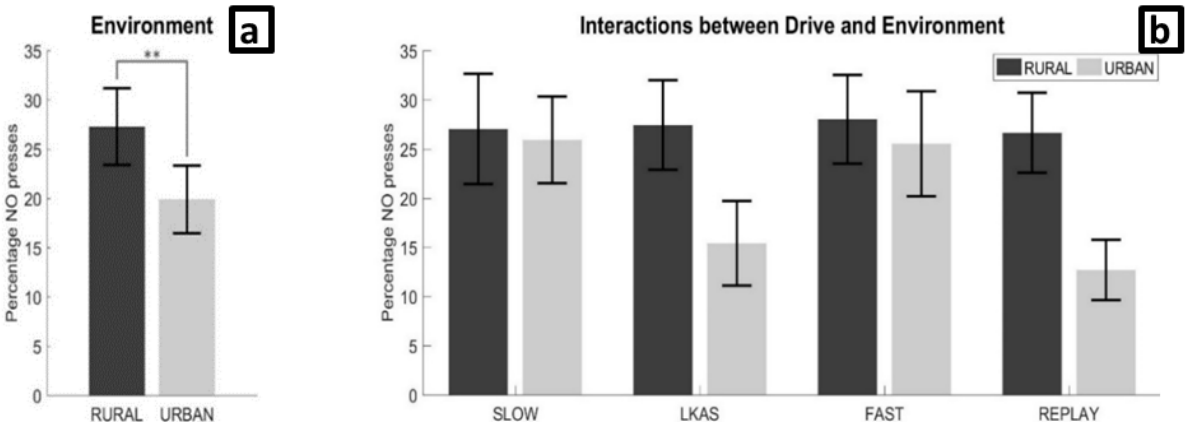

Figure 3.5. Percentage of NO presses: **(a)** across the two environments; **(b)** the interaction between these two factors is shown in the right graph. **p ≤ 0.01. Error bars denote s.e.

3.4 Discussion and Conclusions

This study investigated driver discomfort, from a physiological perspective, and sought to establish whether drivers' physiological state changes in line with the behaviour of different automated vehicle controllers. Drivers' response in manual driving was compared to four automated drives, with each navigating through a range of road geometries and speeds, associated with urban and rural road environments.

Physiological signals can be highly subjective, and therefore individuals may respond slightly differently to a particular stimulus. Additional care must be given whilst interpreting a physiological change to a psychological construct, as a range of constructs could initiate similar physiological responses (Beggiato et al., 2019). In this study, participants were pre-screened for any physiological anomalies that could occur from usage of cardiac stimulants, exercise, or any medication that they were taking. Furthermore, for EDA analysis, we used nSCR/min instead of amplitude sum of each SCRs, and the former is less susceptible to individual differences such as thickness of skin, as each event related SCR is generally initiated as a response to a particular stimuli. This, and, given the fact that our study incorporated a within-subject design, additional standardisation techniques were not applied for processing RMSSD, mean HR and nSCR/min metrics.

Results showed lower RMSSD values, and higher mean HR and nSCR/min values, in the MANUAL drive, compared to at least one of the AV controllers. However, since drivers were not required to evaluate their own driving, by button presses in the MANUAL drive, it is not possible to conclude whether this difference in physiological metric between the MANUAL and automated drives reflects driver discomfort only, or rather, whether it is due to an increased physical and mental demand associated with the manual driving task, or both.

There were no significant main effects in either the physiological metrics, or button press data, between the four automated drives. This may be because overall, the drives had similar resultant acceleration profiles across the whole drive (see **Figure 3.2**). We analysed physiological metrics and subjective button press data for each segment/environment, which were at least 2 min long. Hence, some of the instantaneous variations in controller behaviour may have produced opposing effects, which cancelled each other out when averaged across a larger time window. These

findings are in agreement with Beggiato et al. (2019), where the authors did not find any significant differences in physiological responses between their three automated drives (defensive, aggressive and replay of manual drive). Those authors attributed the lack of difference in physiological responses to high confidence interval bands in their analysis, where missing or opposite effects would have increased the confidence bands dramatically.

In contrast, there were some observable differences, both in terms of physiological metrics nSCR/min), and subjective button presses, for the two road environments, with the rural roads being significantly more uncomfortable than the urban environments. This increase in discomfort is likely attributed to the significantly higher resultant acceleration and jerk experienced in the rural environments, for all drives, which often crossed the 2 m/s^2 and 0.9 m/s^3 threshold for acceleration and jerk, respectively, for a comfortable driving experience, as suggested by Eriksson & Svensson (2015). In other words, the higher speed limits, narrower roads and tighter curves associated with the rural environments, seem to be the main cause of increased driver discomfort in this environment. Although more obstacles were present in the urban sections (10 vs. 4), it seems that the way these were negotiated by the vehicle in the rural sections (i.e., passed at a much higher velocity and on narrower roads), was a significant source of driver discomfort during rural environment. These findings are in line with those of Mourant & Thattacherry (2000), where the authors found higher levels of simulator sickness in high-velocity rural environments, when compared to city environments. These results also suggest that those developing automated vehicle controllers should focus on improving comfort, and thereby minimising jerk, when the vehicle is negotiating higher speed, higher acceleration, road geometries.

While the mean discomfort ratings and nSCR/min seemed to be quite similar across all AV controllers in the rural environments, these were particularly low for the urban section of the LKAS (as seen in both discomfort ratings and nSCR/min) and REPLAY (as seen in the discomfort ratings) drives. This is likely due to the absence of any obstacles in the LKAS drive, resulting in very little variations in velocity and lateral offset (and thus, resultant acceleration). With respect to the REPLAY drive, it is likely that participants visibly recognised their own driving style and preferred this

familiar behaviour during the lower speed urban environment, where their comfort threshold for acceleration forces was not breached. This was also reflected in their subjective ratings. This recognition was indeed noted by some participants, after their REPLAY drive, although not formally recorded. There seems to be incongruence in participants' physiological indicator of discomfort and perceived level of discomfort during the REPLAY drive in urban environments, indicating a bias in rating one's own driving behaviour. These findings suggest that when the resultant acceleration and jerk experienced by the driver remains well below the comfort threshold, other factors that affect discomfort, such as familiarity of the drive or presence of obstacles, become more prominent and noticeable. In contrast, when the resultant acceleration and jerk values moves above the comfort threshold, it seemingly overshadows other determinants of driver discomfort. This warrants further research into understanding drivers' comfort threshold in terms of jerk and acceleration forces, and its impact on other factors that induce discomfort to the driver.

This study was conducted on a dynamic driving simulator (see Section 3.2.2 for more details), and the acceleration and jerk forces experienced by the participants would be similar to that in a real-world scenario. Since acceleration and jerk were two main factors affecting discomfort, we believe a drivers' feeling of discomfort due to these forces is quite similar in a simulator and real-world environment. Johnson et al. (2011) conducted a study on effect of physiological responses in fixed-based simulator vs. real-world driving and concluded that while level of immersion is at an acceptable level to elicit presence and the trends observed in physiological data during simulated driving relative to real-world driving were quite similar, the absolute physiological responses for virtual and real-world environments were significantly different. There is also the possibility of different behavioural responses by drivers in simulator, when compared to a real-world driving situation (Ekanayake et al., 2013). This study incorporated conventional techniques and sensors to measure drivers' physiological data, which were intrusive in nature. However, recent technological advancements have led to non-intrusive (Beggiato et al., 2019; McCarthy et al., 2016a) and even non-contact physiological sensor technologies (Kranjec et al., 2014), which need to be validated with on-road studies.

To conclude, there is a need to measure discomfort objectively, and in real-time, so that future AVs can adapt their driving behaviour and provide a more comfortable and pleasant driving experience for human occupants. The novelty of this study is in understanding and measuring the long-term effects of discomfort, across various road environments and a range of AV controllers, using physiological measures. This study suggests that, compared to HR variability measures, EDA-based SCR values are more sensitive to continuous changes in discomfort inducing stimuli, such as those experienced when a vehicle navigates through different geometric and speed-based scenarios. We observed a moderately positive correlation between participants' nSCR/min and their subjective rating of discomfort. Further research may, therefore, be warranted to investigate the value of this metric for assessing real-time driver discomfort levels, which may be useful when developing more acceptable controllers for future automated vehicles.

3.5 References

Bae, I., Moon, J., & Seo, J. (2019). Toward a comfortable driving experience for a self-driving shuttle bus. *Electronics (Switzerland)*, *8*(9), 943. https://doi.org/10.3390/electronics8090943

Beard, G. F., & Griffin, M. J. (2013). Discomfort caused by low-frequency lateral oscillation, roll oscillation and roll-compensated lateral oscillation. *Ergonomics*, *56*(1), 103–114. https://doi.org/10.1080/00140139.2012.729613

Beggiato, M., Hartwich, F., & Krems, J. (2018). Using Smartbands, Pupillometry and Body Motion to Detect Discomfort in Automated Driving. *Frontiers in Human Neuroscience*, *12*, 338. https://doi.org/10.3389/fnhum.2018.00338

Beggiato, M., Hartwich, F., & Krems, J. (2019). Physiological correlates of discomfort in automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *66*, 445–458. https://doi.org/10.1016/j.trf.2019.09.018

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, *190*(1), 80–91. https://doi.org/10.1016/j.jneumeth.2010.04.028

Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Datos no

normales: ¿es el ANOVA una opción válida? *Psicothema*, *29*(4), 552–557. https://doi.org/10.7334/psicothema2016.383

Boer, E. R. (2016). Satisficing Curve Negotiation: Explaining Drivers' Situated Lateral Position Variability. *IFAC-PapersOnLine*, *49*(19), 183–188. https://doi.org/10.1016/j.ifacol.2016.10.483

Bosetti, P., Da Lio, M., & Saroldi, A. (2014). On the human control of vehicles: An experimental study of acceleration. *European Transport Research Review*, *6*(2), 157–170. https://doi.org/10.1007/s12544-013-0120-2

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2015). *A Guide for Analysing Electrodermal Activity & Skin Conductance Responses (SCRs) for Psychophysiological Experiments*. https://doi.org/10.1017.S0142716405050034

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (Eds.). (2007). *The Handbook of Psychophysiology* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511546396

Cahour, B. (2008). Discomfort, affects and coping strategies in driving activity. *15th European Conference on Cognitive Ergonomics: The Ergonomics of Cool Interaction (ECCE '08)*, *369*, 1–7. https://doi.org/10.1145/1473018.1473046

Carsten, O., & Martens, M. H. (2019). How can humans understand their automated cars? HMI principles, problems and solutions. *Cognition, Technology and Work*, *21*(1), 3–20. https://doi.org/10.1007/s10111-018-0484-0

Cinaz, B. L. ;, Marca, R. ;, Arnrich, B. ;, & Tröster, G. ; (2010). Monitoring of mental workload levels. *IADIS International Conference on E-Health*, 189–193. https://doi.org/10.5167/uzh-66793

Ekanayake, H. B., Backlund, P., Ziemke, T., Ramberg, R., Hewagamage, K. P., & Lebram, M. (2013). Comparing Expert Driving Behavior in Real World and Simulator Contexts. *International Journal of Computer Games Technology*, *2013*. https://doi.org/10.1155/2013/891431

Eriksson, J., & Svensson, L. (2015). *Tuning for Ride Quality in Autonomous Vehicle Application to Linear Quadratic Path Planning Algorithm* [Uppsala University]. http://www.teknat.uu.se/student

ERTRAC. (2017). *Automated Driving Roadmap*. European Road Transport Research Advisory Council. http://www.ertrac.org/uploads/documentsearch/id48/ERTRAC_Automated_Driving_2017.pdf

Foy, H. J., & Chapman, P. (2018). Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation. *Applied Ergonomics*, *73*, 90–99. https://doi.org/10.1016/j.apergo.2018.06.006

Hajiseyedjavadi, F., Merat, N., Romano, R., Paschalidis, E., & Boer, E. (n.d.). *Effect of Environmental and Individual Differences on Subjective Evaluation of Human-Like and Conventional Automated Vehicle Controllers*.

Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems*, *6*(2), 156–166. https://doi.org/10.1109/TITS.2005.848368

Johnson, M. J., Chahal, T., Stinchcombe, A., Mullen, N., Weaver, B., & Bédard, M. (2011). Physiological responses to simulated and on-road driving. *International Journal of Psychophysiology*, *81*(3), 203–208. https://doi.org/10.1016/j.ijpsycho.2011.06.012

Kikhia, B., Stavropoulos, T. G., Andreadis, S., Karvonen, N., Kompatsiaris, I., Sävenstedt, S., Pijl, M., & Melander, C. (2016). Utilizing a wristband sensor to measure the stress level for people with dementia. *Sensors (Switzerland)*, *16*(12). https://doi.org/10.3390/s16121989

Kranjec, J., Beguš, S., Geršak, G., & Drnovšek, J. (2014). Non-contact heart rate and heart rate variability measurements: A review. In *Biomedical Signal Processing and Control* (Vol. 13, Issue 1, pp. 102–112). Elsevier. https://doi.org/10.1016/j.bspc.2014.03.004

Laborde, S., Mosley, E., & Thayer, J. F. (2017). Heart rate variability and cardiac vagal tone in psychophysiological research - Recommendations for experiment planning, data analysis, and data reporting. In *Frontiers in Psychology* (Vol. 8, p. 213). Frontiers. https://doi.org/10.3389/fpsyg.2017.00213

Lal, S. K. L., & Craig, A. (2002). Driver fatigue: Electroencephalography and psychological assessment. *Psychophysiology*, *39*(3), 313–321.

https://doi.org/10.1017/S0048577201393095

Levison, W. H., Campbell, J. L., Kludt, K., Bittner, Alvah C. Jr., I. P., Harwood, D. W., Hutton, J., Gilmore, D., J. Gavin Howe, Chrstos, J. P., Allen, R. W., Kantowitz, B., Robbins, T., & Schreiner, C. (2007). Development of a Driver Vehicle Module for the Interactive Highway Safety Design Model. In *The Federal Highway Administration(FHWA)* (Issue November). https://www.fhwa.dot.gov/publications/research/safety/08019/08019.pdf

Louw, T., Hajiseyedjavadi, F. H., Jamson, H., Romano, R., Boer, E., & Merat, N. (2019). The Relationship between Sensation Seeking and Speed Choice in Road Environments with Different Levels of Risk. *Tenth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 29–35. https://doi.org/10.17077/drivingassessment.1671

Martin, D., & Litwhiler, D. (2008). An investigation of acceleration and jerk profiles of public transportation vehicles. *ASEE Annual Conference and Exposition, Conference Proceedings*. https://www.semanticscholar.org/paper/An-Investigation-Of-Acceleration-And-Jerk-Profiles-Martin-Litwhiler/c2a4d51b2c086eb0e14a71cf1dfa2e19d81a558b

McCarthy, C., Pradhan, N., Redpath, C., & Adler, A. (2016, July 8). Validation of the Empatica E4 wristband. *2016 IEEE EMBS International Student Conference (ISC)*. https://doi.org/10.1109/EMBSISC.2016.7508621

Mehler, B., Reimer, B., Coughlin, J., & Dusek, J. (2009). Impact of Incremental Increases in Cognitive Workload on Physiological Arousal and Performance in Young Adult Drivers. *Transportation Research Record: Journal of the Transportation Research Board*, *2138*, 6–12. https://doi.org/10.3141/2138-02

Mehler, B., Reimer, B., & Wang, Y. (2011). A comparison of heart rate and heart rate variability indices in distinguishing single-task driving and driving under secondary cognitive workload. *Proceedings of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 590–597. https://doi.org/10.17077/drivingassessment.1451

Moon, S., & Yi, K. (2008). Human driving data-based design of a vehicle adaptive cruise control algorithm. *Vehicle System Dynamics*, *46*(8), 661–690.

https://doi.org/10.1080/00423110701576130

Mourant, R. R., & Thattacherry, T. R. (2000). Simulator sickness in a virtual environments driving simulator. *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Association, "Ergonomics for the New Millennium,"* 534–537. https://doi.org/10.1177/154193120004400513

Myers, A. M., Paradis, J. A., & Blanchard, R. A. (2008). Conceptualizing and Measuring Confidence in Older Drivers: Development of the Day and Night Driving Comfort Scales. *Archives of Physical Medicine and Rehabilitation*, *89*(4), 630–640. https://doi.org/10.1016/j.apmr.2007.09.037

Orsila, R., Virtanen, M., Luukkaala, T., Tarvainen, M., Karjalainen, P., Viik, J., & Savinainen, M. (2008). Perceived mental stress and reactions in heart rate variability—a pilot study among employees of an electronics company. *International Journal of Occupational Safety and Ergonomics*, *14*(3), 275–283. https://doi.org/10.1080/10803548.2008.11076767

Powell, J. P., & Palacín, R. (2015). Passenger Stability Within Moving Railway Vehicles: Limits on Maximum Longitudinal Acceleration. *Urban Rail Transit*, *1*(2), 95–103. https://doi.org/10.1007/s40864-015-0012-y

SAE International. (2018). Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. In *SAE International* (Vol. J3016, Issue 201806). https://doi.org/https://doi.org/10.4271/J3016_201806

Siebert, F. W., Oehl, M., Höger, R., & Pfister, H. R. (2013). Discomfort in Automated Driving - The Disco-Scale. *Communications in Computer and Information Science*, *374*(PART II), 337–341. https://doi.org/10.1007/978-3-642-39476-8_69

Slater, K. (1986). The assessment of comfort. *Journal of the Textile Institute*, *77*(3), 157–171. https://doi.org/10.1080/00405008608658406

Summala, H. (2007). Modelling driver behaviour in automotive environments. In P. C. Cacciabue (Ed.), *Modelling Driver Behaviour in Automotive Environments: Critical Issues in Driver Interactions with Intelligent Transport Systems* (pp. 189–207). Springer London. https://doi.org/https://doi.org/10.1007/978-1-84628-618-6_11

Tarvainen, M. P., Niskanen, J. P., Lipponen, J. A., Ranta-aho, P. O., & Karjalainen, P. A. (2014). Kubios HRV - Heart rate variability analysis software. *Computer Methods and Programs in Biomedicine*, *113*(1), 210–220. https://doi.org/10.1016/j.cmpb.2013.07.024

Thakurta, K., Koester, D., Bush, N., & Bachle, S. (1995, February 1). Evaluating short and long term seating comfort. *SAE Technical Papers*. https://doi.org/10.4271/950144

Vogel, H., Kohlhaas, R., & von Baumgarten, R. J. (1982). Dependence of motion sickness in automobiles on the direction of linear acceleration. *European Journal of Applied Physiology and Occupational Physiology*, *48*(3), 399–405. https://doi.org/10.1007/BF00430230

Wertheim, A. H., & Hogema, J. H. (1997). *Thresholds, comfort and maximum acceptability of horizontal accelerations associated with car driving*.

4 MEASURING DRIVER WORKLOAD AT DIFFERENT STAGES OF AUTOMATED DRIVING, INCLUDING TAKEOVERS

ABSTRACT: This driving simulator study, conducted as a part of Horizon2020-funded L3Pilot project, investigated how different car-following situations affected driver workload, within the context of vehicle automation. Electrocardiogram (ECG) and electrodermal activity (EDA)-based physiological metrics were used as objective indicators of workload, along with self-reported workload ratings. A total of 32 drivers were divided into two equal groups, based on whether they engaged in a non-driving related task (NDRT) during automation (SAE Level 3) or monitored the drive (SAE Level 2). Drivers in both groups were exposed to two counterbalanced experimental drives, lasting ~18 minutes each, of Short (0.5 s) and Long (1.5 s) Time Headway conditions during automated car-following (ACF), which was followed by a takeover that happened with or without a lead vehicle. Results showed that driver workload due to the NDRT was significantly higher than both monitoring the drive during ACF and manual car-following (MCF). Furthermore, the results indicated that a lead vehicle maintain a shorter THW can significantly increase driver workload during takeover scenarios, potentially affecting driver safety. This warrants further research into understanding safe time headway thresholds to be maintained by automated vehicles, without placing additional cognitive or attentional demands on the driver. Our results indicated that ECG and EDA signals are sensitive to variations in workload, which warrants further investigation on the value of combining these two signals to assess driver workload in real-time, to help future driver monitoring systems respond appropriately to the limitations of the driver, and predict their performance in the driving task, if and when they have to resume manual control of the vehicle after a period of automated driving.

4.1 Introduction

In recent years, we have seen a gradual increase in the implementation of Advanced Driving Assistance Systems, such as lane-keeping assistance or adaptive cruise control, in vehicles, with manufacturers striving to attain higher levels of vehicle automation capabilities. Highly Automated Driving (HAD) systems such as Traffic Jam Pilot, are currently available on the market (IEEE, 2019) and the use of the first SAE Level 3 (L3; SAE, 2021), Automated Lane Keeping System became legal in the UK in 2021 (Kinnear et al., 2021). Unlike SAE Level 2 (L2) systems, SAE Level 3 does not require constant monitoring of the drive, and drivers can engage in other non-driving

related tasks (NDRTs), and activities. The main rationale for the implementation of HAD is its hypothesised provision of increased comfort and safety (ERTRAC, 2017). However, in terms of the human driver, one of the unwanted consequences of vehicle automation is the out of the loop (OOTL) phenomenon, which, according to Merat et al. (2018), refers to a state where, when automation is engaged, the driver is not monitoring the driving environment, and may or may not be in physical control of the vehicle. Studies have shown that when drivers are out of the loop, a decrement in performance is observed (Endsley & Kiris, 1995) once they are required to resume control of the vehicle, which can lead to reduced safety, for example, if the driver is required to avoid an impending collision (Louw, Madigan, et al., 2017; Wandtner et al., 2018; Zeeb et al., 2016)

Driving performance and safety, after resumption of manual control from HAD, is also affected by driver workload, which describes the relationship between physical and cognitive resources demanded by a task, and those resources available to be supplied by the driver (Dogan et al., 2019; Parasuraman et al., 2008). The relationship between workload and task performance is complex, and follows an 'inverted U-shape' relationship (Bruggen, 2015; de Waard, 1996). For HAD, both high workload (overload - where the task demand exceeds the available resources) and low workload (underload - where the attentional capacity of the driver is reduced due to low task demand and monotony during automation) can result in a performance decrement, and increases the chances of driver error (Parasuraman et al., 2008; Young & Stanton, 2002b). Parasuraman et al. (2008) have also suggested that workload is a better predictor of drivers' future performance, than their current performance. For example, research has shown that a driver's ability to safely resume control from automation is likely to be affected if they are engaged in a high workload task during HAD, such as a demanding NDRT, with worse performance observed compared to a no task period during HAD (Gold et al., 2015; Zeeb et al., 2016). This effect of workload on performance can be especially problematic if there is a sudden increase or change in task demand (such as an unexpected takeover scenario due, e.g. to avoid an obstacle in the lane), compromising safety. For optimal performance in the driving task, reducing the likelihood of errors, drivers are required to maintain a moderate level of workload (Bruggen, 2015; de Waard, 1996). As engagement in NDRTs is likely to increase with higher levels of automation (Carsten et al., 2012; NTSB, 2017),

especially in L3, it is important to understand how driver workload changes during different stages of HAD, in order to provide appropriate mitigation strategies, and reduce performance decrements during transitions of control (Merat et al., 2012; Meteier et al., 2021).

An example of obstacle avoidance after resuming control from automation is preventing a rear-end collision during car following scenarios. Rear-end collisions account for over 31% of all collisions in the US (National Highway Traffic Safety Administration, 2009, p. 56). Car-following refers to the longitudinal following of a lead vehicle by drivers. Given that car-following is a pre-cursor to rear-end collisions (Li et al., 2017), it is of value to understand how different car-following situations in HAD can affect driver workload. Time headway (THW), and time-to-collision, are two safety indicators used in car-following situations to understand drivers' longitudinal driving behaviour (K. Vogel, 2003). THW is defined as the elapsed time between the front of the lead vehicle passing a point on the roadway and the front of the following vehicle passing the same point (Evans, 1991, p. 313).

Research has shown that driver workload can be influenced by the THW maintained by a vehicle. For example, in a manual driving study, conducted in a driving simulator, Liu, Green, & Liu (2019) found that subjective ratings for workload were significantly higher for the 0.5 and 1 s THW conditions, compared to the 2, 2.5 and 3 s THW conditions. In their study on HAD, Siebert & Wallis (2019) reported that drivers were significantly more uncomfortable during THWs under 1.5 s, as reported by their subjective ratings of riskiness due to the THW, compared to longer THWs, while driving at 50 km/h, 100 km/h and 150 km/h respectively. However, participants' subjective ratings of riskiness due to the THW maintained by the vehicle was also dependent on environmental factors such as visibility (fog), and traffic conditions such as following a truck, with driver discomfort increasing with lower visibility, increased traffic, and when following a truck, as opposed to a car. In Louw et al. (2020), we observed, via subjective ratings, that shorter headways maintained by HAD are perceived as riskier or unsafe by drivers, especially when they are not in control of the driving task. Resuming control from automation in the presence of a closer lead vehicle is also likely to be more demanding, especially following engagement in an NDRT (Mehler et al., 2009), further exacerbating the OOTL effect, with studies on HAD

showing that engagement in NDRTs increases driver workload, and negatively affects their driving performance after takeovers (Du et al., 2020; Wandtner et al., 2018; Zeeb et al., 2016).

Given the high inter-individual variability in how people are affected by, or perceive, workload, it can be challenging to accurately measure and interpret it on a moment-to-moment basis. However, the ability to objectively measure workload in real-time during different stages of HAD is crucial, as it can provide insights into drivers' capabilities and limitations, when they are required to resume control of the vehicle, ultimately helping to improve the safety of the automated system. Real-time, minimally-intrusive, and continuous assessment of driver workload can be used to assist the driver, for example, to warn them of dangerous overload or underload situations (Merat et al., 2012). Therefore, in the current paper, we investigated the added value of using physiological signals to objectively measure driver workload, in HAD.

Electrocardiogram (ECG)-based physiological metrics such as heart rate (HR), heart rate variability (HRV), ECG-derived respiration rate (EDR), and metrics derived from electrodermal activity (EDA) signals, have been used extensively to understand and measure workload, in both manual driving and HAD (Biondi et al., 2018; Du et al., 2020; Hidalgo-Muñoz et al., 2019; Mehler et al., 2009).

A general finding is that an increase in drivers' workload is associated with an increase in HR and EDR, the latter of which is the number of breaths a person takes in a minute, as derived from an ECG signal (Hidalgo-Muñoz et al., 2019; Mehler et al., 2009). However, an increase in workload results in a decrease in HRV, which is the physiological phenomenon of variation in time interval between heartbeats (Mehler et al., 2009). A decrease in HRV is reflected by a reduction in drivers' root mean square of successive differences in R-R intervals (RMSSD) (Orsila et al., 2008). An EDA signal consists of a slowly evolving tonic component, called skin conductance level (SCL) and a rapidly evolving phasic component, called skin conductance response (SCR) (Braithwaite et al., 2015; Cacioppo et al., 2007a). SCRs generally have a faster decay time, compared to other EDA and ECG-based metrics, thus making them more sensitive to changes in stimuli that are constantly evolving and/or of short duration (Braithwaite et al., 2015). Both SCL and SCRs are shown to increase with stress and

workload during driving (Du et al., 2020; Foy & Chapman, 2018; Mehler et al., 2009), and the number of SCRs per minute (nSCR/min) has been shown to increase in situations that involve high stress/workload or discomfort for drivers (Foy & Chapman, 2018; Radhakrishnan et al., 2020).

4.1.1 Current study

This study investigated how manipulations of workload, during different stages of L2 and L3 HAD affects drivers' psychophysiological metrics. Workload was manipulated by introducing two lead vehicle conditions (*Lead*/*No Lead*) in an urban driving environment, and two THW conditions when a lead vehicle was present (*Short* and *Long*). To understand how the different vehicle automation states affected driver workload in the context of car-following, our study exposed drivers to automated car-following (*ACF*) segments, manual car-following (*MCF*) segments, and *Takeover* segments, with the latter involving transitions of control from automated to manual driving. To study the effect of NDRT on drivers' workload levels, a between-participant design was used, with one group of drivers (*L2*) asked to monitor the driving environment at all times, and another (*L3*) asked to engage in an NDRT when automation was engaged. ECG- and EDA-based physiological data were collected as objective measures of workload, and compared with drivers' self-reported workload ratings. The following research questions were addressed:

RQ 1. How is L2 drivers' workload, as measured by changes in their physiological state, and self-reported workload ratings, affected by the two THW conditions (*Short* vs *Long*) when monitoring the drive during *ACF*?

RQ 2. Is drivers' workload during *ACF* affected by an NDRT (*L2* vs *L3*)?

RQ 3. Does drivers' workload vary between *ACF* and *MCF*?

RQ 4. Is drivers' workload during the takeover affected by the THWs maintained by the automated controller (*Short* vs *Long*)?

RQ 5. Is drivers' workload during the takeover affected by engagement in an NDRT during automation (*L2* vs *L3*)?

4.2 Materials and Methods

4.2.1 Participants

A total of 32 participants (16 for each level of automation), each with a valid UK driving licence, took part in this driving simulator-based study. A total of 6 participants (3 each from *L2* and *L3*) were excluded from the analysis. Of this, 3 participants did not adhere to the instruction to follow the lead vehicle, and 3 others were excluded due to missing physiological data. For the 13 (4 female, 9 male) participants considered in the *L2* group, the mean age of the participants was 42 ± 17 years, with a mean driving experience of 22 ± 16 years. The 13 participants (3 female, 11 male) of the *L3* group had a mean age of 33 ± 8 years, with a mean driving experience of 14 ± 8 years. Prior to the experiment, participants were instructed to avoid caffeinated products, consumption of alcohol, and engagement in extreme exercise, to control for their effect on physiological data, as recommended in Laborde, Mosley, & Thayer (2017). All participants gave consent to take part in the study, in accordance with the rules and regulations of the University of Leeds ethics committee (LTTRAN-054), and were compensated with £25 for taking part in the study.

4.2.2 Apparatus

The experiment was conducted in the full motion-based University of Leeds Driving Simulator (UoLDS), which consists of a Jaguar S-type cab housed in a 4 m diameter spherical projection dome with a 300º field of view projection system. The simulator also incorporates 8 degrees of freedom electrical motion system. This consists of a 500 mm stroke-length hexapod motion platform, carrying the 2.5 t payload of the dome and vehicle cab combination, and allowing movement in all six orthogonal degrees of freedom of the Cartesian inertial frame. Additionally, the platform is mounted on a railed gantry that allows a further 5 m of effective travel in surge and sway. Drivers' physiological data were collected using Biopac MP35 data acquisition system at 500 Hz, which consisted of ECG electrodes and an EDA sensor.

The Automated Driving System (ADS) was designed to control lateral and longitudinal operation of the vehicle at a speed of 40 mph. The HMI interface on the dashboard showed a red steering wheel symbol when the ADS was inactive (**Figure 4.1a**), and a green steering wheel symbol when the ADS was active (**Figure 4.1b**).

Figure 4.1. HMI Interface on the dashboard: **(a)** when automation was disengaged; **(b)** Automation was engaged (in colour).

4.2.3 Study design

This study incorporated a mixed design, with within-participant factors of Time Headway (*Short*, *Long*), Drive Mode (*ACF*, *MCF*) and Lead Vehicle (*Lead*, *No Lead*), and a between-participant factor of Level of Automation (*L2*, *L3*). All factors, with the exception of Drive Mode, were counterbalanced.

Following a ~10-minute practice drive, each participant experienced two experimental drives. All drives were completed in a single-carriageway urban environment, with a speed limit of 40 mph, and low-density oncoming traffic. For all drivers, their first experimental drive consisted of free driving for ~2 minutes. After around 2 minutes, a lead vehicle joined the driving lane and drivers were instructed to follow the lead vehicle for about 5 minutes, in what we termed as the manual baseline drive. Drivers' baseline car-following behaviour, including their preferred headway, was collected during this segment. Except for the manual baseline segment, which was only present in the first experimental drive, the two experimental drives were the same. There were 4 segments in each experimental drive, which were experienced in the following order: Automated drive 1, Manual drive 1, Automated drive 2, Manual drive 2 (**Figure 4.2**).

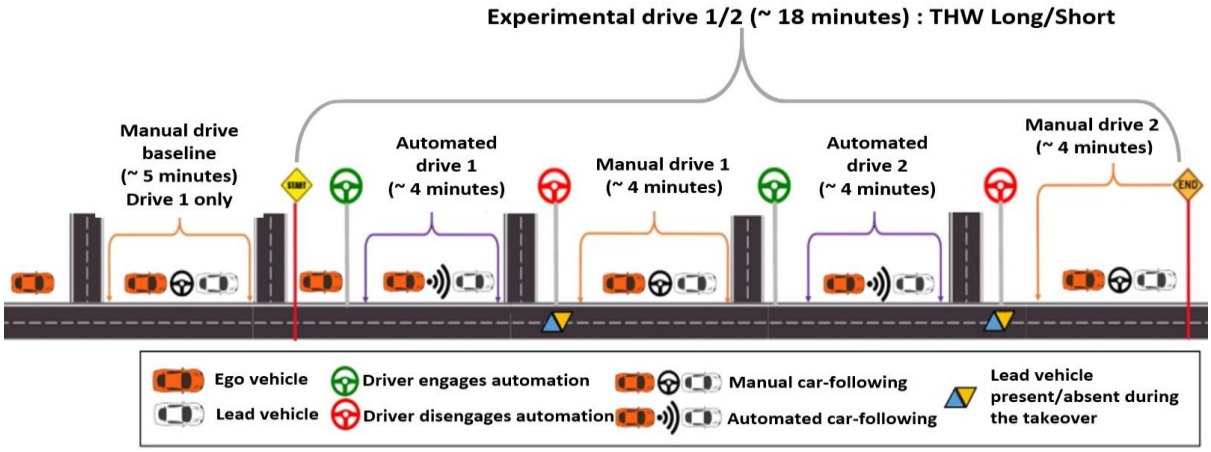

Figure 4.2. Schematic representation of the experimental drives (in colour).

During automation, the participants experienced one of the two THW conditions (0.5 s for the *Short* THW and 1.5 s for the *Long* THW, derived from the 25th and 75th percentile of a driver behaviour model by Ferson et al. (2019), based on naturalistic driving studies incorporating drivers' instantaneous aggressiveness during car-following scenarios), in a counterbalanced order. After the manual baseline drive, or at the start of the second experimental drive, drivers experienced a ~1- minute free drive, after which the ADS was available. Drivers received a verbal-audio prompt: "Attention, engage automation" through the car's speakers, upon which, they could engage the ADS feature by pressing a button on the steering wheel. About a minute into the automated drive, a lead vehicle moved into the driving lane, starting the *ACF* segment.

ACF was followed by an auditory-verbal takeover request: "Attention, get ready to takeover". The takeover request was presented when the ego vehicle reached a section of the road with faded lane markings, representing a system limitation condition for the ADS. The takeover request was followed by a short acoustic tone (1000 Hz, lasting 0.2 s), with increasing frequency, until the driver resumed manual control. The ADS could be disengaged by either pulling the left-hand stalk, rotating the steering wheel by more than 2^{o}, or pressing the brake or accelerator pedals.

To study how the time headway of a lead vehicle affected workload, we introduced two kinds of takeover scenarios in this study, one with the lead vehicle present (*Lead*) and one without the lead vehicle (*No Lead*). In the *No Lead* condition, the lead vehicle exited the road at an intersection shortly before a takeover request

was given, and subsequently, a new lead vehicle joined the driving lane from the next intersection, ~10 s after participants resumed manual control of the vehicle. In the *Lead* condition, the lead vehicle continued in the driving lane when the takeover request was issued.

Each takeover was followed by a period of manual car-following (*MCF*). Each experimental drive consisted of *ACF*, *Takeover* and *MCF*, in that order, repeated twice, and incorporating the Lead Vehicle factor (see **Figure 4.2** and **Figure 4.4**).

Level of Automation determined whether or not participants could engage in NDRTs during automation. In the *L3* group, participants were not required to monitor the drive and were asked to engage in an 'Arrows Task' (Jamson & Merat, 2005) during automation. This task required participants to search and select the upward-facing arrow, in a 4x4 grid of arrows displayed on a touchscreen (**Figure 4.3a**), placed in the centre console, near the gearshift (

Figure 4.3b). The screen showed participants' cumulative score, as well as a "score to beat". To ensure full engagement in the Arrows task, participants were told they would receive an additional £5 if they beat this score, but, for ethical reasons, every participant was paid an additional £5 at the end of the study. This task was only available after engagement of automation, until when the takeover request was given. The instructions for participants in L2 automation was to monitor the road scene at all times, although they removed their hands from the steering wheel and foot off the pedals when automation was engaged.

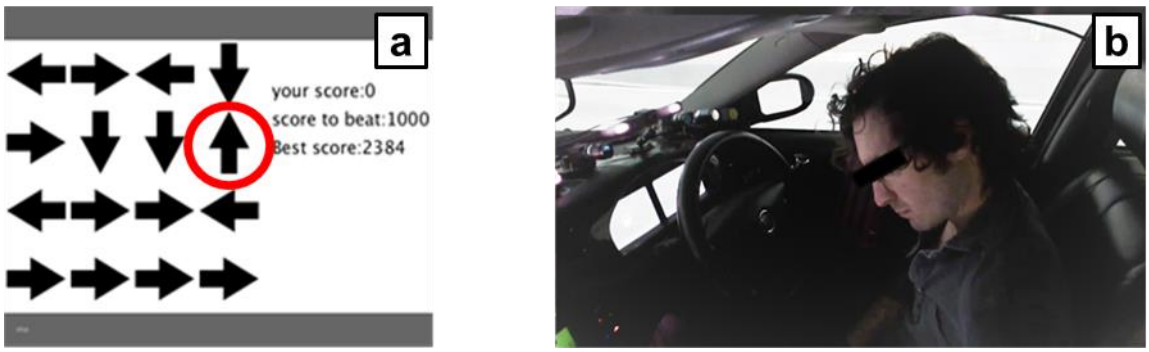

Figure 4.3. **(a)** A representation of the arrows task with the upward facing arrow circled in red; **(b)** A participant engaging in arrows task in L3 group during automation (in colour).

4.2.4 Self-reported workload ratings

Each participant was asked to rate their workload 13 times (7 times in the first experimental drive, including once during the manual baseline drive and 6 times during the second experimental drive). Response was provided verbally on a scale of 1-10, with 10 denoting highest workload. When they were engaged in either *ACF* or *MCF* (two each per experimental drive), they were prompted with the following verbal-auditory message ~ 2.5 minutes after starting *ACF* or *MCF* (roughly halfway through the drive): “Please rate your workload now”. Similarly, when they resumed manual control of the vehicle (two times per experimental drive), they were prompted with the following message 10 s after the takeover: “Please rate your workload during the takeover”.

4.2.5 Procedure

Upon arrival, participants were briefed with a description of the study, after which they were invited to sign a consent form, with an opportunity to ask questions about the study. Three ECG electrodes were then attached to the participant’s chest, and 2 EDA electrode bands were attached to the index and middle finger of their non-dominant hand. Once the participant was seated in the simulator cab, we collected the physiological baseline data, where participants were asked to relax for a period of 7 minutes with their eyes closed, palms on their laps facing upwards (Braithwaite et al., 2015; Laborde et al., 2017). This was used to standardise the experimental physiological data. Participants then performed a practice drive, which included both automated and manual driving. During the practice drive, participants were talked through the various aspects of the vehicle HMI, how to engage and disengage automation and those in the *L3* group practised the Arrows task. After the practice drive, participants experienced the two experimental drives, which lasted ~18 minutes each.

For all manual driving segments, participants were instructed to adhere to the posted speed limit of 40 mph and drive in the centre of the lane. They were also asked not to overtake the lead vehicle, but otherwise, follow the normal rules of the road, ensuring the safe operation of the vehicle, and maintaining their desired distance from the lead vehicle. After each experimental drive, the participants were given a 10-minute break, during which they were asked to complete a set of questionnaires,

including Arnett's Sensation Seeking Questionnaire (Arnett, 1994), traffic locus of control (Özkan & Lajunen, 2005) and the Driver Style Questionnaire (French et al., 1993), see Louw et al. (2020). However, results from these questionnaires are not reported here, since they did not include questions about driver workload.

4.2.6 Data analysis

To analyse drivers' workload during *ACF*, *MCF* and *Takeover*, the physiological data was first segmented into appropriate time windows. Participants provided verbal self-reported workload ratings, for each of the *ACF*, *MCF* and *Takeover* windows, as mentioned in section 2.4, which aligned with the segments used for physiological data collection. The time window for *ACF* was established as the time from when the lead vehicle entered the driving lane during the automated drive, until when the takeover request was given to the driver. The verbal response provided by the driver during this time window was considered as the self-reported workload rating for *ACF*.

Drivers' data from when they received the takeover request, until 10 s after they resumed manual control of the vehicle was classified as data for the *Takeover* window (see **Figure 4.4**). The subjective workload ratings for the *Takeover* window were provided 10 s after drivers resumed manual control of the vehicle. We took 10 s after takeover as the cut-off point, as previous research has shown that the peak in driving performance decrement is observed within 10-15 s after takeover of control (Merat et al., 2014). For the purpose of this study, we classified the time window from 10 s after resuming manual control until re-engaging automation as *MCF*, and the verbal response given by the participant during this window was considered as self-reported workload rating during *MCF*.

Drivers experienced two experimentally similar *ACF* and *MCF* scenarios, in each of the experimental drives, as seen in **Figure 4.2**. A set of t-tests applied to the physiological data, and self-reported workload ratings revealed no significant differences between the two *ACF* and two *MCF* scenarios, within each experimental drive. Therefore, in order to study changes in workload during automated and manual car-following, data for the two *ACF* scenarios, and the two *MCF* scenarios, in each experimental drive, were aggregated to a single representation. This was applied for both physiological data, and self-reported workload ratings.

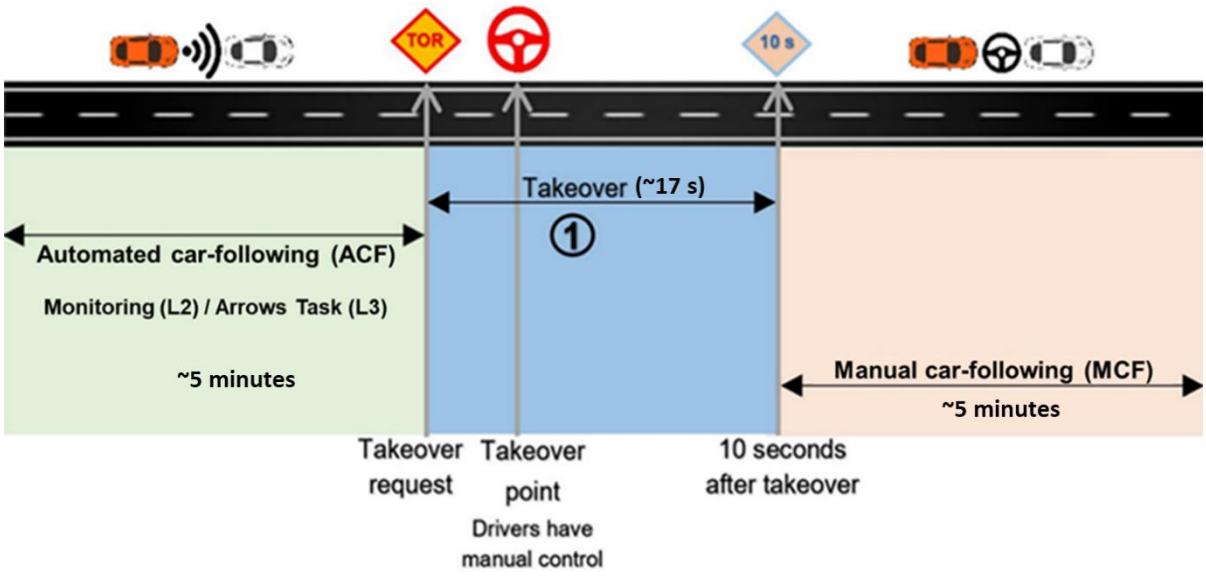

Figure 4.4. Schematic depicting the Time windows used for data analysis.

4.2.7 Data analysis tools

The ECG data was processed on Kubios HRV premium software (Tarvainen et al., 2014). EDA signals were pre-processed, and artefacts were removed using custom algorithms[2] based on recommendations from Braithwaite et al. (2015) and Kikhia et al. (2016), using MATLAB R2016a. The EDA data was analysed using Ledalab v3.9 (Benedek & Kaernbach, 2010), a MATLAB-based software package. The EDA signal was decomposed into tonic and phasic components, using continuous decomposition analysis (CDA; Benedek & Kaernbach, 2010). To identify a phasic event as an SCR, an amplitude threshold of 0.01 μS was used (Braithwaite et al., 2015). nSCR/min was computed as total number of SCRs in the window (above the amplitude threshold), divided by the time duration of the window (in seconds), which was then multiplied by 60, to get the number of SCRs per minute.

4.2.8 Statistical analysis

Statistical analysis was conducted with IBM SPSS Statistics 26. A Shapiro Wilk's test showed that the majority (> 75%) of the group-level estimates were normally distributed for each of the dependent variables used, for every ANOVA test we conducted. For statistical significance, an α-value of 0.05 was used as a limiting

[2] Described in detail in Chapter 2 of this thesis

criterion, and partial eta-squared was computed as an effect size statistic. Degrees of freedom were Greenhouse-Geisser corrected when Mauchly's test showed a violation of sphericity. Homogeneity of data was tested using Levene's test. In cases when the data was heterogeneous, the differences in group sizes were mostly equal (largest/smallest < 1.5), and the ANOVA was sufficiently robust to handle such heterogeneity in the data (Pituch & Stevens, 2016). Data from a participant was identified as an outlier if it was 3 times the interquartile range (IQR) above the 3rd quartile or below the 1st quartile, of the dataset. Due to a technical error with the voice recorder, self-reported workload ratings were missing for 5 participants, and self-reflected workload ratings data from a participant was identified as an outlier, across all the analyses done in this study. Therefore, data from these participants was excluded from the self-reported workload rating analysis. Data from 2 participants (one each from the *L2* and the *L3* group) in the EDR metric, nSCR/min metric and the RMSSD metric, were identified as outliers and excluded from the analysis. For analysis of RQ 3, data from 4 additional participants were identified as outliers, and removed from nSCR/min analysis.

For the self-reported workload ratings, only 50% of the dataset was normally distributed, and Levene's test revealed that workload ratings for the *Long* THW condition during *ACF* across the *L2* and *L3* groups were only slightly heterogeneous (p = .046), with equal group size. The skewness and kurtosis of the non-normally distributed data were within the acceptable range +/- 2 (George & Mallery, 2010). The one way ANOVA and the mixed ANOVA were robust enough to accommodate this violation of normality and homogeneity, with only a small effect on Type I error (Blanca et al., 2017; Pituch & Stevens, 2016).

4.3 Results

Since there is high inter-individual variability in physiological data, and this study incorporated a between-participant design, the physiological data during each of the *ACF*, *MCF* and *Takeover* time windows, for each participant, were standardised by representing the physiological values during *ACF*, *MCF* and *Takeover*, for each participant as a percentage of their respective physiological baseline data, collected before the experiment (see section 4.2.5).

4.3.1 The effect of Time Headway on driver workload during *ACF* in the *L2* group (RQ 1)

To understand how drivers' workload was affected by the two THW conditions, during the monitoring phase in the *L2* group, we performed a one-way ANOVA with repeated measures on drivers' RMSSD, mean HR, EDR, nSCR/min and self-reported workload ratings, with a within-participant factor of Time Headway (*Short*, *Long*).

As shown in **Table 4.1**, there were no main effects of Time Headway, across all the physiological metrics and self-reported workload ratings, with participants showing similar physiological activity, and self-reported workload ratings, in the *Short* and the *Long* THW conditions.

Table 4.1. Results of the one-way ANOVA with repeated measures (RQ1) across various physiological measures and subjective ratings, for the Time Headway condition, in the L2 group.

Predictor	df1	df2	F	p	η_p^2
1. RMSSD	1	11	.154	.702	.014
2. Mean HR	1	12	.001	.976	.000
3. EDR	1	11	1.77	.211	.138
4. nSCR/min	1	11	.041	.843	.004
5. Self-reported workload ratings	1	9	4.06	.075	.311

4.3.2 The effect of an NDRT on driver workload during *ACF* (RQ 2)

To understand how drivers' workload during *ACF* was affected by engagement in an NDRT, compared to just monitoring the drive, we performed a one-way ANOVA on drivers' RMSSD, mean HR, EDR, nSCR/min and self-reported workload ratings, with a between-participant factor of Level of Automation (*L2*, *L3*). As seen in section 4.3.1, in the *L2* group, participants exhibited similar physiological activity and self-reported workload ratings across the two THW conditions during *ACF*. In the *L3* group, participants were engaged in the visual Arrows task during *ACF,* without paying attention to the road and hence, did not observe the two THW conditions. Therefore,

we combined the physiological metrics and self-reported workload ratings during *ACF*, for *Short* and *Long* THW conditions, into a single representation, for each group.

Our results indicated that there was a significant main effect of Level of Automation, across all the physiological metrics (**Figure 4.5, Table 4.2**), during *ACF*, with drivers in the L3 group having significantly higher physiological activation, and hence, workload, when engaged in an NDRT, compared to those in the L2 group who were just monitoring the drive. Although statistically insignificant, a similar trend was observed in drivers' self-reported workload ratings, with drivers reporting higher workload when engaged in an NDRT (*L3* group), compared to monitoring the drive (*L2* group), during *ACF* (**Figure 4.6**).

Table 4.2. Results of the one-way ANOVAs across various physiological measures and self-reported workload ratings during ACF (RQ 2), for the Level of Automation condition.

Predictor	**df1**	**df2**	***F***	***p***	η_p^2
1. RMSSD	1	22	9.616	.005	.304
2. Mean HR	1	24	5.41	.029	.184
3. EDR	1	22	13.34	.001	.377
4. nSCR/min	1	22	5.35	.030	.196
5. Self-reported workload ratings	1	18	1.258	.277	.065

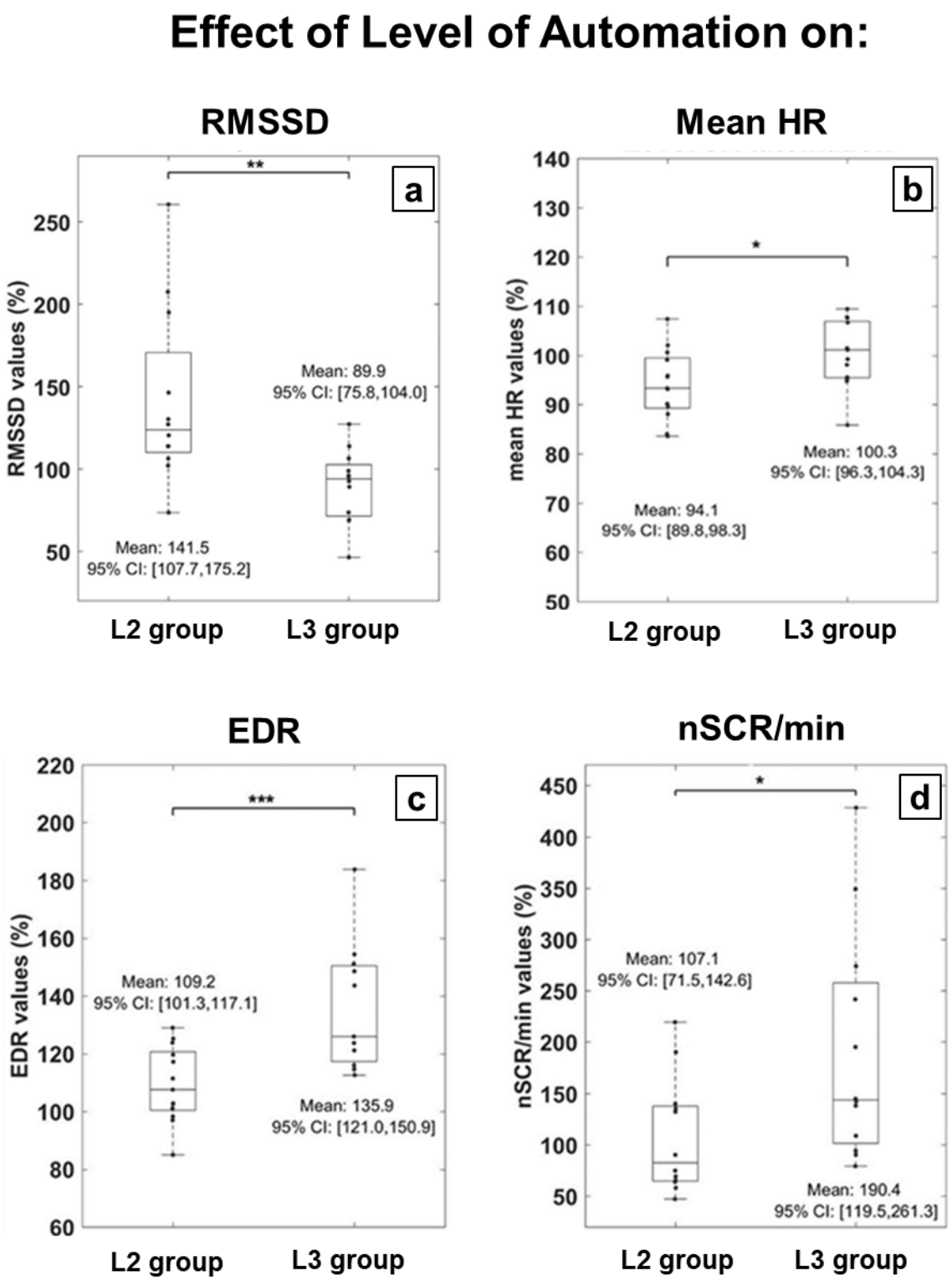

Figure 4.5. Effect of Level of Automation on workload during ACF as reflected by **(a)** RMSSD **(b)** Mean HR **(c)** EDR and **(d)** nSCR/min metric. *p≤ .05 , **p ≤ .01 and ***p≤.001.

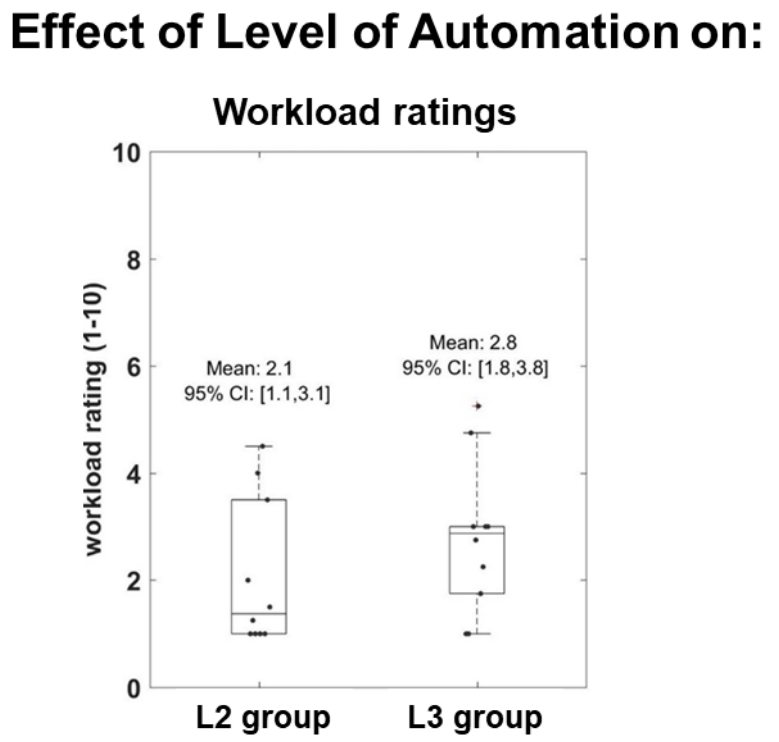

Figure 4.6. Effect of Level of Automation on drivers' self-reported workload ratings during ACF.

4.3.3 The effect of Drive Mode on driver workload (RQ3)

To understand how drivers' workload varied between *ACF* and *MCF*, we conducted a one-way ANOVA with repeated measures on drivers' RMSSD, mean HR, EDR, nSCR/min values, and self-reported workload ratings, with a within-participant factor of Drive Mode (*ACF*, *MCF*). This was done separately for the *L2* and *L3* groups, since drivers in the *L3* group were only exposed to NDRT during *ACF* and not during *MCF*. Since drivers had similar workload across the two THW conditions, for all the physiological metrics and self-reported workload ratings, we combined the values for *Short* and *Long* THW conditions into a single representation, for both *ACF* and *MCF*.

There was an effect of Drive Mode on drivers' EDR values, in the *L2* group (see **Table 4.3** and **Figure 4.8a**), with drivers having significantly higher respiration rates during *MCF*, compared to *ACF*. However, there was no statistically significant effect of Drive Mode on any other physiological metrics, or self-reported workload ratings, for the *L2* group (see **Table 4.3**). On the other hand, a significant effect of Drive Mode was seen across all the physiological metrics evaluated in the *L3* group (see **Table 4.3**). Drivers had significantly higher physiological activity during *ACF,* compared to

MCF (see **Table 4.3**, **Figure 4.7**, **Figure 4.8b**, **Figure 4.9a**), suggesting that a higher level of workload was imposed by engaging in the NDRT. Results from the ANOVA did not reveal any statistically significant differences in self-reported workload ratings between *ACF* and *MCF* in the *L3* group (see **Table 4.3** and **Figure 4.9b**).

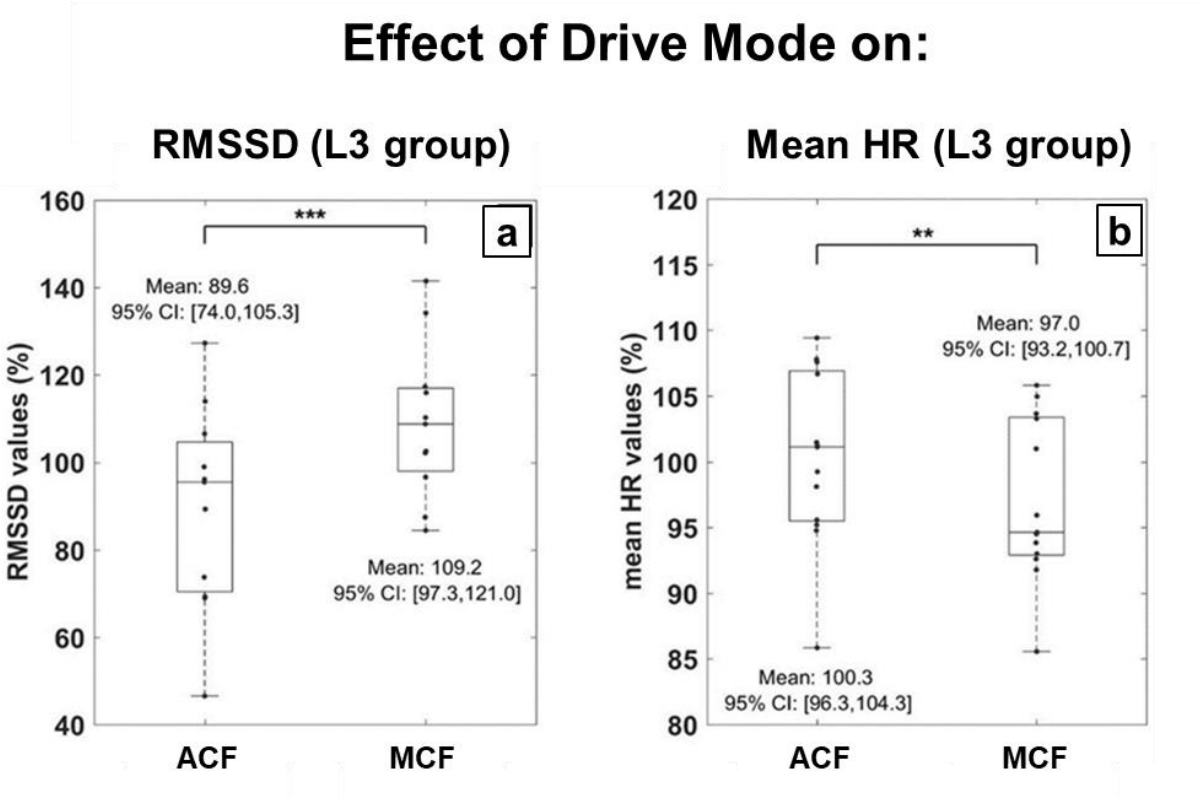

Figure 4.7. Effect of Drive Mode on workload in the L3 Group, as reflected in **(a)** RMSSD metric and **(b)** mean HR metric. **p ≤ .01 and ***p≤ .001.

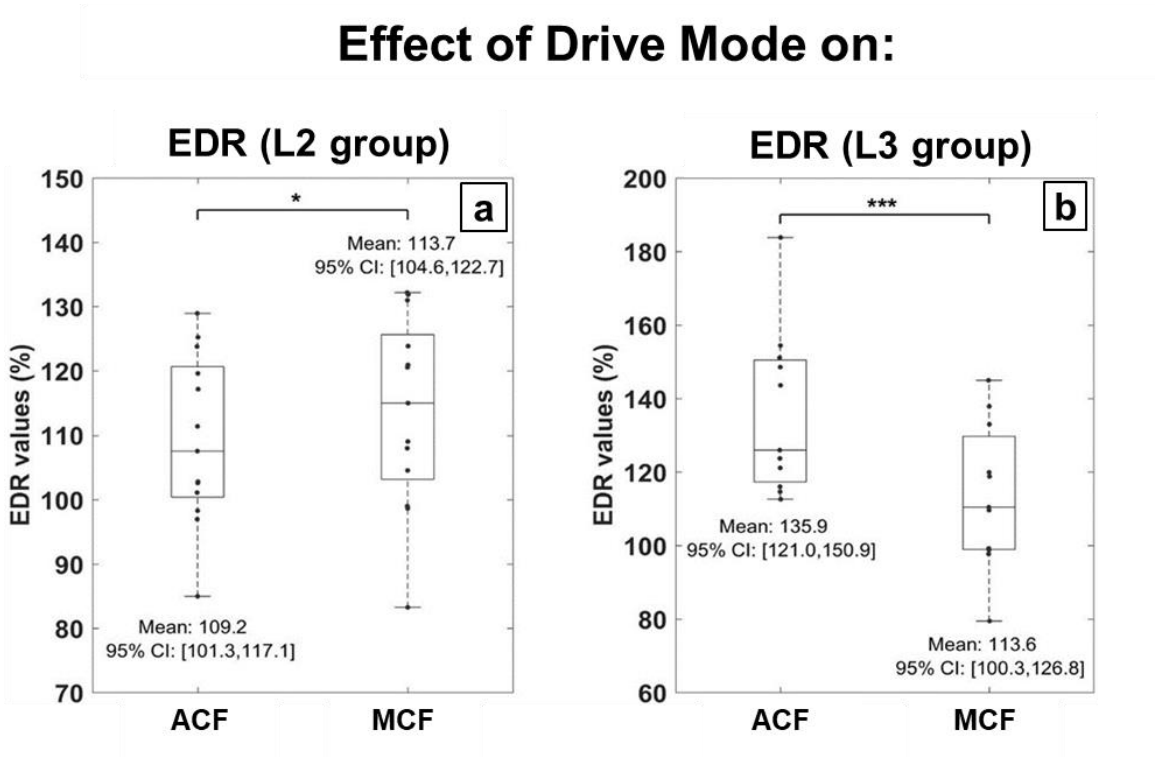

Figure 4.8. Effect of Drive Mode on workload, as reflected in **(a)** EDR metric in the L2 group and **(b)** EDR metric in the L3 group. *p ≤ .05 and **p ≤ .001.

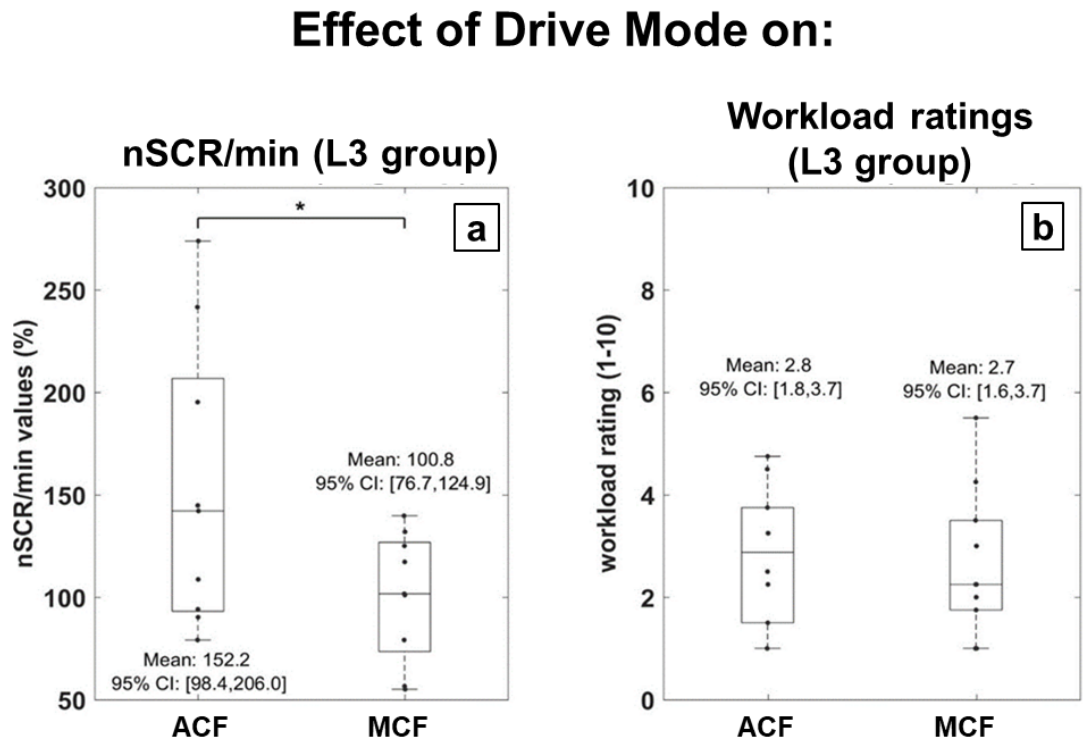

Figure 4.9. Effect of Drive Mode on workload in the L3 group, as reflected in **(a)** nSCR/min metric and **(b)** Self-reported workload ratings. *p ≤ .05.

Table 4.3. Results of one-way ANOVA with repeated measures across various physiological measures, on Drive Mode, in the L2 and the L3 group.

Predictor	df1	df2	*F*	*p*	η_p^2
1. RMSSD					
L2 group	1	10	.140	.716	.014
*L3 group **	1	10	21.99	.001	.687
2. Mean HR					
L2 group	1	12	.383	.548	.031
*L3 group **	1	12	15.99	.002	.571
3. EDR					
*L2 group **	1	10	6.08	.030	.336
*L3 group **	1	10	18.98	.001	.655
4. nSCR/min					
L2 group	1	12	.069	.797	.006
*L3 group **	1	8	7.39	.026	.480
5. Self-reported workload ratings					
L2 group	1	9	.421	.533	.045
L3 group	1	9	.040	.845	.004

* statistically significant at $p \leq .05$

4.3.4 The effect of Takeover on driver workload (RQ 4 and RQ 5)

ECG-based metrics, such as RMSSD, mean HR and EDR, were not analysed around takeovers, as accurate analysis of such metrics requires a minimum time window of 5 minutes (Malik et al., 2012), and in this study, the takeover time windows (shown in light blue**, Figure 4.4**) only lasted 15 - 20 seconds. As mentioned in section 4.2.6, nSCR/min reported in this section was computed using the data for Takeover Window (shown in light blue, **Figure 4.4**), that is, the time period from when they

received the takeover request, until 10 s after they resumed manual control of the vehicle.

To understand how driver workload was affected by the time headway of the lead vehicle during takeovers, and whether their workload level during the takeover was affected by their engagement in an NDRT during automation (which occurred prior to the takeover), we performed a 2x2 mixed ANOVA, including a within-participant factor of Time Headway (*Short*, *Long*) and between-participant factor of Level of Automation (*L2*, *L3*), on drivers' nSCR/min, and self-reported workload ratings. Only the *Lead* vehicle condition was considered here, as the Time Headway factor is irrelevant when there is no lead vehicle. For the nSCR/min metric, there was missing data from 5 participants, and data from 2 other participants were classified as outliers and excluded from the analysis.

As shown in **Table 4.3**, ANOVA results revealed that the effect of Time Headway on drivers' nSCR/min was nearing statistical significance (p = .056) with drivers of both groups having higher mean nSCR/min values during the *Short* THW condition, compared to the *Long* THW condition (see **Figure 4.10a** and **Table 4.4**). There was also an effect of Time Headway on drivers' self-reported workload ratings. Drivers had significantly higher self-reported workload ratings during the *Short* THW condition, compared to the *Long* THW condition (see **Figure 4.10b** and **Table 4.4**), similar to the trend observed in nSCR/min metric. There was no effect of Level of Automation or interaction effects, for either nSCR/min or self-reported workload ratings, suggesting that drivers of both group experienced similar levels of workload during the takeovers.

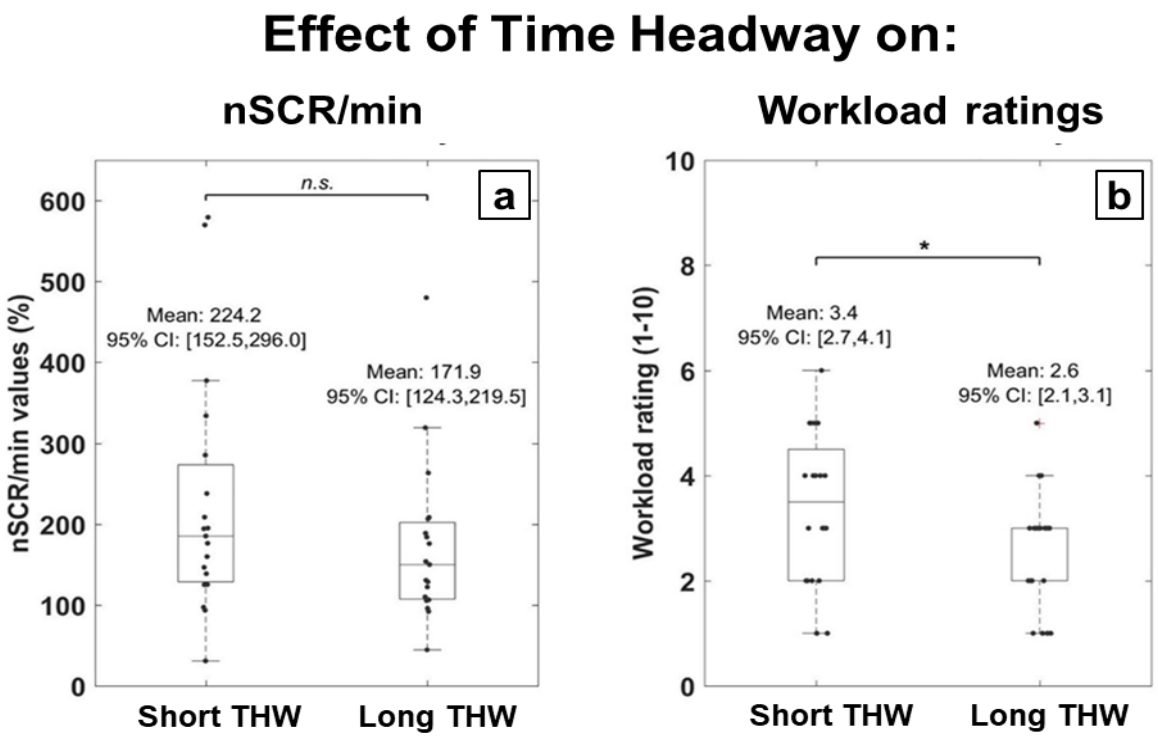

Figure 4.10. Effect of Time Headway condition when a lead vehicle was present, during takeover, **(a)** on drivers' nSCR/min metric and **(b)** self-reported workload ratings during. n.s. nearing significance (p = .056), *p ≤ .05

Table 4.4. Results of mixed ANOVAs across nSCR/min and self-reported workload ratings, during takeovers.

Predictor	df1	df2	*F*	*p*	η_p^2
1. nSCR/min					
Time Headway [n.s.]	1	17	4.21	.056	.198
Level of Automation	1	17	.830	.375	.047
2. Self-reported workload ratings					
Time Headway	1	18	5.36	.033	. 229
Level of Automation	1	18	.045	.834	.003

[n.s.] nearing significance.

4.4 Discussion and Conclusions

This study investigated how changes in driver workload, imposed by different demands from a simulator-based automated driving study, with car-following scenarios, affected drivers' physiological state and self-reported workload ratings. Two groups of participants were recruited to study how the presence of a lead vehicle, and its Time Headway (THW), affected workload during different stages of *L2* and *L3* automated driving. We also investigated whether engagement in an NDRT during *L3* automation increased driver workload.

Both physiological (ECG and EDA-based) metrics, and self-reported workload ratings indicated that the *L2* drivers experienced a similar level of workload when monitoring the lead vehicle during automation, whether this was in the *Short* (0.5 s) or *Long* (1.5 s) THW condition. This result is in contrast to results from a manual driving simulator study by Liu et al. (2019), who found significantly higher subjective feedback of workload for short THW conditions (0.5 and 1 s), compared to longer THWs of 2, 2.5 and 3 s. The absence of an effect of headway in the current study may be because the perceptual difference in the two levels of THW used was not prominent enough to affect drivers' workload levels, especially when they were simply monitoring the driving environment. Therefore, future studies should compare workload experienced at different THWs by both manual and automated driving, to understand if the workload experienced by short THWs is directly related to drivers' responsibility for the driving task.

When automation was engaged, drivers in the *L3* group (who were asked to conduct the Arrows task during automation) had significantly higher levels of workload, as illustrated by all the physiological metrics, when compared to the *L2* group. This suggests that the workload associated with both the physical movement (hand/finger), and the cognitive elements of the Arrows task in the *L3* group, was higher than that experienced by the simple monitoring of the drive by the *L2* group, although the between participant nature of this study must also be taken into account, when considering the implications of these results. These results are in agreement with Müller et al. (2021), who found a significant increase in drivers' workload levels, as reflected in their RMSSD values, when they were engaged in texting, using a touch-screen phone, compared to monitoring the drive, during an automated driving

simulator study. Our participants' self-reported workload ratings during this period of automation were found to be similar for the *L2* and *L3* groups. Hart & Wickens (1990) argue that such self-reported workload ratings reflect subjective impressions of workload, and may differ from the workload reflected by task performance, which may suggest that drivers underestimated their workload levels when performing the Arrows task.

When comparing the physiological metrics observed in manual and automated driving in the *L2* group, results were only significant for the ECG based EDR-metric, with an in increase in respiratory activity during manual driving, compared to automated driving. This is likely due to the sensitivity of the EDR-metric to the physical act of steering and pedal control, required in manual driving (Hidalgo-Muñoz et al., 2019; Omlin et al., 2016). All other physiological metrics, and also the self-reported workload ratings, indicated that drivers had similar cognitive workload levels when monitoring the drive during automation, compared to when they were engaged in manual driving. This result is similar to that of Lohani et al. (2021), and Stapel et al. (2019), who observed that drivers exhibited similar levels of cognitive workload, as indicated by their RMSSD values (Lohani et al., 2021), or subjective ratings (Stapel et al., 2019), during L2 automation and manual driving.

For the *L3* group, drivers showed significantly higher workload, as indicated by all the physiological metrics, when they were engaged in the Arrows task during automated driving, compared to the workload levels observed in manual driving. While it is difficult to accurately separate the physical and cognitive demands of the Arrows task, it is assumed that the physical demand of responding to this simple touchscreen task was lower than that of manual driving. Therefore, higher levels of physiological activity during automation in the *L3* group are more likely to be linked to the cognitive demand from the Arrows task.

To understand the sensitivity of different physiological indices to physical and cognitive demands within the driving context, our study incorporated three distinct scenarios: manual driving, the monitoring task in *L2*, and conducting the Arrows task in *L3*, each of which had varying levels of physical and cognitive demands. In terms of the physical demand, we proposed that the monitoring task had the lowest physical demand, with manual driving likely being more physically demanding than the Arrows

task. Given the similar driving environment across all the scenarios, we argue that the monitoring task and manual driving likely had similar cognitive demands (Lohani et al., 2021; Stapel et al., 2019), with the highest cognitive demand imposed by the Arrows task. The EDR was the only metric that picked up the changes in physical demand between monitoring and manual driving in the *L2* group, along with the differences in cognitive demand between the Arrows task and manual driving in the *L3* group, suggesting that it may be a better indicator of drivers' overall (physical and cognitive) workload levels in this study.

When considering how different Time Headways affected driver workload during takeovers across the *L2* and the *L3* groups, EDA-based nSCR/min were found to be marginally higher for the short THW condition, with the differences between the two THWs only approaching significance (p = .056). However, a significant difference was found for the self-reported ratings, with higher values for the *Short* THW condition. Therefore, although simple monitoring of a lead vehicle, which maintained a short headway during automation, did not seem to affect *L2* drivers' workload, an actual resumption of control while closely following a lead vehicle does seem to have increased drivers' perceived and actual workload levels. In terms of how workload at the takeover was affected by the previous activity (i.e. engaging in the Arrows task for the *L3* group, versus monitoring the drive in the *L2* group), results were similar for both physiological and self-reported workload ratings. This might suggest that the takeover was extremely demanding in itself, masking the effect of any activities that took place before the takeover request. On the other hand, since this was a non-critical takeover scenario, and drivers had adequate time to stop engaging in the Arrows task before resuming control, any additional workload imposed by the Arrows task may have been eliminated by this stage. Further research into safe and acceptable Time Headways maintained by AV controllers is, therefore, warranted, to ensure safe transition of control to the driver when required, without placing additional demand on the driver. Our results share some similarities with the findings of Dogan et al. (2019), who observed an increase in driver workload, measured by subjective ratings, with increases in the criticality of takeovers, influenced by shorter takeover lead time, and presence of an obstacle in the driving lane.

To conclude, our results illustrate the added value of psychophysiological metrics in identifying driver workload, during different stages of an automated drive. Our findings suggest that a demanding NDRT such as the Arrows task, can significantly increase drivers' workload levels, compared to just monitoring the drive, or manual driving. While the time headway conditions did not have an effect on drivers' workload levels when monitoring the drive during automation, we observed that the presence of a lead vehicle maintaining a shorter time headway significantly increased drivers' workload levels during the takeover, when they had to resume manual control of the vehicle. Although physiological signals are susceptible to motion artefacts, we were able to filter these out and objectively monitor drivers' workload levels, in a fully motion-based simulator environment, which emulates motion experienced during real-world driving. However, it is well known that physiological metrics are sensitive to a wide range of stimuli (Backs & Boucsein, 2000). To clearly interpret driver state using physiological metrics, especially in real-world scenarios, it is important to know the specific conditions and context which induced such physiological changes in drivers (Beggiato et al., 2019). Combining physiological metrics with camera-based sensors, and eye tracking data, can help to identify the cause of such physiological changes, and further improve driver state predictions. Therefore, further work is warranted to assess the real-time use of such physiological signals as part of future driver monitoring systems, to support drivers' safe operation of automated vehicles. Finally, it should be noted that in our study, participants in the *L3* group were always required to engage in an NDRT during automation. However, this would not necessarily be the case in real-world SAE L3 driving, where drivers could also choose to observe the driving environment. Our aim here, was to understand the impact of sustained eyes-off-road behaviour on physiological trends, making a clear distinction between the *L2* and *L3* groups. Future studies should investigate the impact of more naturalistic, self-regulated behaviours in SAE L3 driving.

4.5 References

Arnett, J. (1994). Sensation seeking: A new conceptualization and a new scale. *Personality and Individual Differences*, *16*(2), 289–296. https://doi.org/10.1016/0191-8869(94)90165-1

Backs, R. W., & Boucsein, W. (2000). *Engineering psychophysiology : issues and applications*. Lawrence Erlbaum.

Beggiato, M., Hartwich, F., & Krems, J. (2019). Physiological correlates of discomfort in automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *66*, 445–458. https://doi.org/10.1016/j.trf.2019.09.018

Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, *190*(1), 80–91. https://doi.org/10.1016/j.jneumeth.2010.04.028

Biondi, F. N., Lohani, M., Hopman, R., Mills, S., Cooper, J. M., & Strayer, D. L. (2018). 80 MPH and out-of-the-loop: Effects of real-world semi-automated driving on driver workload and arousal. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 1878–1882. https://doi.org/10.1177/1541931218621427

Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Datos no normales: ¿es el ANOVA una opción válida? *Psicothema*, *29*(4), 552–557. https://doi.org/10.7334/psicothema2016.383

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2015). *A Guide for Analysing Electrodermal Activity & Skin Conductance Responses (SCRs) for Psychophysiological Experiments*. https://doi.org/10.1017.S0142716405050034

Bruggen, A. (2015). An empirical investigation of the relationship between workload and performance. *Management Decision*, *53*(10), 2377–2389. https://doi.org/10.1108/MD-02-2015-0063

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (Eds.). (2007). *The Handbook of Psychophysiology* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511546396

Carsten, O., Lai, F. C. H., Barnard, Y., Jamson, A. H., & Merat, N. (2012). Control task substitution in semiautomated driving: Does it matter what aspects are automated? *Human Factors*, *54*(5), 747–761. https://doi.org/10.1177/0018720812460246

de Waard, D. (1996). *The Measurement of Drivers' Mental Workload*. s.n.

Dogan, E., Honnêt, V., Masfrand, S., & Guillaume, A. (2019). Effects of non-driving-related tasks on takeover performance in different takeover situations in conditionally automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *62*, 494–504. https://doi.org/10.1016/j.trf.2019.02.010

Du, N., Yang, X. J., & Zhou, F. (2020). Psychophysiological responses to takeover requests in conditionally automated driving. *Accident Analysis and Prevention*, *148*. https://doi.org/10.1016/j.aap.2020.105804

Endsley, M. R., & Kiris, E. O. (1995). The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(2), 381–394. https://doi.org/10.1518/001872095779064555

ERTRAC. (2017). *Automated Driving Roadmap*. European Road Transport Research Advisory Council. http://www.ertrac.org/uploads/documentsearch/id48/ERTRAC_Automated_Driving_2017.pdf

Evans, L. (1991). *Traffic Safety and the Driver*. Van Nostrand Reinhold.

Ferson, N., Pizzigoni, E., Garnier-Follet, F., & Val, C. (2019). A new methodology to model driver behaviour accounting for the variation in driving manner using naturalistic driving data. *26th International Technical Conference on The Enhanced Safety of Vehicles*, 1–8. https://www.researchgate.net/publication/338549089_A_new_methodology_to_model_driver_behaviour_accounting_for_the_variation_in_driving_manner_using_naturalistic_driving_data

Foy, H. J., & Chapman, P. (2018). Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation. *Applied Ergonomics*, *73*, 90–99. https://doi.org/10.1016/j.apergo.2018.06.006

French, D. J., West, R. J., Elander, J., & Wilding, J. M. (1993). Decision-making style, driving style, and self-reported involvement in road traffic accidents. *Ergonomics*, *36*(6), 627–644. https://doi.org/10.1080/00140139308967925

George, D., & Mallery, P. (2010). SPSS for Windows step by step: A simple guide and

reference,17.0 update. In *Boston: Pearson Education, Inc* (10th ed.). Allyn & Bacon.

Gold, C., Berisha, I., & Bengler, K. (2015). Utilization of drivetime - Performing non-driving related tasks while driving highly automated. *Proceedings of the Human Factors and Ergonomics Society*. https://doi.org/10.1177/1541931215591360

Hart, S. G., & Wickens, C. D. (1990). Workload Assessment and Prediction. In H. R. Booher (Ed.), *Manprint* (pp. 257–296). Springer Netherlands. https://doi.org/10.1007/978-94-009-0437-8_9

Hidalgo-Muñoz, A. R., Béquet, A. J., Astier-Juvenon, M., Pépin, G., Fort, A., Jallais, C., Tattegrain, H., & Gabaude, C. (2019). Respiration and Heart Rate Modulation Due to Competing Cognitive Tasks While Driving. *Frontiers in Human Neuroscience*, *12*, 525. https://doi.org/10.3389/fnhum.2018.00525

IEEE. (2019). *New Level 3 Autonomous Vehicles Hitting the Road in 2020*. IEEE Innovation at Work. https://innovationatwork.ieee.org/new-level-3-autonomous-vehicles-hitting-the-road-in-2020/

Jamson, A. H., & Merat, N. (2005). Surrogate in-vehicle information systems and driver behaviour: Effects of visual and cognitive load in simulated rural driving. *Transportation Research Part F: Traffic Psychology and Behaviour*. https://doi.org/10.1016/j.trf.2005.04.002

Kikhia, B., Stavropoulos, T. G., Andreadis, S., Karvonen, N., Kompatsiaris, I., Sävenstedt, S., Pijl, M., & Melander, C. (2016). Utilizing a wristband sensor to measure the stress level for people with dementia. *Sensors (Switzerland)*, *16*(12). https://doi.org/10.3390/s16121989

Kinnear, N., Stuttard, N., Hynd, D., Helman, S., & Edwards, M. (2021). *Safe performance of other activities in conditionally automated vehicles Automated Lane Keeping System Report details*. Transport Research Laboratory, UK. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/978409/safe-performance-of-other-activities-in-conditionally-automated-vehicles.pdf

Laborde, S., Mosley, E., & Thayer, J. F. (2017). Heart rate variability and cardiac vagal tone in psychophysiological research - Recommendations for experiment

planning, data analysis, and data reporting. In *Frontiers in Psychology* (Vol. 8, p. 213). Frontiers. https://doi.org/10.3389/fpsyg.2017.00213

Li, Y., Wang, H., Wang, W., Xing, L., Liu, S., & Wei, X. (2017). Evaluation of the impacts of cooperative adaptive cruise control on reducing rear-end collision risks on freeways. *Accident Analysis and Prevention*, *98*, 87–95. https://doi.org/10.1016/j.aap.2016.09.015

Liu, K., Green, P., & Liu, Y. (2019). Traffic and Ratings of Driver Workload: The Effect of the Number of Vehicles and Their Distance Headways. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 2134–2138. https://doi.org/10.1177/1071181319631051

Lohani, M., Cooper, J. M., Erickson, G. G., Simmons, T. G., McDonnell, A. S., Carriero, A. E., Crabtree, K. W., & Strayer, D. L. (2021). No Difference in Arousal or Cognitive Demands Between Manual and Partially Automated Driving: A Multi-Method On-Road Study. *Frontiers in Neuroscience*, *15*. https://doi.org/10.3389/fnins.2021.577418

Louw, T., Goncalves, R., Torrao, G., Radhakrishnan, V., Lyu, W., Puente Guillen, P., & Merat, N. (2020). Do drivers change their manual car-following behaviour after automated car-following? *Cognition, Technology and Work*. https://doi.org/10.1007/s10111-020-00658-5

Louw, T., Madigan, R., Carsten, O., & Merat, N. (2017). Were they in the loop during automated driving? Links between visual attention and crash potential. *Injury Prevention*, *23*, 281–286. https://doi.org/10.1136/injuryprev-2016-042155

Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J., & Schwartz, P. J. (2012). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, *17*(3), 354–381. https://doi.org/10.1093/oxfordjournals.eurheartj.a014868

Mehler, B., Reimer, B., Coughlin, J., & Dusek, J. (2009). Impact of Incremental Increases in Cognitive Workload on Physiological Arousal and Performance in Young Adult Drivers. *Transportation Research Record: Journal of the Transportation Research Board*, *2138*, 6–12. https://doi.org/10.3141/2138-02

Merat, N., Jamson, A. H., Lai, F. C. H., & Carsten, O. (2012). Highly automated driving,

secondary task performance, and driver state. *Human Factors*, *54*(5), 762–771. https://doi.org/10.1177/0018720812442087

Merat, N., Jamson, A. H., Lai, F. C. H., Daly, M., & Carsten, O. M. J. (2014). Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*(PB), 274–282. https://doi.org/10.1016/j.trf.2014.09.005

Merat, N., Seppelt, B., Louw, T., Engström, J., Lee, J. D., Johansson, E., Green, C. A., Katazaki, S., Monk, C., Itoh, M., McGehee, D., Sunda, T., Unoura, K., Victor, T., Schieben, A., & Keinath, A. (2018). The “Out-of-the-Loop” concept in automated driving: proposed definition, measures and implications. *Cognition, Technology & Work*, 1–12. https://doi.org/10.1007/s10111-018-0525-8

Meteier, Q., Capallera, M., Ruffieux, S., Angelini, L., Abou Khaled, O., Mugellini, E., Widmer, M., & Sonderegger, A. (2021). Classification of Drivers’ Workload Using Physiological Signals in Conditional Automation. *Frontiers in Psychology*, *12*, 596038. https://doi.org/10.3389/fpsyg.2021.596038

Müller, A. L., Fernandes-Estrela, N., Hetfleisch, R., Zecha, L., & Abendroth, B. (2021). Effects of non-driving related tasks on mental workload and take-over times during conditional automated driving. *European Transport Research Review*, *13*(1), 1–15. https://doi.org/10.1186/s12544-021-00475-5

National Highway Traffic Safety Administration. (2009). *Traffic Safety Facts 2009: A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System*. U.S. Department of Transportation. http://www.nhtsa.gov/NCSA.

NTSB. (2017). *Highway Accident Report: Collision Between a Car Operating With Automated Vehicle Control Systems and a Tractor-Semitrailer Truck Near Williston, Florida, May 7, 2016*. https://www.ntsb.gov/investigations/AccidentReports/Reports/HAR1702.pdf%0A https://trid.trb.org/view/1485316

Omlin, X., Crivelli, F., Heinicke, L., Zaunseder, S., Achermann, P., & Riener, R. (2016). Effect of rocking movements on respiration. *PLoS ONE*, *11*(3). https://doi.org/10.1371/journal.pone.0150581

Orsila, R., Virtanen, M., Luukkaala, T., Tarvainen, M., Karjalainen, P., Viik, J., & Savinainen, M. (2008). Perceived mental stress and reactions in heart rate variability—a pilot study among employees of an electronics company. *International Journal of Occupational Safety and Ergonomics*, *14*(3), 275–283. https://doi.org/10.1080/10803548.2008.11076767

Özkan, T., & Lajunen, T. (2005). Multidimensional Traffic Locus of Control Scale (T-LOC): Factor structure and relationship to risky driving. *Personality and Individual Differences*, *38*(3), 533–545. https://doi.org/10.1016/j.paid.2004.05.007

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs. *Journal of Cognitive Engineering and Decision Making*, *2*(2), 140–160. https://doi.org/10.1518/155534308X284417

Pituch, K. A., & Stevens, J. P. (2016). Applied Multivariate Statistics for the Social Sciences. In *Applied Multivariate Statistics for the Social Sciences* (Sixth). Routledge, Taylor and Francis group. www.routledge.com/9780415836661

Radhakrishnan, V., Merat, N., Louw, T., Lenné, M. G., Romano, R., Paschalidis, E., Hajiseyedjavadi, F., Wei, C., & Boer, E. R. (2020). Measuring drivers' physiological response to different vehicle controllers in highly automated driving (HAD): Opportunities for establishing real-time values of driver discomfort. *Information (Switzerland)*, *11*(8), 390. https://doi.org/10.3390/INFO11080390

SAE International. (2021). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. https://www.sae.org/standards/content/j3016_202104/

Siebert, F. W., & Wallis, F. L. (2019). How speed and visibility influence preferred headway distances in highly automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *64*, 485–494. https://doi.org/10.1016/j.trf.2019.06.009

Stapel, J., Mullakkal-Babu, F. A., & Happee, R. (2019). Automated driving reduces perceived workload, but monitoring causes higher cognitive load than manual driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *60*, 590–605. https://doi.org/10.1016/j.trf.2018.11.006

Tarvainen, M. P., Niskanen, J. P., Lipponen, J. A., Ranta-aho, P. O., & Karjalainen, P. A. (2014). Kubios HRV - Heart rate variability analysis software. *Computer Methods and Programs in Biomedicine*, *113*(1), 210–220. https://doi.org/10.1016/j.cmpb.2013.07.024

Vogel, K. (2003). A comparison of headway and time to collision as safety indicators. *Accident Analysis and Prevention*, *35*(3), 427–433. https://doi.org/10.1016/S0001-4575(02)00022-2

Wandtner, B., Schömig, N., & Schmidt, G. (2018). Effects of Non-Driving Related Task Modalities on Takeover Performance in Highly Automated Driving. *Human Factors*, *60*(6), 870–881. https://doi.org/10.1177/0018720818768199

Young, M. S., & Stanton, N. A. (2002). Malleable Attentional Resources Theory: A New Explanation for the Effects of Mental Underload on Performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *44*(3), 365–375. https://doi.org/10.1518/0018720024497709

Zeeb, K., Buchner, A., & Schrauf, M. (2016). Is take-over time all that matters? the impact of visual-cognitive load on driver take-over quality after conditionally automated driving. *Accident Analysis and Prevention*, *92*, 230–239. https://doi.org/10.1016/j.aap.2016.04.002

5 EFFECT OF DRIVERS' ATTENTIONAL DEMAND ON MENTAL WORKLOAD, AT DIFFERENT STAGES OF AUTOMATION

ABSTRACT: This Horizon2020-funded driving simulator-based study on automated driving investigated the effect of different car-following scenarios, and takeover situations, on drivers' mental workload, as measured by eye tracking-based metrics of pupil diameter and self-reported workload ratings. This study incorporated a mixed design format, with 16 drivers recruited for the SAE Level 2 (L2; SAE International, 2021) automation group, who were asked to monitor the driving and road environment during automation, and 16 drivers in the Level 3 (L3) automation group, who engaged in a non-driving related task (NDRT) during automation. Drivers in each group undertook two experimental drives, lasting about 18 minutes each. To manipulate perceived workload, difficulty of the driving task was controlled by incorporating a lead vehicle which maintained either a Short (0.5 s) or Long (1.5 s) Time Headway (THW) condition during automated car-following (ACF). Each ACF session was followed by a subsequent request to takeover, which happened either in the presence or absence of a lead vehicle. Results from standard deviation of pupil diameter values indicated that drivers' mental workload levels fluctuated significantly more when monitoring the drive during L2 ACF, compared to manual car-following (MCF). Additionally, we found that drivers' mental workload, as indicated by their mean pupil diameter, increased steeply around takeovers, and was further exacerbated by the presence of a lead vehicle during the takeovers, especially in the Short THW condition, for both groups. Pupil diameter was found to be sensitive to subtle variations in mental workload, and closely resembled the trend seen in self-reported workload ratings. Further research is warranted to assess the feasibility of using eye tracking-based metrics along with other physiological sensors, in real-world settings, to aid real-time indications of drivers' mental workload, for further improvement of future driver state monitoring systems.

5.1 Introduction

With the recent push towards more automated control of vehicles, human-factors research into how this affects driver state has also gained momentum. However, we are still a long way from achieving full autonomy of the driving task, where drivers are not required to intervene at all. Especially for automation levels that require some form of driver intervention, such as SAE Level 2 (L2) or Level 3 (L3; SAE International, 2021), drivers are required to have appropriate "readiness levels" (Gold

et al., 2016; Zeeb et al., 2016) to safely resume control of the vehicle, when the system reaches limiting criteria, based on its Operational Design Domain, or when an unexpected event/fault causes the automated system to relinquish control to the driver (Mioch et al., 2017; Parasuraman et al., 2008).

Inappropriate levels of driver workload (underload or overload) can influence drivers' performance, safety, and readiness, especially when they have to resume manual control of an automated vehicle (Dogan et al., 2019; Parasuraman et al., 2008). Mental workload is described as the relation between the mental resources demanded by a task/activity, and an individual's information processing capacity, or the mental resources that are available to be supplied to the driver (de Waard, 1996; Parasuraman et al., 2008). Mental workload and task performance follow a complex and 'inverted U-shape' relationship (Bruggen, 2015; de Waard, 1996). Drivers have limited cognitive resources at their disposal, including a central resource pool that is used to perform all tasks (Kahneman, 1973), and additional multiple resources (such as visual or auditory), that are utilised, based on the task demand and modality. (Wickens, 1984). The size of this resource pool can vary in capacity, based on the task demand (Young & Stanton, 2002b). For example, exceptionally low workload levels (underload), resulting from low task demand, and/or the monotony imposed by automation, can reduce drivers' vigilance, or ability to maintain sustained attention (Young & Stanton, 2002b). This also reduces drivers' ability to detect or perceive risk/hazards, resulting in deteriorating performance of the driving task, if and when drivers are required to resume control of the vehicle (Heikoop et al., 2019). Similarly, very high workload (overload) and increased demand from a competing task, such as engaging in a demanding non-driving related task (NDRT), can also result in a performance decrement, because drivers' physical and mental attention is taken away from the primary driving task (Gold et al., 2015; Zeeb et al., 2016). Therefore, in order to successfully resume control from automation and maintain appropriate readiness levels, drivers should maintain moderate workload and arousal levels, to reduce the likelihood of driving errors (Bruggen, 2015; de Waard, 1996). This paper focuses on understanding how drivers' mental workload is influenced by a series of car-following situations in manual and automated driving, and takeover scenarios after automation, in a simulator-based study on highly automated driving (HAD), investigating whether

eye-based metrics can be used as an objective, non-invasive, indicator of drivers' mental workload.

Studies that have compared drivers' mental workload between monitoring the drive during automation and manual driving, suggest that drivers had similar levels of mental workload, and did not experience automation-induced underload during a monitoring phase of the drive, compared to manual driving (Lohani et al., 2020; Stapel et al., 2019). However, these studies measured average workload over a ~30-minute time window, and relied on subjective or heart-rate based measures. Observation studies within the aviation domain that have used subjective ratings, and a detection response task as performance indicators of workload (Wiener, 1989; Sarter, Woods, & Billings, 1997), have suggested that automation results in an uneven distribution of mental workload, rather than simply reducing or increasing it. In a recent study on workload in unmanned aerial vehicle (UAV) operators, Boehm et al. (2021) observed that mental workload can vary rapidly (in a matter of seconds) as a function of fluctuating task demands, and affects the performance of the operator. However, there are limited studies that have focused on understanding the moment-to-moment or phasic changes in drivers' mental workload (that is, rapid fluctuations in their workload levels) while monitoring the drive during automation, how this is different to fluctuations in mental workload during manual driving, and whether it affects driving performance during the resumption of control from automation.

Another factor influencing drivers' mental workload is car-following scenarios, where changes in the distance maintained from a lead vehicle can affect drivers' attentional demands (Liu et al., 2019; Siebert & Wallis, 2019). Car-following refers to the longitudinal following of a lead vehicle by a driver or automated system, and is a pre-cursor to rear-end collisions, which constitutes over 31% of all collisions in the US (National Highway Traffic Safety Administration, 2009, p. 56). Time headway (THW), which is the elapsed time between the front of a lead vehicle passing the road, and the front of the ego vehicle passing the same point, has been used as a behavioural measure for understanding car-following (K. Vogel, 2003). Using a simulator study with manual driving, Liu et al. (2019), found that drivers reported significantly higher workload when the lead vehicle maintained shorter THWs of 1 s or less, compared to longer THWs of 1.5 s or more. Shorter THWs maintained by the ego vehicle have also

been associated with higher risk perception (Louw et al., 2020) and discomfort (Siebert & Wallis, 2019), which can also lead to an increase in drivers' workload levels (Beggiato et al., 2019), during automated car-following scenarios. Therefore, to understand how drivers' mental workload was affected by the THW of a lead vehicle during both automated (ACF) and manual car-following (MCF), as well as the takeover period, this study included a series of car-following scenarios with different THWs.

Task switching within the automated driving context, for example, from monitoring the drive in a passive supervisory role, to resuming control of the vehicle in an active role as a driver, can increase driver workload (Monsell, 2003; Wickens et al., 2015). When a takeover request is issued, drivers are required to refocus their attention to the driving environment, and pay more attention to the road and any potential hazards. This sudden change in attentional demand from simply monitoring the drive, to the attention and effort required to perform a successful takeover, can result in a physical and mental overload during takeovers, where the task demand exceeds the resources available to the driver.

Accurate, objective, measurement of drivers' mental workload levels can be challenging, due to the high inter-individual variability of how people perceive, and are affected by, different task demands. Nevertheless, from a performance and safety standpoint, it is crucial that we are able to objectively measure and understand variations in driver workload during HAD, as this can inform the automated system about the capabilities and limitations of the driver, especially if they have to resume control of the vehicle. Eventually, the vehicle can establish if the driver is indeed capable of taking back control of the vehicle, or if other measures, such as a Minimum Risk Manoeuvre (Thorn et al., 2018) should be considered to bring the car to a safe stop (Yu et al., 2021).

Skin conductance and heart-based sensors have been used extensively within the driving context, as objective indicators of mental workload (Foy & Chapman, 2018; Radhakrishnan et al., 2022; Reimer & Mehler, 2011), stress (Healey & Picard, 2005) or discomfort (Beggiato et al., 2019; Radhakrishnan et al., 2020). However, as a stand-alone measure, such skin conductance and heart-based measures are unable to identify the causal factor (such as stress, attention, workload, or arousal) that induce physiological changes, and the addition of other sensors, such as eye tracking, can

aid our understanding of driver state. Also, skin conductance and heart-based sensors are affected by physical load, and in tasks such as manual driving, which involves both physical and mental load, it would be challenging to distinguish between physical and mental workload, using skin conductance or heart-based sensors alone. Moreover, some heart-based sensors can be intrusive in nature, and require extended preparation, prior to a study. Dash-based eye-trackers can be both non-invasive and unobtrusive (Marquart et al., 2015; Merat et al., 2012; Tsai et al., 2007), and eye tracking-based metrics, such as pupil diameter, blink frequency and blink duration, have been used in the past as indicators of mental workload, where shorter blink durations lead to high blink frequency, and vice versa. The relationship between blink frequency and duration, and the demand from different tasks is not always clear. For example, visual tasks can reduce blink frequency (Veltman & Gaillard, 1996), whereas a non-visual cognitive task (n-back) can increase blink frequency (Recarte et al., 2008). Given this ambiguity between blink duration and frequency and their relationship to workload in this paper, we only assessed pupil diameter, and its feasibility as an indicator of real-time driver workload.

Studies have shown that pupil diameter generally increases with an increase in drivers' mental load (Tsai et al., 2007). Mean pupil diameter has been used as an indicator of tonic dilation/constriction, with mean pupil diameter values increasing with an increase in mental load (Appel et al., 2018; Steinhauer et al., 2004), although these are generally 11% or lower, compared to baseline values (Batmaz & Öztürk, 2008; John A. Stern, 1997). However, one disadvantage of using mean pupil diameter is that it fails to capture the phasic fluctuations in pupil diameter, as the high and low fluctuations can cancel each other out, when taking the mean value of pupil diameter over the entire time series (Buettner, 2013). The standard deviation of pupil diameter has therefore been used as an indicator phasic fluctuations, with higher standard deviation of pupil diameter suggesting a higher fluctuation in drivers' workload levels (Beatty, 1982; Buettner, 2013). Pupil diameter is also known to be affected by individual and environmental differences, with variations in pupil diameter due to changes in light intensity being significantly higher than variations due to mental workload (Mathôt, 2018). For example, pupil diameter can increase in darker conditions to accommodate more light into the eyes (pupil light response, Mathôt, 2018; Spector, 1990). A shift in focus between near objects to those further away (pupil

near response or accommodation reflux), can also result in pupil dilation (Mathôt, 2018), while a sudden change in the environment, for example, caused by sounds, movement or touch (orienting response), can also cause a small increase in magnitude or duration of pupil dilation, within 0.5 s to 1s of stimulus onset (Mathôt, 2018).

5.1.1 Current study

This study was funded by the European Commission Horizon 2020 program under the L3Pilot project, grant agreement number 723051, and investigated how drivers' mental workload, as measured by pupil diameter changes, varied during different stages of manual and automated driving. This study is also discussed in Radhakrishnan et al. (2022)[3], where the authors investigated the effect of driver workload on heart-rate and skin conductance-based physiological signals. Using a mixed design, drivers were asked to monitor the drive during automation (L2) or engage in an NDRT (L3). Driver workload was also manipulated using two THW conditions, in a series of car-following scenarios (Short and Long), and two different types of takeovers, that happened either in the presence or absence of a lead vehicle (Lead/No Lead). To understand how drivers' mental workload varied during the monitoring phase, we compared drivers' mental workload levels during automated car-following (ACF) and manual car-following (MCF). To understand how drivers' mental workload varied at different points during the transition of control from automation to manual driving, we analysed drivers' mental workload levels before (Pre-Takeover), during (Takeover) and after resumption of control (Post-Takeover). We also investigated whether the presence of a lead vehicle, or shorter headway conditions, increased their mental workload during takeovers. Additionally, we investigated whether prior engagement in an NDRT (L3), as opposed to monitoring the drive during automation (L2), affected drivers' mental workload during the takeover. Eye tracking-based measure of pupil diameter was used as an objective measure of drivers' mental workload, and drivers were also asked to report their perceived workload using a

[3] The work mentioned here represents Chapter 4 of this thesis.

subjective workload ratings scale (1-10, with 10 being extremely high workload). The research questions addressed by this paper were:

RQ 1. Does drivers' mental workload, as indicated by pupil diameter and self-reported workload ratings, vary between automated car-following (ACF) and manual car-following (MCF), in the L2 group (ACF vs MCF)?

RQ 2. Is drivers' mental workload during the takeover different for the different takeover windows in the L2 group (Pre-Takeover, Takeover, and Post-Takeover)?

RQ 3. Is drivers' mental workload during the takeover affected by the THWs maintained by the automated controller (Short vs Long)?

RQ 4. Is drivers' mental workload during the takeover affected by the presence of a lead vehicle during the takeover (Lead vs No Lead)?

RQ 5. Is drivers' mental workload during the takeover affected by prior engagement in an NDRT during automation (L2 vs L3)?

5.2 Materials and Methods

5.2.1 Participants

For this driving simulator study, a total of 32 participants were recruited, with 16 participants each in the L2 and L3 groups. However, the data from 3 participants from each group were discarded due to either missing data, or because they did not follow the instructions of the study. The remaining 13 participants (4 females, 9 males) in the L2 group had a mean age of 42 ± 17 years, with a mean driving experience of 22 ± 16 years. For the L3 group, the 13 participants (3 females, 11 males) had a mean age of 33 ± 8 years, with a mean driving experience of 14 ± 8 years. In accordance with the rules and regulations of the University of Leeds ethics committee (LTTRAN-054), all participants gave prior consent to take part in the study. Additionally, they were compensated with £25 upon completion of the experiment.

5.2.2 Apparatus

The University of Leeds Driving Simulator (UoLDS), which is a full motion-based driving simulator, was used to conduct the experiment. UoLDS consists of a 4m diameter spherical projection dome, which has a 300º field of view projection system, and also includes a Jaguar S-type cab, housed within the dome. The electrical motion

system for the simulator has 8 degrees of freedom, and consists of a 500 mm stroke-length hexapod motion platform, carrying the 2.5 T payload of the dome and vehicle cab combination, and allowing movement in all six orthogonal degrees of freedom of the Cartesian inertial frame. The electrical motion platform is mounted on a railed gantry, which further allows 5m of effective motion in surge and sway. Drivers' eye tracking data was recorded using Seeing Machines FOVIO eye tracking hardware system and Seeing Machines PC-DMS software package.

When engaged, the Automated Driving System (ADS) used in this study was designed to control both lateral and longitudinal aspects of the driving task, driving at a constant speed of 40 mph. When the control of the vehicle was with the driver during manual driving, and the ADS was inactive, the HMI interface on the dashboard displayed a red steering icon (**Figure 1a**), When the ADS was activated, the HMI interface displayed a green steering icon (**Figure 1b**).

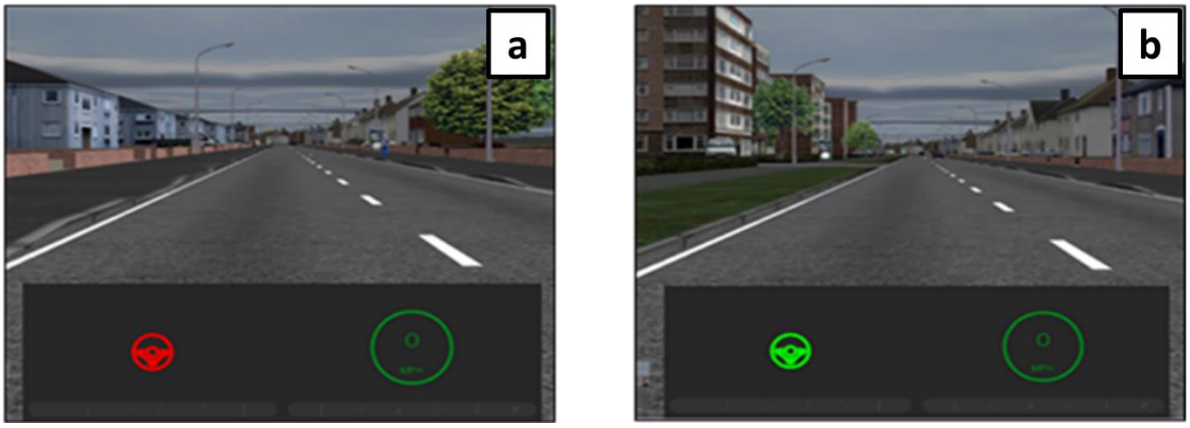

Figure 5.1. HMI Interface on the dashboard: **(a)** when automation was disengaged; **(b)** Automation was engaged.

5.2.3 Study design

This study has already been presented in Radhakrishnan et al. (2022). A mixed design was incorporated in this study, with between-participant factor of Level of Automation (L2, L3) and a within-participant factor of Time Headway (Short, Long), Drive Mode (ACF, MCF) and Lead Vehicle (Lead, No Lead). With the exception of Drive Mode, all the other factors were counterbalanced.

Each participant undertook a ~10-minute practice drive, to become familiar with the simulator environment, and the driving controls. After the practice drive, each

participant experienced two experimental drives, with a 5-minute break in between. All the experimental drives were performed in a single-carriageway urban environment, with a 40 mph speed limit, and low-density oncoming traffic. The first experimental drive consisted of free driving for ~2 minutes, following which a lead vehicle joined the driving lane. The participants were instructed to follow the lead vehicle as they normally would in such a scenario, without overtaking, for about 5 minutes. This was termed the manual baseline drive, and was used to collect drivers' baseline car-following behaviour, prior to any experimental manipulations. With the exception of the manual baseline drive, which was only present in the first experimental drive, the two experimental drives were similar. Each experimental drive consisted of 4 segments in the following order: Automated drive 1, Manual drive 1, Automated drive 2, and Manual drive 2 (**Figure 4.2**).

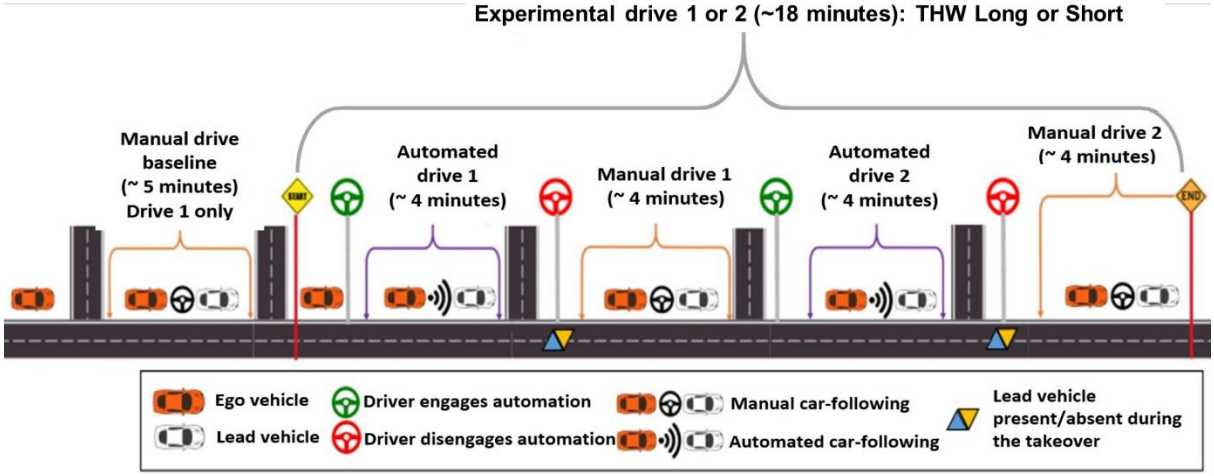

Figure 5.2. Schematic representation of the experimental drives (adapted from Radhakrishnan et al., 2022).

At the end of the manual baseline drive, the lead vehicle exited the driving lane at an intersection, following which drivers experienced free driving for ~1 minute. At the end of the free drive, participants were prompted with an auditory handover message, requesting them to engage the ADS: "Attention, engage automation". Drivers engaged the ADS using a button on the steering wheel. After about 1 minute into the automated drive, a lead vehicle joined the driving lane, starting what we termed the automated car-following (ACF) segment. During the ACF segment, participants were exposed to one of two THW conditions (0.5 s for Short THW and 1.5 s for Long THW), in a counterbalanced order across the two experimental drives.

The end of an ACF segment was followed by an auditory takeover request: “Attention, get ready to takeover”. The takeover request was issued when the ADS reached a system limitation criterion, linked to faded lane markings on the road. The takeover request was followed by a short acoustic tone (1000 Hz), lasting 0.2 s, with increasing frequency (by a factor of 1.2 after each cycle), until the driver resumed manual control. To disengage the ADS, drivers pulled the left-hand indicator stalk, or rotated the steering wheel by more than 2^{o}, or pressed the accelerator or brake pedals.

There were two types of takeover situations in this study, one with a lead vehicle present during the takeover (Lead), and the other without a lead vehicle when the takeover request was issued (No Lead). Each of the two experimental drives included the two lead vehicle conditions, with participants experiencing them after automation in both drives. The order in which participants experienced each of the two takeover situations was counterbalanced. In the Lead condition, the lead vehicle remained in the driving lane, when the takeover request was issued, and the participants took control of the vehicle. In the No Lead condition, the lead vehicle exited the driving lane at an intersection, about 5 s before the takeover request was given. About 10 s after the participants took control of the vehicle, a lead vehicle joined the driving lane, starting the manual car-following segment (MCF). Each experimental drive included an ACF, a Takeover and an MCF, repeated twice, in that order.

In this study, the Level of Automation determined whether the driver had to monitor the road during automated driving (L2), or could engage in an NDRT (L3). The Arrows task (Jamson & Merat, 2005) was used as the NDRT in this study. Participants were required to select an upward facing arrow, from a 4x4 grid of arrows, as shown in **Figure 5.3a**. The arrows were displayed on a touchscreen, placed near the centre console and the gearshift, as shown in **Figure 5.3b**. The screen also displayed participants’ cumulative score (each correct selection awarded them with one point) as well as a “score to beat”. To ensure that participants fully engaged in the Arrows Task, the participant briefing sheet offered an additional £5 reward to anyone who was able to beat the “score to beat”. However, for ethical reasons, every participant was paid this extra £5. Drivers in the L2 group were instructed to monitor the drive, when the ADS was engaged, with their hands off the steering wheel and foot off the pedals.

Figure 5.3. **(a)** A representation of the arrows task with the upward facing arrow circled in red; **(b)** A participant engaging in arrows task in L3 group during automation.

5.2.4 Self-reported workload ratings

In addition to eye tracking based metrics, we collected participants' subjective workload ratings, with each participant asked to rate their level of perceived workload, 13 times across the two drives (7 times in the first experimental drive, including once during the manual baseline drive, and 6 times in the second experimental drive). When prompted, participants rated their workload verbally, on a scale of 1-10, with 10 corresponding to the highest workload.

5.2.5 Procedure

Participants were briefed about the study upon arrival, after which they were requested to sign the consent form, and ask any questions they had about the study. After they signed the consent form, they entered the driving simulator for the practice drive. The practice drive included automated and manual driving, as well as a takeover situation. Participants were shown how to control the vehicle, engage and disengage the ADS, and provided information about the HMI. Participants in the L3 group were also given an opportunity to practise the Arrows Task. The practice drive was followed by the two experimental drives, lasting about 18 minutes each.

During the experimental drives, whenever participants were engaged in manual driving, they were advised to adhere to the speed limit of 40 mph, and asked to avoid overtaking the lead vehicle. A ~ 10-minute break followed each of the experimental drives, allowing participants to answer a set of questions about the experimental drive, including Arnett's Sensation Seeking Questionnaire (Arnett, 1994), traffic locus of control (Özkan & Lajunen, 2005) and the Driver Style Questionnaire (French et al.,

1993), see Louw et al. (2020). The scope of the questionnaire data is beyond the objectives of this paper, and as such, were not included in our analysis.

5.2.6 Data analysis

The eye tracking metrics were pre-processed using Seeing Machines' PC-DMS system, and all the data segmentation and plotting was done using MATLAB R2016a. Mean values were used to investigate drivers' overall mental workload levels as depicted by eye-based metrics, and standard deviations were used to show the variation or spread in mental workload, for each participant, across ~4-minute time windows of ACF and MCF. The time window for ACF was considered as the time from which the lead vehicle entered the driving lane after the drivers engaged automation, until the takeover request. The MCF time window started 10 s after taking over manual control of the vehicle, until ADS was re-engaged. As stated earlier, there were two ACF segments and two MCF segments, in each experimental drive (see **Figure 4.2**). A set of t-tests on drivers' mean pupil diameter, and self-reported workload ratings, did not reveal any statistically significant differences between the two ACF and the two MCF segments, within each of the experimental drives. Therefore, the data for the two similar ACF and MCF segments were aggregated to a single representation, for both eye-metrics and self-reported ratings of workload, in the L2 group.

To analyse how drivers' mental workload varied around the takeover, we segmented the eye-metrics into 3 distinct time windows: Pre-Takeover, Takeover and Post-Takeover (**Figure 4.4**). Only drivers in the L2 group were considered in this analysis, as drivers in the L3 group were engaged in the Arrows tasks, and their eyes were not captured by the eye tracker. The Pre-Takeover window was the time window from 10s before the takeover request, until the takeover request was given. Takeover was the time window from when the takeover request was given, until the driver resumed manual control of the vehicle. This Takeover time window varied for each participant. Post-Takeover was the time window from when the driver resumed manual control of the vehicle, until 10s after they resumed manual control. This time window was chosen because previous studies have shown that the peak in driving performance decrement was observed within 10 s after takeover of control (Merat et al., 2014).

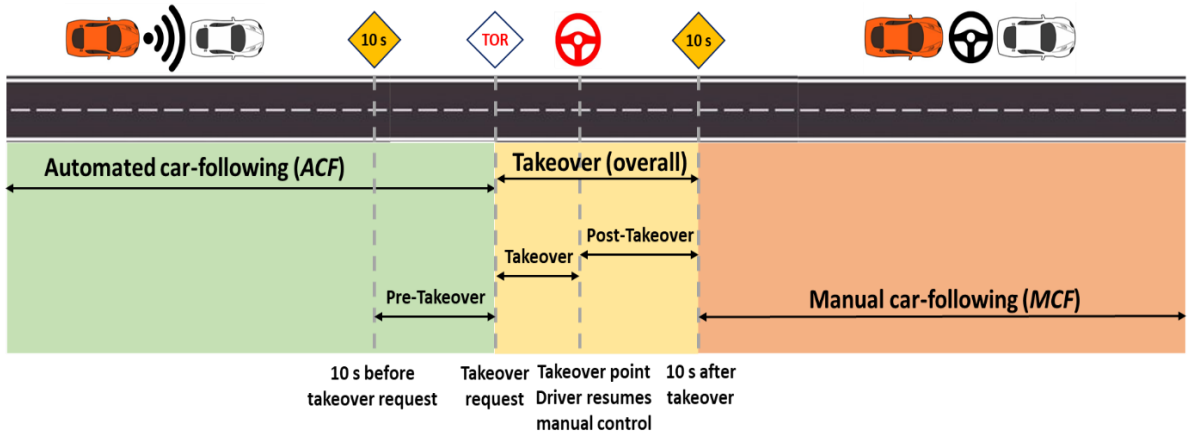

Figure 5.4. Schematic depicting different takeover windows.

To understand how drivers' mental workload levels during takeovers were affected by the presence of a lead vehicle (RQ 4), the THW maintained by this lead vehicle (RQ 3), and engagement in an NDRT during automation (RQ 5), we aggregated the eye-metrics across the Takeover and Post-Takeover windows, into a single window labelled Takeover (overall) window. This analysis was done across drivers in the L2 and L3 groups, as drivers in the L3 group redirected their eyes onto the road environment, which was within the field of view of the eye-tracker, upon hearing the takeover request. The time from which a takeover request was given, until 10 s after they resumed manual control of the vehicle was considered as the overall takeover window (**Figure 4.4**, yellow segment). Since pupil diameter has a rapid signal decay time, and is sensitive to instantaneous fluctuations in drivers' mental workload (Buettner, 2013; Mathôt, 2018), it is less likely to be influenced by previous workload inducing stimuli. Therefore, during the two No Lead takeover conditions in this study, presence of a lead vehicle in either Short or Long THW condition during automation, would not have influenced drivers' mean pupil diameter values at a later stage during the takeover. Additionally, paired-sample t-tests between the two No Lead takeover conditions did not reveal any significant differences, for either mean pupil diameter or self-reported workload ratings, across the L2 and the L3 groups. Hence, the data for the two No Lead conditions were combined to form a single representation, and labelled Infinite THW.

We computed both mean and standard deviation values of pupil diameter, during ACF and MCF segments. However, only mean values were used when

comparing performance in the takeover segments. This was because the ACF and MCF segments were comparatively longer (~ 4 minute) compared to the relatively short Takeover windows (maximum window width of ~ 15 s), and therefore, any fluctuations (high and low) in phasic pupil diameter values would have cancelled each other out, when taking only mean values. Standard deviation of pupil diameter is indicative of fluctuations in pupil diameter, and therefore, fluctuations in drivers' mental workload levels (Buettner, 2013).

5.2.7 Statistical analysis

IBM SPSS Statistics 26 was used to conduct all the statistical tests in this study. Shapiro Wilk's test revealed that the majority (> 90%) of the group-level estimates were normally distributed for each of the dependent variables used, for every ANOVA test we conducted. For statistical significance, an α-value of 0.05 was used as a limiting criterion, and partial eta-squared was computed as an effect size statistic. Mauchly's test was used to test the sphericity of data. In cases where the data was not spherical, the degrees of freedom were corrected using Greenhouse-Geisser corrections. Levene's test was used to test whether the data was homogenous. Pair-wise comparisons were used to determine differences between the different factors.

Only 50% of the data from self-reported workload ratings were normally distributed. However, the skewness and kurtosis of the non-normally distributed data were within the acceptable range +/- 2 (George & Mallery, 2010). The repeated measures ANOVA was robust enough to accommodate this violation of normality and homogeneity, with only a small effect on Type I error (Blanca et al., 2017; Pituch & Stevens, 2016).

5.3 Results

5.3.1 The effect of Drive Mode on driver workload in the L2 group (RQ 1)

To understand how drivers' mental workload during monitoring the drive in automated car-following (ACF) compared to manual car-following (MCF), we a performed one-way ANOVA, with within-participant factor of Drive Mode (ACF, MCF), on drivers' mean pupil diameter, standard deviation of pupil diameter, and self-

reported workload ratings, for the L2 group. The use of Time Headway (THW) as a factor was excluded in these analyses, because drivers controlled their own THW in the MCF condition, as opposed to the fixed values imposed during ACF. Additionally, paired sample t-tests revealed no significant differences in pupil diameter values, or self-reported workload ratings, across the Short and Long THW conditions, during ACF. Therefore, we combined the values for the Short and Long THW conditions into a single representation, for both ACF and MCF. Since the ACF and MCF segments were about 4 minutes in length, we analysed the mean and standard deviation of pupil diameter, to understand if there were any overall differences and/or variations in mental workload within the ~ 4-minute segment.

Results showed no significant main effect of Drive Mode on the mean pupil diameter values (**Table 5.1**). This may be because the time window used for analysis of data was relatively long (~4 minutes), such that the phasic fluctuations in pupil diameter might have cancelled out when using the mean values for this metric. Previous research has shown that the standard deviation of pupil diameter provides a better indication of the phasic and dynamic aspect of pupil dilation, and is therefore a better measure for highlighting fluctuations in mental workload (Buettner, 2013). As seen in **Table 5.1**, our results showed a significant main effect of Drive Mode on the standard deviation of drivers' pupil diameter, with drivers showing significantly higher standard deviation of pupil diameter during ACF, compared to MCF (**Figure 5.5**, **Figure 5.6**).

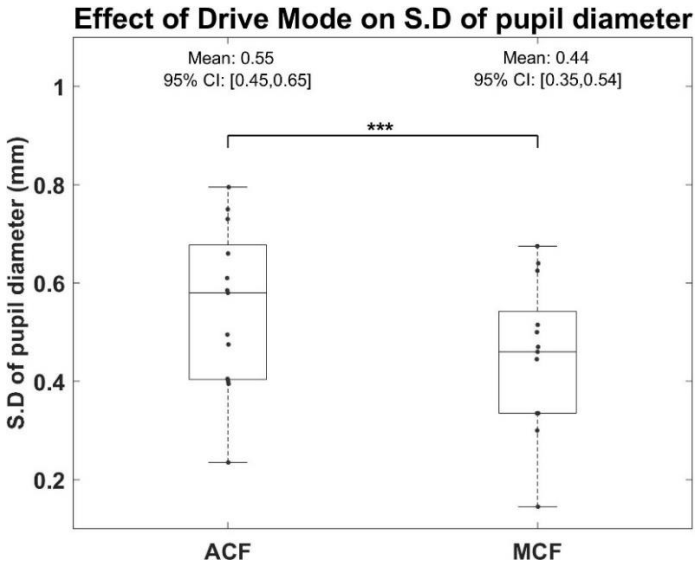

Figure 5.5. Effect of Drive Mode on standard deviation of pupil diameter, in the L2 group. ***p≤ .001

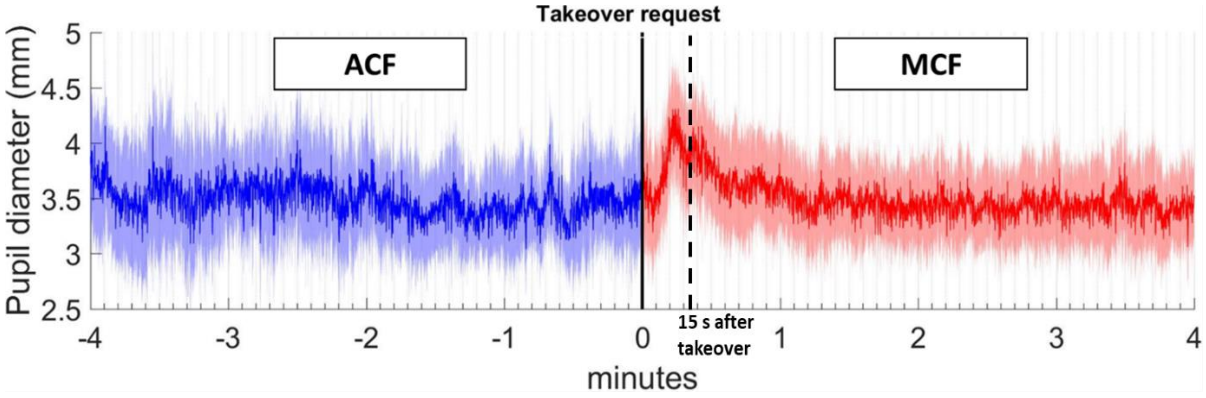

Figure 5.6. Variation of drivers' pupil diameters, during ACF and MCF, in the L2 group. The darker blue and red lines denote mean values across all drivers in the L2 group, and the lighter blue and pink regions denote the 95% confidence interval bands. Note that the MCF window used in the analysis only starts 10 s after drivers resumed manual control (~ 15 s after takeover request is issued), to filter out any variations due to the takeover.

Table 5.1. Effect of Drive Mode on drivers' pupil diameter, and self-reported workload ratings, in the L2 group.

Predictor	df1	df2	*F*	*p*	η_p^2
1. Pupil diameter (mean)	1	12	3.01	.108	.200
2. Pupil diameter (standard deviation)	1	12	59.09	<.001	.831
3. Self-reported workload ratings	1	9	.421	.533	.045

To further understand why pupil diameter values fluctuated more during ACF, compared to MCF, we visualised drivers' raw gaze data during ACF and MCF, across all participants and all drives, in the L2 group, using a 3D gaze contour plot. Results from the gaze contour plot suggest that there was a larger spread in gaze, across the driving scene, with drivers looking around more, when monitoring the drive during ACF in the L2 group (**Figure 5.7a**). However, when they were in control of the vehicle during MCF, their gaze was mostly concentrated around the centre of the driving scene, or approximately around the road centre area, suggesting they were more attentive to the driving task while being engaged in Manual driving (**Figure 5.7b**).

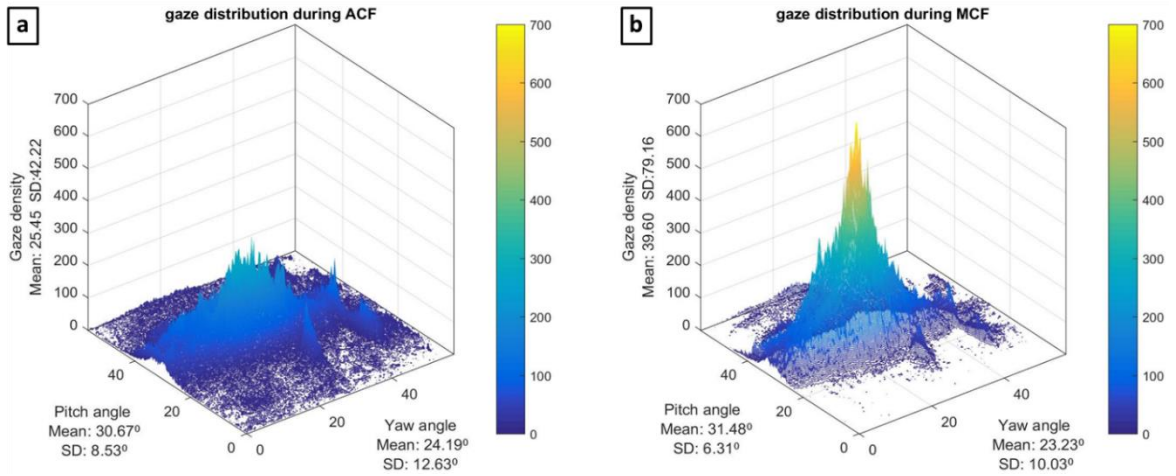

Figure 5.7. 3D gaze contour plots across **(a)** ACF and **(b)** MCF, in the L2 group. Colour-bar scale indicates number of gaze points in a particular bin, with the size of each square bin used to create the contour grid being 0.285º.

Results from drivers' self-reported workload ratings did not reveal any significant effect of Drive mode, suggesting that drivers experienced similar levels of

perceived workload, across both ACF and MCF, in the L2 group. However, there was only one workload rating for the entire duration of ACF or MCF, and hence, we could not compare the moment-to-moment changes in perceived workload, during the whole ACF or MCF period, using self-reported workload ratings.

5.3.2 Changes in workload during different stages of a takeover in the L2 group (RQ 2)

To understand how drivers' mental workload varied around the actual takeovers in the L2 group, and whether the presence of a lead vehicle affected this workload, we performed two 3x2 repeated measures ANOVAs with within-participant factors of Takeover Window (Pre-Takeover, Takeover, Post-Takeover) and Lead Vehicle (No Lead, Lead), on drivers' mean pupil diameter. A separate ANOVA was conducted for the Short and Long THW conditions, as adding THW as a factor would have been inaccurate because drivers did not experience any THW conditions, during the No Lead condition. Self-reported workload ratings were not captured separately during the three Takeover Windows considered in this analysis, and therefore, not included in the analysis.

There was a main effect of Takeover Window on drivers' mean pupil diameter in both the Short and Long THW conditions (**Table 5.2**). Given each of these takeover windows were under 10s, standard deviation of pupil diameter was not analysed here, as mean pupil diameter was sufficient to reflect drivers' workload levels during such shorter time windows. Post-hoc tests revealed that drivers' mean pupil diameter increased significantly from the Pre-Takeover time window to the Takeover time window, as well as from the Pre-Takeover time window to the Post-Takeover time window, for both the Short and Long THW conditions (**Figure 5.8a** and **Figure 8b**). Additionally, the mean pupil diameter increased significantly from the Takeover time window to the Post-Takeover time window, for the Short THW condition. Taken together, our results suggest that drivers' mental workload increased during the takeover event, increasing from the Pre-Takeover time window to the Post-Takeover time window, across both the Short and Long THW conditions. There were no other main effects or interaction effects.

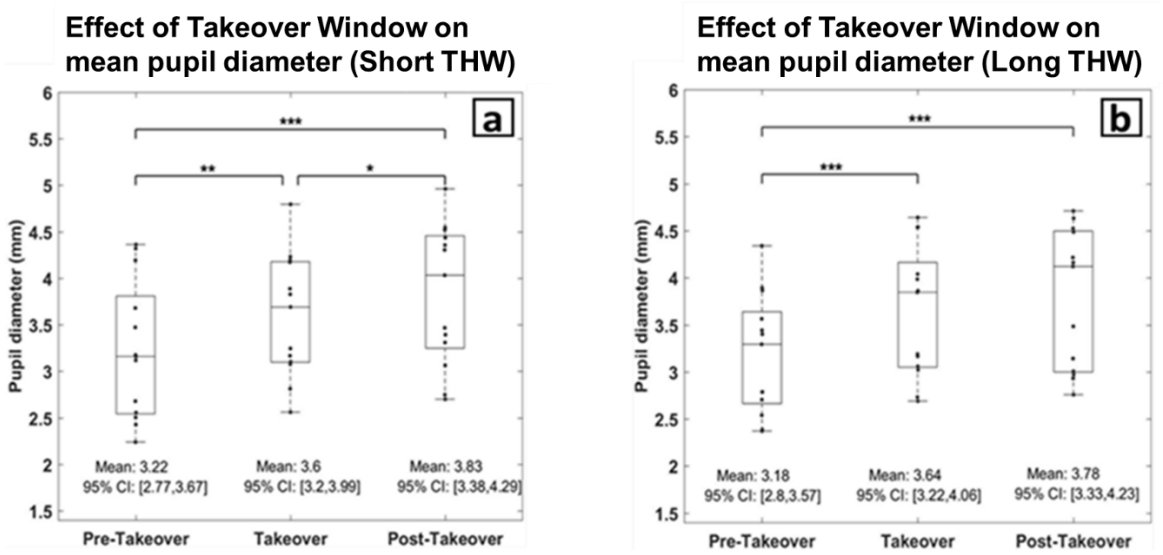

Figure 5.8. Effect of Takeover Window on pupil diameter values during **(a)** Short and **(b)** Long THW conditions. *p≤ .05, **p≤ .01, ***p≤ .001

Table 5.2. Effect of Takeover Window on mean pupil diameter values, during Short and Long THW conditions, in the L2 group.

Predictor	df1	df2	*F*	*p*	η_p^2
Effect of Takeover Window					
1. Mean pupil diameter (Short THW)	2	24	26.95	<.001	.692
2. Mean pupil diameter (Long THW)	1	12	35.78	.015	.386
Effect of Lead Vehicle					
1. Mean pupil diameter (Short THW)	1	12	.064	.805	.005
2. Mean pupil diameter (Long THW)	1	12	1.365	.265	.102

5.3.3 Effect of Time Headway, Lead Vehicle and Level of Automation on driver workload during takeovers (RQ 3, RQ4 and RQ 5)

To understand how drivers' mental workload was affected by the presence of a lead vehicle during the takeover, and its Time Headway, and whether their workload level during the takeover was affected by their prior engagement in an NDRT during automation, we performed a 3x2 mixed ANOVA on drivers' mean pupil diameter and self-reported workload ratings, using a within-participant factor of Time Headway (Short, Long, Infinite) and a between-participant factor of Level of Automation (L2, L3). As mentioned in **section 5.2.6**, the takeover window for this analysis is considered as the time from which the takeover request was given, to 10 s after they resumed manual control of the vehicle. For the Short and Long THW conditions, only the takeovers where a lead vehicle was present during the takeover, in either the Short or Long THW condition, was included. Additionally, the two No Lead conditions were consolidated to a single presentation labelled Infinite THW, as explained in **section 5.2.6**. We could not analyse the L3 drivers' eye-metrics before the takeover request was issued (such as the Pre-Takeover window seen in **section 5.3.2**) as they were engaged in the Arrows task, and, hence their eyes were occluded from the field of view of the fixed-base eye tracker used in this study.

Results showed a main effect of the Time Headway condition on drivers' mean pupil diameter values, and self-reported workload ratings (**Table 5.3**, **Figure 5.9**). Post-hoc tests revealed that drivers exhibited significantly higher mental workload levels in the Short THW condition, compared to when there was no Lead vehicle present during the takeover (Infinite THW), as revealed by both mean pupil diameter values (p <.001, **Figure 5.9a**) and self-reported workload ratings (p =.007, **Figure 5.9b**). Although statistically insignificant, drivers had higher mean pupil diameter values during the takeover in the Short THW condition, compared to the Long THW condition (p =.098). Drivers also reported higher workload levels during takeover in the Short THW condition, compared to the Long THW condition (p =.033, **Figure 5.9b**). There were no other main effects or interaction effects on drivers' mean pupil diameter values, or self-reported workload ratings.

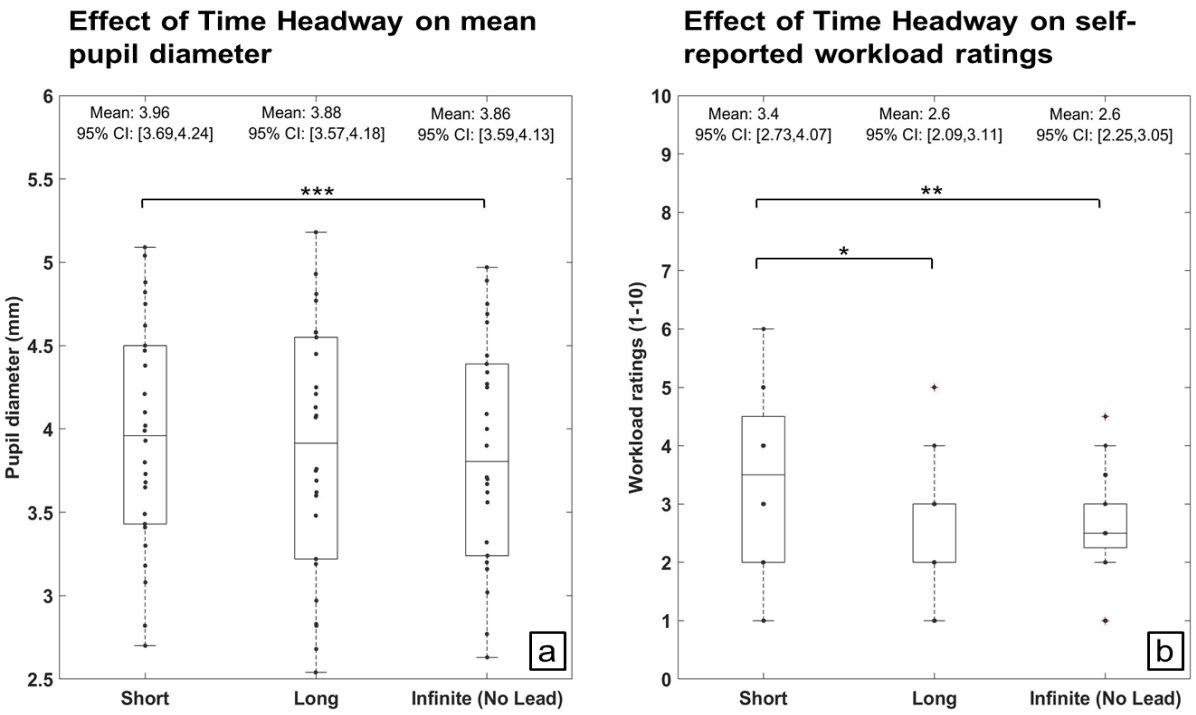

Figure 5.9. Effect of Time Headway on driver workload, as reflected in **(a)** mean pupil diameter and **(b)** self-reported workload ratings, with drivers' self-reported workload increasing from 1 to 10. *p≤ .05, **p≤ .01, ***p≤ .001

Table 5.3. Effect of Time Headway (including No Lead condition represented as Infinite THW) and Level of Automation on drivers' mean pupil diameter values and self-reported workload ratings, during takeovers.

Predictor	df1	df2	*F*	*p*	η_p^2
Effect of Time Headway					
1. Mean pupil diameter	2	48	3.29	.046	.120
2. Self-reported workload ratings	2	36	4.02	.011	.220
Effect of Level of Automation					
1. Mean pupil diameter	1	24	1.92	.179	.074
2. Self-reported workload ratings	1	18	.01	.937	.0004

5.4 Discussion and Conclusions

This study investigated changes in drivers' mental workload during a series of car-following situations in manual and automated driving, and examined how factors such as presence of a lead vehicle during the takeover, the Time Headway (THW) maintained by the lead vehicle, and prior engagement in an NDRT during automation affected drivers' mental workload. Drivers' self-reported levels of workload, and eye tracking based metrics, such as mean pupil diameter and standard deviation of pupil diameter, were compared at different stages of automated and manual driving.

Paired-sampled t-tests revealed no significant differences in mean pupil diameter values or self-reported workload ratings, due to the different THW conditions, during the automated car-following (ACF). It is likely that the perceptual difference between the two THW conditions used in this study, was not prominent enough to affect drivers' mental workload levels, especially when they were not in control of the driving task. Additionally, drivers' mean pupil diameter values, and self-reported workload ratings, indicated that they experienced similar levels of overall mental workload across the entire automated drive, (which last around 4 minutes), when the L2 drivers were simply monitoring the driving environment during ACF, compared to manual car-following (MCF). This is consistent with results from past studies which report similar levels of mental workload, as indicated by physiological data (Lohani et al., 2021; Radhakrishnan et al., 2022), or subjective ratings (Stapel et al., 2019).

However, a comparison of the standard deviation of pupil diameter values in the L2 group suggests that drivers had a higher moment to moment variation in pupil diameter, when monitoring the drive during ACF, compared to manual driving during MCF. Such fluctuations in pupil diameter could be attributed to a host of factors, such as variations in light intensity (Ellis, 1981), pupil near response, that is, the pupillary response when the pupil constricts or dilates to focus on an object (Mathôt, 2018), a startle response (Mathôt, 2018) or variations in mental workload (Buettner, 2013; Kahneman & Beatty, 1966). To understand this result further, we compared drivers' gaze patterns during the monitoring task (in ACF) and manual driving (MCF). 3D Gaze contour plots revealed that drivers focused their attention more towards the road-centre area when they were in manual control of the vehicle, as would be expected. However, in agreement with previous studies in our lab (Louw, Madigan, et al., 2017;

Louw, Markkula, et al., 2017), and that of others (Noble et al., 2021; Schartmüller et al., 2021; Yang et al., 2022), a higher spread in gaze was observed during automation, with drivers looking away from the road centre. The higher spread in gaze during automation could have been caused by drivers looking around the driving scene more and shifting and refocusing their attention between objects near and further away in their field of view, resulting in pupil near response, also known as accommodation reflex, and explains the higher phasic fluctuations seen in our study (Fincham, 1951; Mathôt, 2018). Since the ACF and MCF driving environments in our study were quite similar in terms of lighting conditions, it is unlikely that variations in pupil diameter were due to startle response or variations in light intensity. At present, it is not possible to establish whether these higher fluctuations in pupil diameter during automation were due to variations in mental workload, or the movement of gaze to different regions, or both. Our results are in agreement with findings of Wiener (1989), who showed uneven distributions of attention for airline pilots during automation, as indicated by subjective ratings, and performance of the detection response task. This uneven distribution of drivers' mental workload during automation (e.g. extreme underload or overload) can be detrimental for a safe takeover (Merat et al., 2014), and highlights the value of accurate driver monitoring systems that can assist drivers with keeping the right level of attention for supervision of the automated system, as is requested by the SAE guidelines (SAE International, 2021).

We also analysed the L2 drivers' mental workload, as indicated by pupil diameter and subjective ratings, across the three takeover windows: Pre-Takeover, Takeover and Post-Takeover. Mean pupil diameter values steadily increased from the Pre-Takeover window to the Post-Takeover window. This clear increase in pupil diameter at each of the three stages, from monitoring the drive, to hearing the takeover request, to taking over shows how drivers' vigilance and attention levels, and likely their workload, is affected as they are asked to come back into the driving loop (Merat et al., 2018), and take on the responsibility of the driving task. Similar results have been observed for skin conductance responses, when workload was found to steadily increase upon the issuance of a takeover request (Du et al., 2020). In our study, drivers' workload levels, as reflected by their pupil diameter values, peaked at around 15 s after they resumed manual control of the vehicle (see **Figure 5.6**). This is in line with driving simulator studies which show a peak in performance decrement around

15 seconds after a resumption of control from automated driving (Merat et al., 2014), after which the driver is able to stabilise the vehicle (Bueno et al., 2016).

In terms of differences in workload experienced between the two groups, engagement in the Arrows tasks did not seem to affect drivers' workload levels when they resumed control from automation, with similar pupil diameter values seen for the L2 and L3 group at the Pre-takeover stage. Since this study incorporated a non-critical takeover scenario, it is likely that drivers had adequate time to stop engaging in the Arrows task before resuming control of the vehicle, thereby eliminating the effect of any additional workload demands placed by the Arrows task on drivers in the L3 group, by the time they resumed manual control. Our results also indicated the presence of lead vehicle, especially at shorter THWs, significantly increased driver workload during the takeover, when compared to longer THWs, or takeovers without a lead vehicle present (Infinite THW). We did not observe any differences in drivers' mental workload, as reflected by mean pupil diameter values and self-reported workload ratings, between takeovers with a lead vehicle in the Long THW condition and takeovers without a lead vehicle, which suggests that the longer (1.5 s) THW conditions felt more comfortable for drivers during these non-critical takeover scenarios. We have found a similar pattern for drivers' skin conductance responses (SCR), which are known to be sensitive to stress and workload in driving (Du et al., 2020; Foy & Chapman, 2018; Mehler et al., 2009), with higher SCRs observed during takeovers that were preceded by a lead vehicle maintaining a short THW (0.5 s), compared to those with longer THWs (1.5 s) (see Radhakrishnan et al., 2022). Overall, these results suggest that drivers' physiological response and self-reported workload ratings are sensitive to the demands of a takeover after automated driving, and especially prior to more (potentially) critical scenarios.

To conclude, our findings indicate that pupil diameter can be used as an objective indicator of drivers' mental workload levels, as well as to indicate phasic variations in workload. One of the potential limitations of this study is that it was conducted in a controlled driving simulator environment, where the driving scene and its brightness levels were relatively similar throughout the drive. Further research is warranted to understand how pupil diameter is affected by different stages of automated driving in more real world settings. In this study, we were unable to use the

eye tracker to objectively capture drivers' mental workload when they were engaged in the Arrows task during automation (L3 group), since drivers were looking away from the eye tracking cameras (see **Figure 5.3b)**. Since the chances of engaging in other activities will increase with higher levels of automation, it is important that drivers' workload levels and attention are monitored during automation, because, according to guidelines from SAE (SAE International, 2021), Level 3 driving still requires drivers to resume control from automation "when required". Therefore, combining eye tracking measures with physiological sensors, can provide a more comprehensive, accurate and continuous prediction of drivers' mental workload and attention levels, even when drivers' eyes are occluded from the eye tracking sensors. Further research is required to understand the value of these metrics, and their fusion, for a wider range of takeover scenarios, to help with the creation of more reliable driver monitoring systems.

Author contribution:

Conceptualisation: V.R, N.M, T.L, G.T, R.G **Methodology:** V.R, N.M, T.L, M.L; **Software:** V.R; **Validation:** N.M, T.L, M.L; **Formal analysis:** V.R, N.M, T.L; **Investigation:** G.T, R.G, V.R; **Resources:** M.L, N.M; **Data curation:** V.R; **Writing – Original Draft:** V.R; **Writing – Review & Editing:** N.M, T.L, M.L; **Visualisation:** V.R; **Supervision:** N.M, T.L, M.L; **Project Administration:** N.M; **Funding acquisition:** N.M.

Funding: This study was supported by funding from the European Commission Horizon 2020 program under the project L3Pilot, grant agreement number 723051. The main author's PhD is co-funded by Seeing Machines Pty Ltd, Canberra, Australia.

Acknowledgement: The authors would like to thank all partners within L3Pilot and the University of Leeds Driving Simulator for their cooperation and valuable contribution. Responsibility for the information and views set out in this publication lies entirely with the authors.

5.5 References

Appel, T., Scharinger, C., Gerjets, P., & Kasneci, E. (2018). Cross-subject workload classification using pupil-related measures. *Eye Tracking Research and Applications Symposium (ETRA)*, *18*. https://doi.org/10.1145/3204493.3204531

Arnett, J. (1994). Sensation seeking: A new conceptualization and a new scale. *Personality and Individual Differences*, *16*(2), 289–296. https://doi.org/10.1016/0191-8869(94)90165-1

Batmaz, I., & Öztürk, M. (2008). Using pupil diameter changes for measuring mental workload under mental processing. *Journal of Applied Sciences*, *8*(1), 68–76. https://doi.org/10.3923/jas.2008.68.76

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276–292. https://doi.org/10.1037/0033-2909.91.2.276

Beggiato, M., Hartwich, F., & Krems, J. (2019). Physiological correlates of discomfort in automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *66*, 445–458. https://doi.org/10.1016/j.trf.2019.09.018

Blanca, M. J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Datos no normales: ¿es el ANOVA una opción válida? *Psicothema*, *29*(4), 552–557. https://doi.org/10.7334/psicothema2016.383

Boehm, U., Matzke, D., Gretton, M., Castro, S., Cooper, J., Skinner, M., Strayer, D., & Heathcote, A. (2021). Real-time prediction of short-timescale fluctuations in cognitive workload. *Cognitive Research: Principles and Implications*, *6*(1), 30. https://doi.org/10.1186/s41235-021-00289-y

Bruggen, A. (2015). An empirical investigation of the relationship between workload and performance. *Management Decision*, *53*(10), 2377–2389. https://doi.org/10.1108/MD-02-2015-0063

Bueno, M., Dogan, E., Selem, F. H., Monacelli, E., Boverie, S., & Guillaume, A. (2016). How different mental workload levels affect the take-over control after automated driving. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2040–2045. https://doi.org/10.1109/ITSC.2016.7795886

Buettner, R. (2013). Cognitive workload of humans using artificial intelligence systems: Towards objective measurement applying eye-tracking technology. In I. J. Timm & M. Thimm (Eds.), *Advances in Artificial Intelligence. KI 2013. Lecture Notes in Computer Science* (Vol. 8077, pp. 37–48). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-40942-4_4

de Waard, D. (1996). *The Measurement of Drivers' Mental Workload*. s.n.

Dogan, E., Honnêt, V., Masfrand, S., & Guillaume, A. (2019). Effects of non-driving-related tasks on takeover performance in different takeover situations in conditionally automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *62*, 494–504. https://doi.org/10.1016/j.trf.2019.02.010

Du, N., Yang, X. J., & Zhou, F. (2020). Psychophysiological responses to takeover requests in conditionally automated driving. *Accident Analysis and Prevention*, *148*. https://doi.org/10.1016/j.aap.2020.105804

Ellis, C. J. K. (1981). The pupillary light reflex in normal subjects. *British Journal of Ophthalmology*, *65*(11), 754–759. https://doi.org/10.1136/bjo.65.11.754

Fincham, E. F. (1951). The accommodation reflex and its stimulus. *British Journal of Ophthalmology*, *35*(7), 381–393. https://doi.org/10.1136/bjo.35.7.381

Foy, H. J., & Chapman, P. (2018). Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation. *Applied Ergonomics*, *73*, 90–99. https://doi.org/10.1016/j.apergo.2018.06.006

French, D. J., West, R. J., Elander, J., & Wilding, J. M. (1993). Decision-making style, driving style, and self-reported involvement in road traffic accidents. *Ergonomics*, *36*(6), 627–644. https://doi.org/10.1080/00140139308967925

George, D., & Mallery, P. (2010). SPSS for Windows step by step: A simple guide and reference,17.0 update. In *Boston: Pearson Education, Inc* (10th ed.). Allyn & Bacon.

Gold, C., Berisha, I., & Bengler, K. (2015). Utilization of drivetime - Performing non-driving related tasks while driving highly automated. *Proceedings of the Human Factors and Ergonomics Society*. https://doi.org/10.1177/1541931215591360

Gold, C., Körber, M., Lechner, D., & Bengler, K. (2016). Taking over Control from Highly Automated Vehicles in Complex Traffic Situations. *Human Factors*, *58*(4), 642–652. https://doi.org/10.1177/0018720816634226

Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems*, *6*(2), 156–166. https://doi.org/10.1109/TITS.2005.848368

Heikoop, D. D., de Winter, J. C. F., van Arem, B., & Stanton, N. A. (2019). Acclimatizing to automation: Driver workload and stress during partially automated car following in real traffic. *Transportation Research Part F: Traffic Psychology and Behaviour*, *65*, 503–517. https://doi.org/10.1016/j.trf.2019.07.024

Jamson, A. H., & Merat, N. (2005). Surrogate in-vehicle information systems and driver behaviour: Effects of visual and cognitive load in simulated rural driving. *Transportation Research Part F: Traffic Psychology and Behaviour*. https://doi.org/10.1016/j.trf.2005.04.002

Kahneman, D. (1973). Attention and Effort. In *Prentice-Hall series in experimental psychology*. Prentice-Hall. https://doi.org/10.2307/1421603

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(3756), 1583–1585. https://doi.org/10.1126/science.154.3756.1583

Liu, K., Green, P., & Liu, Y. (2019). Traffic and Ratings of Driver Workload: The Effect of the Number of Vehicles and Their Distance Headways. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *63*(1), 2134–2138. https://doi.org/10.1177/1071181319631051

Lohani, M., Cooper, J. M., Erickson, G. G., Simmons, T. G., McDonnell, A. S., Carriero, A. E., Crabtree, K. W., & Strayer, D. L. (2020). Driver Arousal and Workload Under Partial Vehicle Automation: A Pilot Study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *64*(1), 1955–1959. https://doi.org/10.1177/1071181320641471

Lohani, M., Cooper, J. M., Erickson, G. G., Simmons, T. G., McDonnell, A. S., Carriero, A. E., Crabtree, K. W., & Strayer, D. L. (2021). No Difference in Arousal or Cognitive Demands Between Manual and Partially Automated Driving: A Multi-

Method On-Road Study. *Frontiers in Neuroscience*, *15*. https://doi.org/10.3389/fnins.2021.577418

Louw, T., Goncalves, R., Torrao, G., Radhakrishnan, V., Lyu, W., Puente Guillen, P., & Merat, N. (2020). Do drivers change their manual car-following behaviour after automated car-following? *Cognition, Technology and Work*. https://doi.org/10.1007/s10111-020-00658-5

Louw, T., Madigan, R., Carsten, O., & Merat, N. (2017). Were they in the loop during automated driving? Links between visual attention and crash potential. *Injury Prevention*, *23*, 281–286. https://doi.org/10.1136/injuryprev-2016-042155

Louw, T., Markkula, G., Boer, E., Madigan, R., Carsten, O., & Merat, N. (2017). Coming back into the loop: Drivers' perceptual-motor performance in critical events after automated driving. *Accident Analysis and Prevention*, *108*, 9–18. https://doi.org/10.1016/j.aap.2017.08.011

Marquart, G., Cabrall, C., & de Winter, J. (2015). Review of Eye-related Measures of Drivers' Mental Workload. *Procedia Manufacturing*, *3*, 2854–2861. https://doi.org/10.1016/j.promfg.2015.07.783

Mathôt, S. (2018). Pupillometry: Psychology, Physiology, and Function. *Journal of Cognition*, *1*(1), 1–23. https://doi.org/10.5334/joc.18

Merat, N., Jamson, A. H., Lai, F. C. H., & Carsten, O. (2012). Highly automated driving, secondary task performance, and driver state. *Human Factors*, *54*(5), 762–771. https://doi.org/10.1177/0018720812442087

Merat, N., Jamson, A. H., Lai, F. C. H., Daly, M., & Carsten, O. M. J. (2014). Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*(PB), 274–282. https://doi.org/10.1016/j.trf.2014.09.005

Merat, N., Seppelt, B., Louw, T., Engström, J., Lee, J. D., Johansson, E., Green, C. A., Katazaki, S., Monk, C., Itoh, M., McGehee, D., Sunda, T., Unoura, K., Victor, T., Schieben, A., & Keinath, A. (2018). The "Out-of-the-Loop" concept in automated driving: proposed definition, measures and implications. *Cognition, Technology & Work*, 1–12. https://doi.org/10.1007/s10111-018-0525-8

Mioch, T., Kroon, L., & Neerincx, M. A. (2017). Driver Readiness Model for Regulating the Transfer from Automation to Human Control. *Proceedings of the 22nd International Conference on Intelligent User Interfaces - IUI '17*, *March*, 205–213. https://doi.org/10.1145/3025171.3025199

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134–140. https://doi.org/10.1016/S1364-6613(03)00028-7

National Highway Traffic Safety Administration. (2009). *Traffic Safety Facts 2009: A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System*. U.S. Department of Transportation. http://www.nhtsa.gov/NCSA.

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

Özkan, T., & Lajunen, T. (2005). Multidimensional Traffic Locus of Control Scale (T-LOC): Factor structure and relationship to risky driving. *Personality and Individual Differences*, *38*(3), 533–545. https://doi.org/10.1016/j.paid.2004.05.007

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs. *Journal of Cognitive Engineering and Decision Making*, *2*(2), 140–160. https://doi.org/10.1518/155534308X284417

Pituch, K. A., & Stevens, J. P. (2016). Applied Multivariate Statistics for the Social Sciences. In *Applied Multivariate Statistics for the Social Sciences* (Sixth). Routledge, Taylor and Francis group. www.routledge.com/9780415836661

Radhakrishnan, V., Merat, N., Louw, T., Gonçalves, R. C., Torrao, G., Lyu, W., Puente Guillen, P., & Lenné, M. G. (2022). Physiological indicators of driver workload during car-following scenarios and takeovers in highly automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *87*, 149–163. https://doi.org/10.1016/J.TRF.2022.04.002

Radhakrishnan, V., Merat, N., Louw, T., Lenné, M. G., Romano, R., Paschalidis, E., Hajiseyedjavadi, F., Wei, C., & Boer, E. R. (2020). Measuring drivers'

physiological response to different vehicle controllers in highly automated driving (HAD): Opportunities for establishing real-time values of driver discomfort. *Information (Switzerland)*, *11*(8), 390. https://doi.org/10.3390/INFO11080390

Recarte, M. Á., Ṕerez, E., Conchillo, Á., & Nunes, L. M. (2008). Mental workload and visual impairment: Differences between pupil, blink, and subjective rating. *Spanish Journal of Psychology*, *11*(2), 374–385. https://doi.org/10.1017/s1138741600004406

Reimer, B., & Mehler, B. (2011). The impact of cognitive workload on physiological arousal in young adult drivers: A field study and simulation validation. *Ergonomics*. https://doi.org/10.1080/00140139.2011.604431

SAE International. (2021). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. https://www.sae.org/standards/content/j3016_202104/

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). AUTOMATION SURPRISES. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed., pp. 1926–1943). Wiley.

Schartmüller, C., Weigl, K., Löcken, A., Wintersberger, P., Steinhauser, M., & Riener, A. (2021). Displays for productive non-driving related tasks: Visual behavior and its impact in conditionally automated driving. *Multimodal Technologies and Interaction*, *5*(4), 21. https://doi.org/10.3390/MTI5040021

Siebert, F. W., & Wallis, F. L. (2019). How speed and visibility influence preferred headway distances in highly automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *64*, 485–494. https://doi.org/10.1016/j.trf.2019.06.009

Spector, R. H. (1990). The pupils. In H. K. Walker, W. D. Hall, & J. W. Hurst (Eds.), *Clinical Methods: The History, Physical, and Laboratory Examinations* (3rd ed.). Butterworths. https://www.ncbi.nlm.nih.gov/books/NBK381/

Stapel, J., Mullakkal-Babu, F. A., & Happee, R. (2019). Automated driving reduces perceived workload, but monitoring causes higher cognitive load than manual driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *60*, 590–605. https://doi.org/10.1016/j.trf.2018.11.006

Steinhauer, S. R., Siegle, G. J., Condray, R., & Pless, M. (2004). Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *International Journal of Psychophysiology*, *52*(1), 77–86. https://doi.org/10.1016/J.IJPSYCHO.2003.12.005

Stern, J. A. (1997). *The pupil of the eye: what can it tell us about 'mental processes'?* Human Engineering for Quality Life (HQL) Quarterly 8.

Thorn, E., Kimmel, S., & Chaka, M. (2018). A Framework for Automated Driving System Testable Cases and Scenarios (Report No DOT HS 812 623). In *National Highway Traffic Safety Administration*. U.S. Department of Transportation. www.ntis.gov.

Tsai, Y. F., Viirre, E., Strychacz, C., Chase, B., & Jung, T. P. (2007). Task performance and eye activity: Predicting behavior relating to cognitive workload. *Aviation Space and Environmental Medicine*, *78*(5 II).

Veltman, J. A., & Gaillard, A. W. K. (1996). Physiological indices of workload in a simulated flight task. *Biological Psychology*, *42*(3), 323–342. https://doi.org/10.1016/0301-0511(95)05165-1

Vogel, K. (2003). A comparison of headway and time to collision as safety indicators. *Accident Analysis and Prevention*, *35*(3), 427–433. https://doi.org/10.1016/S0001-4575(02)00022-2

Wickens, C. D. (1984). Processing Resources in Attention. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 63–102). Academic Press. https://doi.org/R. Parasuraman, R. Davies

Wickens, C. D., Gutzwiller, R. S., & Santamaria, A. (2015). Discrete task switching in overload: A meta-analyses and a model. *International Journal of Human Computer Studies*, *79*, 79–84. https://doi.org/10.1016/j.ijhcs.2015.01.002

Wiener, E. L. (1989). *Human Factors of Advanced Technology ("Glass Cockpit") Transport Aircraft*. NASA Contractor Report 177528. https://human-factors.arc.nasa.gov/publications/HF_AdvTech_Aircraft.pdf

Yang, S., Wilson, K., Roady, T., Kuo, J., & Lenné, M. G. (2022). Beyond gaze fixation: Modeling peripheral vision in relation to speed, Tesla Autopilot, cognitive load,

and age in highway driving. *Accident Analysis & Prevention*, *171*, 106670. https://doi.org/10.1016/j.aap.2022.106670

Young, M. S., & Stanton, N. A. (2002). Malleable Attentional Resources Theory: A New Explanation for the Effects of Mental Underload on Performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *44*(3), 365–375. https://doi.org/10.1518/0018720024497709

Yu, D., Park, C., Choi, H., Kim, D., & Hwang, S. H. (2021). Takeover safety analysis with driver monitoring systems and driver-vehicle interfaces in highly automated vehicles. *Applied Sciences (Switzerland)*, *11*(15), 6685. https://doi.org/10.3390/app11156685

Zeeb, K., Buchner, A., & Schrauf, M. (2016). Is take-over time all that matters? the impact of visual-cognitive load on driver take-over quality after conditionally automated driving. *Accident Analysis and Prevention*, *92*, 230–239. https://doi.org/10.1016/j.aap.2016.04.002

6 FINAL DISCUSSION AND CONCLUSION

This research was conducted on an EPSRC funded studentship, with iCASE partner Seeing Machines Inc, Canberra, Australia. The overall objective of this project was to evaluate the feasibility of using different physiological sensors, such as EDA, ECG and eye tracking, for understanding and measuring driver states during vehicle automation.

The work carried out in this research involved studies conducted as part of other larger projects, such as the Innovate UK-funded HumanDrive project, and the Horizon2020-funded L3Pilot project. The main aim of the HumanDrive project was to develop human-like autonomous driving systems that can drive 200+ miles across the UK, in live traffic and in natural conditions, using artificial intelligence and machine learning. As part of this project, the study reported in Chapter 3, investigated factors affecting driving comfort, across a wide range of conditions, in automated driving. The EU-funded L3Pilot project focused on testing the viability of SAE L3 automated driving, as a safe and efficient means of transportation on public roads. As part of this project, and in partnership with one of the consortium partners, Toyota Motor Europe, the study reported in Chapters 4 and 5 investigated how driver workload is affected by a host of factors, during different stages of automated driving, especially around car-following scenarios.

The main aim of this thesis was to understand and measure drivers' psychological states such as discomfort and workload, using physiological signals, to better understand how automation affects these states, at different stages of the drive. For example, I investigated how different road environments and automated vehicle driving styles can affect driver discomfort. Experiencing high levels of discomfort during automation can negatively impact drivers' trust and wider acceptance of automation features. To measure the driver states of comfort and workload objectively, and in real-time, physiological metrics derived from electrocardiogram (ECG), electrodermal activity (EDA), and eye tracking, were used. ECG, EDA and eye tracking signals were selected due to the minimally-intrusive nature of these metrics, and prior work suggesting their effective usage as indicators of comfort and workload in manual driving.

This thesis focused on understanding and evaluating the effects of automation on the human driver, and not on the technology underlying such automated systems. Neither was the focus on validating human-machine interfaces and their effects on driver state. The design of the experiments were, on the one hand, guided by answering some of the practical questions in the literature regarding how certain situations and scenarios affect driver state, and on the other hand, measuring such driver states objectively, in a controlled simulator environment. Given the nature of physiological signals, and its many-to-one relationship with psychological states (for example, an increase in EDA can be indicative of stress, arousal, high workload, or, discomfort), the experiments were controlled for other confounding factors such as caffeine intake or physical exercise, and as such, may not reflect precise real-world use cases. However, the results obtained in our studies have allowed us to draw practical conclusions regarding how different physiological metrics, especially when combined, can be used to objectively identify, and distinguish between, different driver states, in real-world conditions. This can help future automated vehicles to better understand the capabilities and limitations of the driver, and assist by providing a more tailored intervention strategy, based on the individual and their current psychological state.

6.1 Summary of main findings from the experiments

As discussed at the end of Chapter 1, four specific research questions were formulated as guiding questions for this thesis. Chapters 2, 3, 4 and 5 of this thesis, were structured to answer these research questions, the findings of which, have been summarised below.

1. **Can motion artefacts be removed from EDA signals, in a dynamic driving environment, to assess drivers' comfort and workload levels in real-world driving?**

The literature on physiological signals of ECG and EDA, suggest that these signals are highly susceptible to motion artefacts (Boucsein, 2012; Kikhia et al., 2016; van Gent et al., 2018). The majority of studies on ECG and EDA signals have been

conducted in stationary laboratory environments, including studies on driving, many of which have been conducted in fixed-base driving simulators (Beggiato et al., 2019; Foy & Chapman, 2018; Mehler et al., 2009; Müller et al., 2021). This gap acted as a starting point for my research, with a preliminary investigation focused on whether such motion artefacts can be successfully removed from ECG and EDA signals, in driving environments, to make useful interpretations in real-world settings. As mentioned in the introduction, ECG signals, especially R-R peaks, were less susceptible to motion-artefacts in driving environments, when the signal was captured at sampling rates greater than 500 Hz. Therefore, I focused on removing motion artefacts from EDA signals.

Chapter 2 of this thesis is a methodology paper, that introduces a novel and efficient method for removing motion artefacts from EDA signals, with the potential to be used in real-time. The filter designed is essentially a shape-based filter, which detects whether the signal exceeds certain threshold values, based on recommendations in Boucsein (2012) and Kikhia et al. (2016). Additionally, a slope-based boundary constraint was applied, derived from the recommendations in Boucsein (2012) and Kikhia et al. (2016). A case study for this method was conducted using a dataset from a simulator study conducted in the full-motion based University of Leeds Driving Simulator (UoLDS).

From visual inspection, it was observed that my proposed algorithm successfully removed motion artefacts from the EDA signal, while closely retaining the shape of skin conductance response (SCR) events in the EDA signal. An algorithm based on this method has been used to remove motion artefacts from EDA signals for the studies mentioned in Chapters 3, 4 and 5. This code will be made available for download as a free MATLAB-based function on GitHub soon, and is also provided in the appendix section of this thesis.

2. What are the primary factors contributing to driver discomfort in HAD, and are these reflected in drivers' physiological state?

Chapter 3 of this thesis focused on understanding factors affecting driver discomfort, and whether the changes in drivers' discomfort levels during different types

of automated driving styles can be measured using physiological metrics such as RMSSD, mean HR and number of SCRs per minute (nSCR/min). I measured driver discomfort (i.e. the lack of comfort), since signs of discomfort tend to be more well-defined and pronounced, compared to an unaroused/relaxed state of comfort (Siebert et al., 2013).

The study in Chapter 3 focused on understanding how different automated driving styles, including human-like driving controllers, and a replay of drivers' own manual driving style, affected driver comfort, while negotiating varying road geometries and obstacles in urban and rural environments. Physiological metrics of RMSSD, mean HR and nSCR/min, captured using the BIOPAC MP35 data acquisition system, were used as objective indicators of driver discomfort, along with subjective discomfort ratings.

Results from the study indicated that there were no significant differences in driver discomfort, as measured by physiological metrics and subjective ratings, across the different automated driving controllers. A possible explanation for this finding could be the similar resultant acceleration profiles across the different automated controllers for the whole drive, where it only exceeded the comfort threshold for acceleration (2 m/s^2; Eriksson & Svensson, 2015) in the rural (and not urban) sections. A similar result was observed in the study by Beggiato et al. (2019), where the authors did not observe any significant changes in driver discomfort across three different automated driving styles, when considering the whole drive, which was 9.2 km in length and consisted of intersections, exit ramps, obstacles, rural road and construction zones.

Results from the study in Chapter 3 also showed significant differences in terms of physiological activity across all the metrics, between manual driving and at least one of the automated driving styles. However, it could not be determined if this was due to increased physical and mental demand required for manual driving, the higher discomfort experienced during manual driving, or both. I also observed higher discomfort, for all controllers, during the higher speed and narrower sections of the rural environment, as opposed to the urban environment, as indicated by both subjective ratings, and nSCR/min values. Similar observations were seen by Mourant & Thattacherry (2000) in their study on manual driving, who suggested that higher speeds and narrower roads can induce discomfort for the driver.

One crucial finding of this study was the importance of acceleration forces and how these affect driver discomfort. Based on our findings, it is likely that when the resultant acceleration and jerk forces experienced by the driver is well below the comfort thresholds, other factors that affect comfort, such as familiarity of the drive, or presence of obstacles, become more prominent and noticeable to drivers and affects their comfort rating. In contrast, when resultant acceleration and jerk forces are well above drivers' comfort threshold, it seemingly overshadows other determinants of driver discomfort, such as familiarity of the drive or presence of obstacles. Results also showed that the EDA-based nSCR/min metric was more sensitive to continuous changes in discomfort-inducing stimuli, compared to ECG-based metrics. This could be because ECG-based measures of RMSSD, and mean HR require larger time windows (minimum of two minutes) for accurate analysis.

In the study reported in Chapter 3, factors affecting driver discomfort, during automation, were investigated when drivers were simply monitoring the drive. However, it should be noted that discomfort shares similarities, and overlaps with other related driver states, such as stress and mental workload (Beggiato et al., 2019). For example, driver discomfort induced when the vehicle does not keep adequate safety margins from another vehicle, can also result in an increase in driver workload, and drivers are likely to become more vigilant and pay more attention to the actions of the automated vehicle and the driving environment, when their safety thresholds are breached, thereby increasing their workload levels.

High workload (overload) is also know to affect both the performance and safety of the driver, for example, if and when they have to resume manual control of the vehicle, after a period of automated driving. To investigate how drivers' workload levels are affected by different sections of an automated driving scenario, and whether this can be measured using physiological metrics, the next study focussed on measuring eye tracking and physiological metrics in an automated driving study, involving a range of take over scenario, as detailed below.

3. **Are changes in workload levels during different stages of HAD, including the transition of control, reflected in drivers' electrodermal activity, and electrocardiogram-based physiological metrics?**

Chapter 4 of this thesis focused on understanding how different manipulations during an automated car following scenario affected driver workload. Additionally, the use of physiological metrics as objective indicators of driver workload was investigated. Accurate detection of drivers' workload levels can help future driver state monitoring systems to be better informed about drivers' capabilities and limitations, especially when they are required to take over from automation. Driver workload was manipulated in a set of car-following situations by varying the time headway from a lead vehicle, at different stages of automated driving, and during takeovers. Car-following situations were of particular interest, as they are known to be pre-cursors to rear end collisions, which account for over 31.5% of all collisions in the US (National Highway Traffic Safety Administration, 2009).

In Chapter 4, I focused on whether physiological metrics of RMSSD, mean HR, ECG-derived respiration rate (EDR) and nSCR/min, can be used as objective indicators of drivers' workload levels, along with self-reported ratings of workload. The physical and cognitive demands imposed by different tasks, such as monitoring the drive, engaging in an NDRT, or manual driving, could be distinguished using physiological signals. I have hypothesised and ranked the physical, cognitive, and overall (physical plus cognitive) workload levels of the different conditions in our study, with manual driving being the task with the highest physical workload, and the Arrows task having the highest cognitive and overall workload (see **Figure 6.1**).

The trends observed in drivers' RMSSD, mean HR, and nSCR/min values suggest that engagement in the Arrows task during SAE L3 drive resulted in higher workload levels than monitoring during SAE L2 drive, or manual driving. The trends in drivers' RMSSD, mean HR and nSCR/min also suggest that monitoring the drive and manual driving had similar levels of workload. This trend is similar to that observed for congitive task demands of the three tasks, as seen in **Figure 6.1**, suggesting RMMSD, mean HR and nSCR/min are more senstive to cognitive demands of a task, within driving context. However, the trend seen in the EDR metric was similar to that of overall workload imposed by monitoring the drive, manual driving and the arrows task, Drivers' had significant higher workload levels as indicated by the EDR metric, during manual driving, compared to monitoring during SAE L2 drive. This could be due to the

higher senstivity of breathing rate to even minor physical load/activity, such as those associated with hand and leg movements in manual driving (Hidalgo-Muñoz et al., 2019). Their workload levels, as indicated by the EDR metric, when engaging in the Arrows task during SAE L3 drive, was significantly higher than workload levels during both monitoring in SAE L2, and manual driving. Therefore, by combining EDR metric, with RMSSD, mean HR, and nSCR/min metric, one can distinguish tasks which are purely cognitive in nature, with task that might have physical workload, within driving context. I believe this to be a novel finding, which is of great importance to psychophysiological research, within the driving context. This ablity to distinguish between the physical and cognitive demands posed by different tasks, using such physiological metrics, can vastly improve workload detection, and drivers' performance predictions in future driver state monitoring systems. While I was able to see differences between the different conditions seen in our study, it is to be seen whether these metrics are able to distinguish more subtle changes in physical or cognitive task demands, in a wider range of physical and cognitive loads, within driving context. Therefore, further research is warranted to understanding the sensitivity of different physiological metrics, to physical and cognitive demands of different tasks, encountered in the driving context.

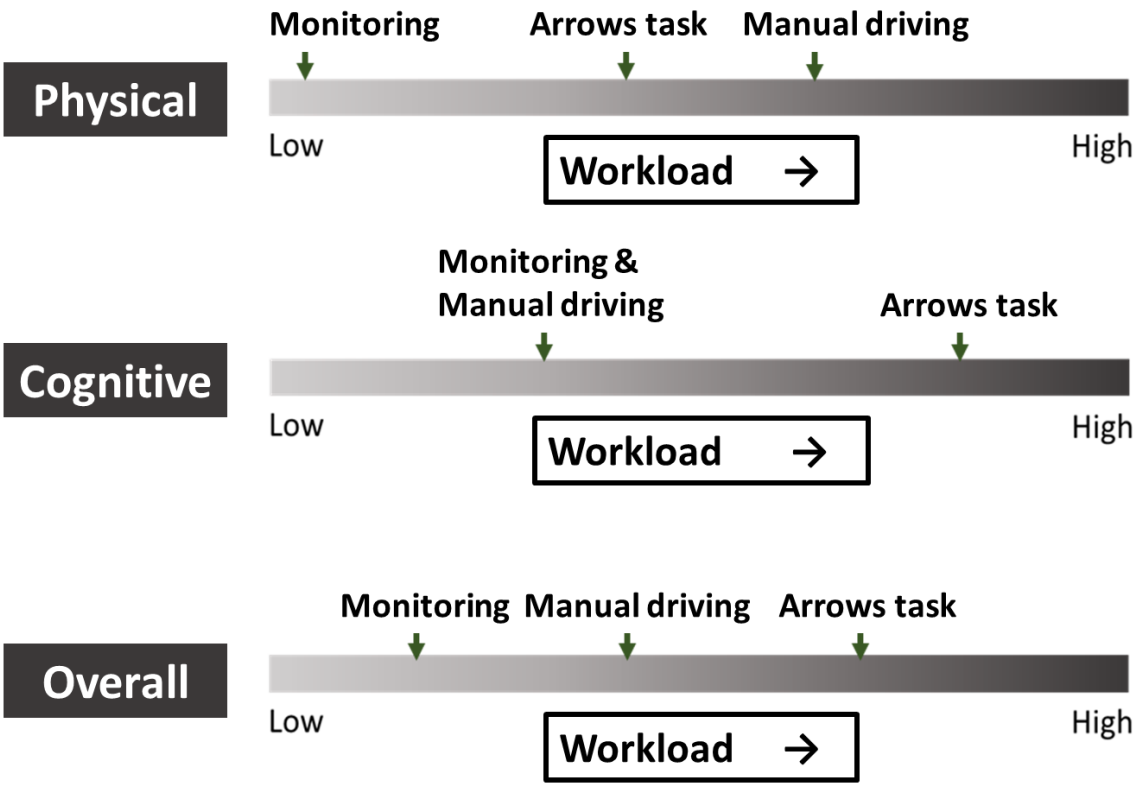

Figure 6.1. Schematic representation of the hypothesised relationship between physical, cognitive and overall (physical + cognitive) workload of different tasks during driving – based on the studies reported in Chapters 4 and 5.

Results from nSCR/min and subjective workload ratings indicated that the different Time Headways (THW) maintained by a lead vehicle, did not affect drivers' workload levels during automation, when they were a passive observer of the driving task. Only the nSCR/min metric was used as an indicator of driver workload around takeovers, as ECG-based metrics require longer time windows (over at least 2 minutes), for accurate analysis (Laborde et al., 2017).However, the presence of a lead vehicle, especially one maintaining shorter THWs, seemed to increase driver workload as indicated by nSCR/min and subjective workload ratings, when a resumption of control by drivers was required. It is likely that drivers were less bothered by the safety margins in terms of time headways, maintained by the automated driving controller, as long as they were not required to control the driving task. Further research into the safe and acceptable time headways maintained by an AV controller is therefore

warranted, to ensure that a safe transition of control is achieved, without too much additional attentional and workload demands on the driver.

Drivers' engagement in an activity during automation (either completing the arrows task, or monitoring the drive), did not seem to affect driver workload during the takeovers, as indicated by their nSCR/min data and subjective workload ratings. This is in contrast to the findings of Zeeb et al. (2016), and Wandtner et al. (2018), where the authors observed significant differences in workload, as indicated by subjective ratings, and takeover performance, as indicated by response times, minimum time to collision and deviation from the lane centre, due to prior engagement in NDRTs. There are several possible explanations for this. It could be due to the between-participant nature of the prior activity, in the current study. Also, since this study involved a non-critical takeover scenario, drivers had adequate time to stop engaging in the Arrows task, before resuming control, thereby nullifying any additional workload that was imposed by the Arrows task. Finally, perhaps the takeover task was quite demanding in itself, masking the effects of any activities that took place before the takeover request.

4. How is drivers' attentional demand, and workload levels, affected at different stages of HAD, as reflected by pupil-diameter values?

Chapter 5 of this thesis focused specifically on how attentional demand (in terms of how they process the visual driving scene) at different stages of automation affected driver workload, as indicated by eye tracking based metrics. Pupil diameter was used as an indicator of attentional demand and cognitive load, and 3-D gaze contour plots were used to understand where drivers allocated their visual attention, during different stages of automation and manual driving. In agreement with results from other physiological metrics, outlined in Chapter 4, mean pupil diameter values were shown to be similar when monitoring the task during automation, as that seen in manual driving. The similar pupil diameter values suggests that drivers had similar overall levels of cognitive workload, when monitoring the driving task in automation, as that seen during manual driving. However, a different result was found for the standard deviation of pupil diameter, which showed a fluctuation of pupil diameter

when drivers were engaged in monitoring the drive during automation, compared to when they were in control of the vehicle in manual driving. Since the brightness levels were similar during automation and manual driving, fluctuation in pupil diameter values could be narrowed down to either accommodation reflex (shifting of focus between near and far objects) or variations in workload. To further investigate the causal factor of this pupil diameter fluctuation, I compared where drivers' visual attention, during automation and during manual driving. For example, if the drivers' shifted their attention across different objects in the driving scene, this could indicate accommodation reflex. Although, it does not rule out variations in cognitive load, and processing different elements across the driving scene could indeed result in varying levels of cognitive load. 3D gaze contour plots revealed that drivers looked around the driving scene more, with more glances away from the road centre, during automation, compared to manual driving, results which are in agreement with previous research in this context (Louw, Markkula, et al., 2017; Schartmüller et al., 2021; Yang et al., 2022). However, it was not possible to conclude whether the larger fluctuations in pupil diameter values during automation were a result of variations in attentional demand and cognitive workload, or due to an accommodation reflex (Mathôt, 2018), or both. Research from the aviation industry has also indicated that airline pilots experience uneven distribution of attention and workload, as indicated by subjective ratings and performance in detection response tasks, during automation (N B Sarter et al., 1997; Wiener, 1989).

I also found that pupil diameter values increased during takeovers, when the takeover happened in presence of a lead vehicle maintaining shorter THWs, indicating an increase in drivers' workload levels. This is in agreement with the findings in Chapter 4. Additionally, how drivers' workload, based on their pupil diameter, varied around the actual takeover, from the pre-takeover section, to the takeover itself, and the period after takeover, was also analysed. As predicted, there was a steady increase in pupil diameter values from the pre-takeover window to the takeover window, increasing further after the takeover. This suggests that the takeover process is in itself a complex and mentally demanding task, which can be identified by changes in pupil diameter. A peak in the pupil diameter was observed around 15 seconds after the drivers resumed manual control of the vehicle, suggesting their workload was highest around this time. This is in line with previous research, which shows a peak

performance decrement around 15 seconds after a resumption of control from automated driving, as indicated by metrics such as maximum steering wheel angles, speed and deviation from lane centre (Bueno et al., 2016; Merat et al., 2014).

Overall, I was able to establish the validity of physiological metrics, for measuring drivers' discomfort and workload levels objectively, in a number of driving simulator studies involving manual and automated driving. In section 6.3, I discuss how future driver state monitoring systems can build on the findings from this research, to provide real-time and accurate detection of drivers' comfort and workload levels. However, there are some merits, as well as limitations in using such metrics in real-world settings, as opposed to more controlled environments, which will be discussed in the next section.

6.2 Reflections on methodology and measures

6.2.1 Methodology used and limitations

The EDA signal was collected in a high-fidelity, motion-based driving simulator, which is comparable to motion experienced in real-world driving. Our algorithm for motion artefact removal is actually adapted from well-cited research, which suggests that EDA signal does not increase by more than 20% or decrease by less than 10%, in a 1 second time window (Boucsein, 2012; Kikhia et al., 2016). Additionally, a slope constraint was derived from this boundary condition, to ensure there were no unwanted sharp spikes in the signal, especially when the EDA signal is collected at high sampling rates. Given the nature of the shape of an EDA signal, as described by Boucsein (2012) and Braithwaite et al. (2015), the Modified Akima cubic Hermite interpolation, also referred to as Makima interpolation technique, was found to be the best technique for fill the missing/noisy data points, and preserving the shape of the signal. Future studies could further validate this method by collecting and comparing EDA data in stationary as well as full-motion environments.

In Chapter 3, I wanted to investigate how driver discomfort is affected by the presence of obstacles, such as parked-cars and roadworks, that were incorporated into the road-geometry of the study. It is likely that the key differences would have been around how the different automated controllers negotiated these obstacles. However, I was unable to analyse the physiological data around the short time

windows where the controllers were negotiating these obstacles, due to a synchronisation error between the BIOPAC MP35 system and the simulator data. This failed synchronisation was likely to cause a sync-time difference of up to 5 seconds, which was unacceptable, given the obstacles were negotiated in time windows lasting less than 5 seconds.

Studies in Chapter 4 and Chapter 5 incorporated a complex mixed-design, to test the various workload-inducing conditions on driver physiological metrics and subjective response. The synchronisation issue for BIOPAC signals was rectified prior to this study.

Due to time limitations, and to avoid a long drive for each participant, a between participant design was used for this study, which meant that half the drivers engaged in the Arrows task and half were asked to monitor the driving environment during automation. Given that both physiological indices, and self-reported workload ratings can be highly subjective in nature, interpretation of the results need to consider this caveat. Additionally, data from 6 participants were discarded, due to a failure to follow the instructions of the study, or due to missing data. While this affected the statistical power of some of our analyses, similar trends were observed across different independent sensors (such as ECG, EDA, eye tracking), which validate the effectiveness of our workload manipulations.

6.2.2 Limitations of the ECG- and EDA-based measures

At the start of this project, the Empatica E4 (McCarthy et al., 2016b) wristband was used to collect physiological data. Empatica E4 included a Photoplethysmography (PPG) based heart rate sensor, and EDA sensor at the wrist. However, initial analysis from a simulator study indicated that the PPG data was too noisy, and could not be used in dynamic driving environments, such as those in the simulator. PPG uses a light sensor, which is highly susceptible to even small movements, corrupting the signal. Moreover, the sampling rate for Empatica E4 for PPG data was at 64 Hz, and that for EDA data was at 4 Hz, which was well below the recommended values (Braithwaite et al., 2015; Laborde et al., 2017). Therefore, in the following experiments, ECG and EDA signals were collected using a BIOPAC MP35 data acquisition system. One disadvantage of using this system was that it was more intrusive than the Empatica E4 system. Recent technological advancements have introduced more non-

intrusive (Beggiato et al., 2018), and even non-contact physiological sensors (Kranjec et al., 2014), that can be considered in future on-road studies.

Initially, there was an error in synchronising the BIOPAC MP35 system, with the simulator data, during the study reported in Chapter 3. This was identified during the data analysis stage, and therefore, certain adjustments had to be made to data analysis, as mentioned in section **6.2.1**. This issue was resolved prior to the next experiment. While motion and noise artefacts did affect both ECG and EDA signals, the methods mentioned in Chapter 2 helped remove these artefacts, providing clean signals, for accurate interpretations. Additionally, baseline (at rest) physiological data were collected, to account for any inter-individual differences arising from the physiological signals.

A large number of EDA and ECG-derived metrics (more details on the different metrics used are provided in the appendix) were analysed during this research. However, the metrics that showed the highest sensitivity to our experimental manipulations, and those that have been used in similar situations in prior research, (i.e. RMSSD, mean HR and nSCR/min), were carefully selected and included for analysis.

One of the drawbacks of using physiological data is that changes in drivers' physiological state could be caused by a host of psychological and/or environmental factors. For example, stress, workload, discomfort, arousal or ambient temperature could all result in an increase in drivers' EDA values. Therefore, the studies in this thesis had to be carefully designed, with strict experimental control, to avoid any confounders that could influence drivers' physiological state. For example, cardiac medications, prior physical activity, caffeine, differences in temperature in the simulator cabin, were all controlled for, to avoid any confounding effect on the ECG or EDA data. However, in a real-life situation, such control is not feasible. In section 6.3.3, I provide some thoughts on how these physiological signals can be used in real-world driving to correctly identify driver state.

6.2.3 Eye tracking based measures

For capturing drivers' eye-movements, the study reported in Chapter 5 used the Seeing Machines Fovio Driver Monitoring System. While creating visualisation plots, the sampling rate of 60 Hz was found not be consistent, with some frames having

a higher and some a lower sampling rate, with the mean value hovering around 60 Hz. This created problems for visualising the time-series data. Therefore, the data was interpolated to create a uniform time-width between consecutive frames in the eye-tracker. Another limitation of results from eye tracking data is that pupil diameter is sensitive to environmental factors, such as brightness levels. This was controlled in this study by ensuring drivers were driving through an urban environment across the whole drive, which had similar background lighting conditions, throughout. Therefore, the use of pupil diameter for measuring workload in real-world driving studies may pose some challenges, based on more variable lighting conditions.

6.2.4 The University of Leeds Driving Simulator (UoLDS)

There are number of advantages in using driving simulators, such as experimenting otherwise dangerous driving scenarios without putting the driver at risk, stricter experimental control by manipulating the virtual driving environment (de Winter et al., 2012), and providing repeated scenarios. This thesis focused on the use of physiological metrics as an objective indicator of driver state in real-world environments. Therefore, my experiments required full motion simulation, where drivers are subject to forces and movements similar to that observed in the real world, to validate whether such physiological metrics can be used in real-world driving settings. Due to its advanced motion-system, the UoLDS was able to elicit motion that was similar to that experienced in real-world driving scenarios. However, in addition to the sensitivity of physiological metrics to motion artefacts, which were partly resolved as highlighted above, the absence of complete realism provided by such simulators, when compared to real world studies, is acknowledged and must be taken into account when interpreting the results.

6.3 Contribution to the field and outlook

This thesis provides an understanding of how driver states of discomfort and workload are affected by different levels and stages of automated driving, indicating how these affect driver physiological metrics, and subjective feedback. While the development of various wearable sensors has increased the interest and ease of conducting psychophysiological research, this work is still limited within the automated driving domain. Furthermore, given the sensitivity of physiological signals such as ECG or

EDA to motion artefacts, further validation within dynamic driving environments ia required. The work presented in this thesis does not provide a comprehensive investigation of all driver states, or the link between all the human factors-related issues associated with such driver states. Rather, this research provides specific insights into driver states of discomfort and workload during HAD, factors influencing and affecting these states, and its measurement using subjective feedback, and physiological signals such as ECG, EDA and eye tracking. An overview of this these relationships is provided in **Figure 6.2**.

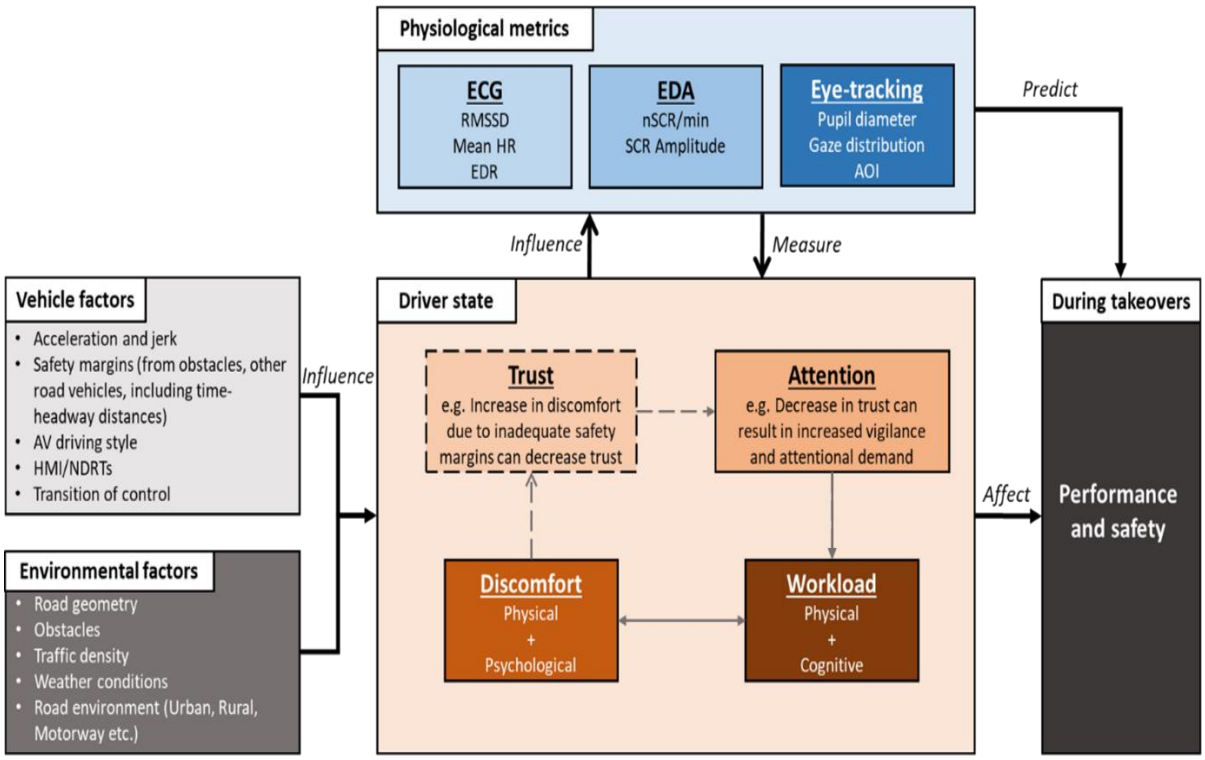

Figure 6.2. Schematic depiction of the relationship between vehicle/environmental factors, driver states and physiological metrics

This figure consists of four main components: 1) Factors that affect driver states; 2) Driver states of discomfort and workload; 3) Physiological metrics to measure driver states; 4) Outcomes in terms of performance and safety during takeovers. Each of these is further elaborated in the subsections below:

6.3.1 Factors affecting driver state

This figures provides a depiction of the various environmental and vehicle-based factors that influence driver discomfort, and driver workload, incorporating the

influence of additional factors interacting with these concepts, to influence driver state – using the current state of the art. On the left side of the **Figure 6.2**, there are vehicle-based and environment-based factors that influence driver state. Vehicle based factors such as acceleration forces, safety margins, AV driving style, HMI, NDRTs and request to takeover, can affect drivers' state. Similarly, various factors from the driving environment, such as road geometries, obstacles, traffic density, whether conditions, road type (e.g. rural roads, urban roads, motorway driving) influence driver state. However, the influence of vehicle-based and environmental-based factors on driver state is not necessarily mutually exclusive. For example, safety margins maintained by a vehicle depend on how the automated driving controller negotiates an obstacle or a road environment, and this interaction between the vehicle-based and environmental-based factors influence driver state. In the next subsection, I expand on the second theme in the figure, around driver states of discomfort and workload.

6.3.2 Driver states of discomfort and workload

In this thesis, I focused on understanding and measuring driver states of discomfort and workload during different stages of automated driving. On a broad level, there are physical (such as acceleration and jerk forces) and psychological factors (such as safety margins, naturalness of the drive) that affect driver discomfort. It is not necessary that the physical and psychological factors act separately, to cause driver discomfort. For example, the way in which and AV controller negotiates an obstacle, with harsh braking and steering controls (physical discomfort), while maintaining poor safety margins (psychological discomfort), can result in both physical and psychological discomfort. Similarly, workload has both physical and psychological components. Physical workload requires the usage of physical resources to perform a task, such as lifting an object. Psychological workload, requires mental resources to perform a task, such as a memory recall n-back task. Certain tasks, such as manual driving, is a psychomotor (combining cognitive functions and physical movements) task, that uses both physical and mental resources.

The relationship between driver discomfort and workload, is not well defined. Drivers can be affected by multiple driver states, at the same instant, and some of these driver states share conceptual similarities. For example, let us consider the safety margins factor and how it affects driver discomfort and workload. Safety

margins are a set of apparent minimum distances (based on the driver's judgement) to be maintained by vehicle from road edges, other road users, obstacles or hazards, for safe and smooth operation of the vehicle. In manual driving, drivers ascertain the safety by maintaining adequate safety margins. However, in automated driving, the automated vehicle controller controls and modulates these safety margins, with little or no input from the driver (Cahour, 2008). When a drivers' threshold for what is deemed a safe margin is breached, such as when an automated vehicle controller is following a lead vehicle quite closely (Siebert & Wallis, 2019), it can result in driver discomfort. It can also negatively affect drivers' trust (Carsten & Martens, 2019), as shown using the dotted grey arrow in **Figure 6.2**. Since trust was not the main focus of this thesis, the relationship between discomfort and trust is shown using dotted-grey arrows as it is an implicit inference. The lack of trust in the automated system's driving capabilities, can cause the driver to become more vigilant and monitor the driving task, resulting in usage of additional attentional resources. Again this is indicated by dotted-grey lines in **Figure 6.2**, as it is an implicit inference. However, using additional resources can also result in increasing drivers' workload levels, as indicated by the grew arrow in **Figure 6.2**. This result was also seen in Chapters 4 and 5 of this thesis, where shorter headways maintained by the automated vehicle from a lead vehicle, increased drivers' workload levels. Therefore, a breach in safety margins, can simultaneously result in both discomfort and increased workload to the driver. This mutual relationship between discomfort and workload, where discomfort to the driver can also result in increased workload, and vice-versa, is indicated by the double headed grey arrow in **Figure 6.2**. While driver states of trust and attention were not the main focus of this thesis, I believe that particularly around safety margins, understanding the potential implications on drivers' trust and attention helps in better highlighting the relationship between workload and discomfort. Future studies on driver states of discomfort and workload should also consider related driver states such as trust and attention, to better understand the nature of this inter-relationship existing between the two driver states.

With regards to the inter-relationship existing between trust, attention and workload, Heikoop et al. (2016) suggested that trust can negatively influence drivers' attention, which is observed in our example above. That is, a decrease in trust can increase driver's attention. They also suggested a U-shaped causality existing

between workload and attention. That is, low or high workload can result in diminished attention at the driving task, or in our example, increased attention can lead to higher levels of workload. It is also clear that a driver can be affected by multiple driver states at the same time. For example, when the safety margins are breached by the automated vehicle controller, this can result in driver having increased discomfort, decreased trust, higher attention and vigilance, higher stress, and higher workload levels. The exact nature of this inter-relationship between the different driver states still remains unclear. In order to accurately determine drivers' capabilities and limitations, it is important to incorporate a holistic approach, and consider all relevant states that affect the driver at a specific instant, and their implications. Future research should take this into account, when creating predictor models that can predict drivers' future performance, by objectively measuring their current and previous driver states.

6.3.3 Physiological metrics and driver state

In this research, I used physiological data from ECG, EDA and eye tracking sensors, to understand and measure driver states of workload and discomfort. It should be noted that a plethora of metrics can be derived from each of these sensors. For example, Shaffer & Ginsberg (2017) lists 29 different ECG-derived metrics, and MATLAB-based EDA software Ledalab (Benedek, 2015) includes 15 different metrics for EDA. However, after careful review of the literature, following an analysis of the different metrics, and removal of highly correlated metrics, RMSSD, mean HR and EDR from ECG, and nSCR/min from EDA, were selected to highlight the changes observed in drivers' discomfort and workload levels. For eye tracking measures, previous research was used to select pupil diameter as the best metric for identifying changes in drivers' workload levels. Driver's gaze was used to analyse and understand where their visual attention was allocated during the experimental manipulations.

Results from Chapter 3 indicated the nSCR/min is a better indicator for driver discomfort than ECG-based metrics, as it captures both long-term variations, as well as instantaneous changes in discomfort, brought about by environmental and driving condition. This is partly because ECG signals are more susceptible to prior manipulations, owing to larger decay times in this signal. However, this can be overcome by collecting baseline and post-event recovery data for the ECG signal, and

modelling the signal's recovery pattern to account for carry-over effects from prior manipulations (Laborde et al., 2017; Radhakrishnan et al., 2020).

To understand and measure workload, RMSSD, mean HR, EDR, nSCR/min and pupil diameter metrics, were used. Results from Chapters 4 and 5, indicated that ECG-based metrics such as RMSSD, mean HR and EDR, were better indicators of longer-term variations in driver workload, such as those due to monitoring the drive, or engagement in NDRTs, during periods of automation. However, given that ECG signals require at least 2 minute windows for accurate measurement (Bourdillon et al., 2017), and also because they have carry-over effects from prior stimuli, they were not deemed suitable for identifying short-term workload variations. On the other hand, both nSCR/min and pupil diameter were able to capture short- and long-term variations in driver workload.

I was also able to use physiological metrics to separate the physical and cognitive aspects of driver workload, based on its sensitivity to different physiological metrics, which I believe is a novel finding. RMSSD, mean HR, nSCR/min and pupil diameter were all more sensitive to the cognitive element of workload experienced during driving. On the other hand, only the EDR metric, which is indicative of respiration rate, was sensitive to the physical demands imposed by the more physically demanding tasks, such as manual driving, as well as the cognitive demands posed by various workload-inducing manipulations in the study, such as monitoring the drive, engaging in the Arrows tasks. Therefore, the EDR metric is considered to be a better indicator of drivers' overall (physical + cognitive) workload. I believe future research can incorporate the EDR metric, along with indicators of cognitive workload such as RMSSD, mean HR, pupil diameter and nSCR/min, to distinguish between the physical and cognitive elements of workload.

6.3.4 Driver state, performance and safety

An important observation from my research was that data from a single sensor is not enough to correctly identify driver state in a real-world setting. For example, an increase in drivers' RMSSD values, mean HR, EDR or nSCR/min could be due to discomfort, stress, increase in workload, arousal or even physical activity. It is important to combine these metrics with eye tracking and vehicle data, to pinpoint the causal factor for physiological changes. Additionally, eye tracking data can become

unavailable in certain use-cases, especially for higher levels of vehicle automation, such as when drivers are allowed to engage in an NDRT during Level 3 automation, which means their eyes are likely to be occluded from the eye tracking cameras. Physiological signals such as EDA and ECG are then useful at this stage, as they can monitor drivers' workload levels. Since there is a desire to increase the capability of automated vehicles and drivers are more and more likely to engage in NDRTs in higher levels of automation, future research should investigate the value of sensor-fusion for different physiological sensors. Combined with suitable machine learning algorithms, this technique may be valuable for identifying drivers' state, establishing their capability to resume control, by predicting the likelihood of impending performance decrements, and providing appropriate support to the driver, should they be required to resume control from automation, for example at the end of an ODD.

6.4 Conclusion

To conclude, this thesis provides a deeper understanding of driver states of discomfort and workload, and the factors influencing these driver states during automated driving. I was able to validate the use of physiological metrics such as RMSSD, mean HR, EDR, pupil diameter and gaze metric for understanding, and measuring, driver states of discomfort and workload. Future work can build on this research by incorporating sensor fusion of ECG and EDA based data, along with eye tracking, to help improve the accuracy and capabilities of future driver state monitoring systems.

6.5 References

Beggiato, M., Hartwich, F., & Krems, J. (2018). Using Smartbands, Pupillometry and Body Motion to Detect Discomfort in Automated Driving. *Frontiers in Human Neuroscience*, *12*, 338. https://doi.org/10.3389/fnhum.2018.00338

Beggiato, M., Hartwich, F., & Krems, J. (2019). Physiological correlates of discomfort in automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *66*, 445–458. https://doi.org/10.1016/j.trf.2019.09.018

Benedek, M. (2015). *Ledalab V3.4.9* (3.4.9).

Boucsein, W. (2012). *Electrodermal activity* (Second). Springer. https://doi.org/https://doi.org/10.1007/978-1-4614-1126-0

Bourdillon, N., Schmitt, L., Yazdani, S., Vesin, J. M., & Millet, G. P. (2017). Minimal window duration for accurate HRV recording in athletes. *Frontiers in Neuroscience*, *11*(AUG), 456. https://doi.org/10.3389/fnins.2017.00456

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2015). *A Guide for Analysing Electrodermal Activity & Skin Conductance Responses (SCRs) for Psychophysiological Experiments*. https://doi.org/10.1017.S0142716405050034

Bueno, M., Dogan, E., Selem, F. H., Monacelli, E., Boverie, S., & Guillaume, A. (2016). How different mental workload levels affect the take-over control after automated driving. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2040–2045. https://doi.org/10.1109/ITSC.2016.7795886

Cahour, B. (2008). Discomfort, affects and coping strategies in driving activity. *15th European Conference on Cognitive Ergonomics: The Ergonomics of Cool Interaction (ECCE '08)*, *369*, 1–7. https://doi.org/10.1145/1473018.1473046

Carsten, O., & Martens, M. H. (2019). How can humans understand their automated cars? HMI principles, problems and solutions. *Cognition, Technology and Work*, *21*(1), 3–20. https://doi.org/10.1007/s10111-018-0484-0

de Winter, J. C. F., van Leeuwen, P. M., & Happee, R. (2012). Advantages and Disadvantages of Driving Simulators: A Discussion. *Proceedings of the Measuring Behavior Conference*, 47–50. www.noldus.com

Eriksson, J., & Svensson, L. (2015). *Tuning for Ride Quality in Autonomous Vehicle Application to Linear Quadratic Path Planning Algorithm* [Uppsala University]. http://www.teknat.uu.se/student

Foy, H. J., & Chapman, P. (2018). Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation. *Applied Ergonomics*, *73*, 90–99. https://doi.org/10.1016/j.apergo.2018.06.006

Heikoop, D. D., de Winter, J. C. F., van Arem, B., & Stanton, N. A. (2016). Psychological constructs in driving automation: a consensus model and critical comment on construct proliferation. *Theoretical Issues in Ergonomics Science*,

17(3), 284–303. https://doi.org/10.1080/1463922X.2015.1101507

Hidalgo-Muñoz, A. R., Béquet, A. J., Astier-Juvenon, M., Pépin, G., Fort, A., Jallais, C., Tattegrain, H., & Gabaude, C. (2019). Respiration and Heart Rate Modulation Due to Competing Cognitive Tasks While Driving. *Frontiers in Human Neuroscience*, *12*, 525. https://doi.org/10.3389/fnhum.2018.00525

Kikhia, B., Stavropoulos, T. G., Andreadis, S., Karvonen, N., Kompatsiaris, I., Sävenstedt, S., Pijl, M., & Melander, C. (2016). Utilizing a wristband sensor to measure the stress level for people with dementia. *Sensors (Switzerland)*, *16*(12). https://doi.org/10.3390/s16121989

Kranjec, J., Beguš, S., Geršak, G., & Drnovšek, J. (2014). Non-contact heart rate and heart rate variability measurements: A review. In *Biomedical Signal Processing and Control* (Vol. 13, Issue 1, pp. 102–112). Elsevier. https://doi.org/10.1016/j.bspc.2014.03.004

Laborde, S., Mosley, E., & Thayer, J. F. (2017). Heart rate variability and cardiac vagal tone in psychophysiological research - Recommendations for experiment planning, data analysis, and data reporting. In *Frontiers in Psychology* (Vol. 8, p. 213). Frontiers. https://doi.org/10.3389/fpsyg.2017.00213

Louw, T., Markkula, G., Boer, E., Madigan, R., Carsten, O., & Merat, N. (2017). Coming back into the loop: Drivers' perceptual-motor performance in critical events after automated driving. *Accident Analysis and Prevention*, *108*, 9–18. https://doi.org/10.1016/j.aap.2017.08.011

Mathôt, S. (2018). Pupillometry: Psychology, Physiology, and Function. *Journal of Cognition*, *1*(1), 1–23. https://doi.org/10.5334/joc.18

McCarthy, C., Pradhan, N., Redpath, C., & Adler, A. (2016). Validation of the Empatica E4 wristband. *2016 IEEE EMBS International Student Conference: Expanding the Boundaries of Biomedical Engineering and Healthcare, ISC 2016 - Proceedings*, 1–4. https://doi.org/10.1109/EMBSISC.2016.7508621

Mehler, B., Reimer, B., Coughlin, J., & Dusek, J. (2009). Impact of Incremental Increases in Cognitive Workload on Physiological Arousal and Performance in Young Adult Drivers. *Transportation Research Record: Journal of the Transportation Research Board*, *2138*, 6–12. https://doi.org/10.3141/2138-02

Merat, N., Jamson, A. H., Lai, F. C. H., Daly, M., & Carsten, O. M. J. (2014). Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*(PB), 274–282. https://doi.org/10.1016/j.trf.2014.09.005

Mourant, R. R., & Thattacherry, T. R. (2000). Simulator sickness in a virtual environments driving simulator. *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Association, "Ergonomics for the New Millennium,"* 534–537. https://doi.org/10.1177/154193120004400513

Müller, A. L., Fernandes-Estrela, N., Hetfleisch, R., Zecha, L., & Abendroth, B. (2021). Effects of non-driving related tasks on mental workload and take-over times during conditional automated driving. *European Transport Research Review*, *13*(1), 1–15. https://doi.org/10.1186/s12544-021-00475-5

National Highway Traffic Safety Administration. (2009). *Traffic Safety Facts 2009: A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System*. U.S. Department of Transportation. http://www.nhtsa.gov/NCSA.

Radhakrishnan, V., Merat, N., Louw, T., Lenné, M. G., Romano, R., Paschalidis, E., Hajiseyedjavadi, F., Wei, C., & Boer, E. R. (2020). Measuring drivers' physiological response to different vehicle controllers in highly automated driving (HAD): Opportunities for establishing real-time values of driver discomfort. *Information (Switzerland)*, *11*(8), 390. https://doi.org/10.3390/INFO11080390

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). AUTOMATION SURPRISES. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (2nd ed., pp. 1926–1943). Wiley.

Schartmüller, C., Weigl, K., Löcken, A., Wintersberger, P., Steinhauser, M., & Riener, A. (2021). Displays for productive non-driving related tasks: Visual behavior and its impact in conditionally automated driving. *Multimodal Technologies and Interaction*, *5*(4), 21. https://doi.org/10.3390/MTI5040021

Shaffer, F., & Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and Norms. In *Frontiers in Public Health* (Vol. 5, p. 258). Frontiers Media SA.

https://doi.org/10.3389/fpubh.2017.00258

Siebert, F. W., Oehl, M., Höger, R., & Pfister, H. R. (2013). Discomfort in Automated Driving - The Disco-Scale. *Communications in Computer and Information Science*, *374*(PART II), 337–341. https://doi.org/10.1007/978-3-642-39476-8_69

Siebert, F. W., & Wallis, F. L. (2019). How speed and visibility influence preferred headway distances in highly automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *64*, 485–494. https://doi.org/10.1016/j.trf.2019.06.009

van Gent, P., Farah, H., van Nes, N., & van Arem, B. (2018). Heart Rate Analysis for Human Factors: Development and Validation of an Open Source Toolkit for Noisy Naturalistic Heart Rate Data. *Humanist 2018 Conference*, *June*.

Wandtner, B., Schömig, N., & Schmidt, G. (2018). Effects of Non-Driving Related Task Modalities on Takeover Performance in Highly Automated Driving. *Human Factors*, *60*(6), 870–881. https://doi.org/10.1177/0018720818768199

Wiener, E. L. (1989). *Human Factors of Advanced Technology ("Glass Cockpit") Transport Aircraft*. NASA Contractor Report 177528. https://human-factors.arc.nasa.gov/publications/HF_AdvTech_Aircraft.pdf

Yang, S., Wilson, K., Roady, T., Kuo, J., & Lenné, M. G. (2022). Beyond gaze fixation: Modeling peripheral vision in relation to speed, Tesla Autopilot, cognitive load, and age in highway driving. *Accident Analysis & Prevention*, *171*, 106670. https://doi.org/10.1016/j.aap.2022.106670

Zeeb, K., Buchner, A., & Schrauf, M. (2016). Is take-over time all that matters? the impact of visual-cognitive load on driver take-over quality after conditionally automated driving. *Accident Analysis and Prevention*, *92*, 230–239. https://doi.org/10.1016/j.aap.2016.04.002

APPENDIX A

Contents (DOI: 10.13140/RG.2.2.24091.54560)

EDA filtering

```
% Author : Vishnu Radhakrishnan
% Licence: mn16vr@leeds.ac.uk
% copyright protected under the author licence.

% For this code to be copied, modified and distributed,  prior permission
% from the author
% is to be requested AND is subject to the following conditions:-
%
% - The code is not to be used for commercial gain.
% - The code and use thereof will be attributed to the authors where
%   appropriate (including papers, presentations and demonstrations which
%   rely on its use).
% - All modified, distributions of the source files will retain this
%   header.
%
% Usage of this code in analysis of any published work requires prior
% permission from the author, and the user is required to cite the
% following paper:
% Radhakrishnan,V.; Solernou, A.; Merat, N. (2022). Artefact Correction for
Electrodermal Activity Signals in Dynamic Driving Environments. Pre-print. DOI:
10.13140/RG.2.2.24091.54560

% ************************************************************************

function [eda_filtered, slope]= eda_filter(eda_data,varargin)
```

Input arguments

```
% Input arguments = edafilter(eda_data,seconds,slope_constraint,...
% max_slope,min_slope,contact_loss,linear_value,time_window,smoothing,...
% cutoff_freq,filt_order)

% eda_data an nx2 matrix, with first column containing time in seconds, and
% second column containing EDA values in µS (microsiemens)

% varargin can contain upto 8 optional arguments

% second argument is seconds (numeric value, a number between 1 and 6), and
% it details for what window-length the shape-based algorithm is performed
% for. If it is 6 (default value), then the shape-based algorithm considers
% window-lengths from 1:6 seconds.

% third argument argument is slope constraint ('Yes' or 'No'), and details
% whether a slope-based noise detection algorithm is perfromed for 1 second
```

```
% windows, where the slope of the signal cannot exceed max_slope, or be
% below min_slope.

% fourth argument argument is percentile ('Yes' or 'No'), and finds,
% max_slope and min_slope values, based on linear fitting for 1s windows,
% for all non-nan values after impletementing a 1 second window shape based
% algorithms, deriving the 5th and 95th percentile values for the entire
% data set, to be used as max-slope and min_slope respectively.

% fifth input agrument is max_slope (between 0 and 1), which defines the
% maximum allowable slope for the slope constraint (default at 0.7).

% sixth input argument is min_slope (between 0 and -1), which defines the
% minimum allowable slope for the slope constraint (default at -0.35).

% seventh input arugment is contact loss, a decimal number (default 0.05),
% where below its value, the electrodes is considered to have lost contact
% with skin and replaced with NaN

% eigth input argument is linear value (default is No), which specifics
% whether if there are consecutive NaN values for more than time-window
% seconds, then whether or not a linear interopolation is done ('Yes', does
% not affect SCRs, but will affect SCL), or removed ('No', data points will
% be replaced by NaN values, and wont be continous).

% ninth imput argument is time-window (default 5) in seconds, which
% specifies for what time window if the data remains as NaN values, a
% linear inpterpolation is done.

% tenth input argument is smoothing (default is Yes), which determines
% whether ('Yes') or not ('No') a low-pass iir butterworth filter is
% applied to smooth the data or not, and remove high frequency noise.

% eleventh input argument is cutoff_freq (default 3 Hz), which determines the
% cutoff frequency for the butterworth filter

% twelth input arguement is filt_order(default is 2), which determines
% the filter order of the butterworth filter.
```

Output arguments

```
% the function returns eda_filtered, an nx2, matrix, with timestamps on the
% first column
```

Code

```
% check for maximum number of optional input arguments
numvarargs = length(varargin);
if numvarargs > 11
    error('myfuns:edafilter:TooManyInputs', ...
        'too many input arguments (must have at least 1, and at most 12)');
end

% check for size of eda_data
no_columns = size(eda_data,2);
if no_columns ~= 2
    error('myfuns:edafilter:MatrixDimension', ...
        'size of eda_data is incorrect (must be an nx2 matrix)');
end
```

```
% set defaults for optional inputs
optargs = {3 'No' 'Yes' 0.4 -0.2 0.05 'No' 5 'Yes' 3 2};

% replace default values with optinal input arguments, if entered
[optargs{1:numvarargs}] = varargin{:};
[seconds,slope_constraint,percentile,max_slope,min_slope,contact_loss,linear_value,
...
    time_window,smoothing,cutoff_freq,filt_order] = optargs{:};

% check for time window of the shape based filter
if (rem(seconds,1)~=0 || seconds>8 || seconds<1)
    error('myfuns:edafilter:IncorrectValue', ...
        'seconds (2nd argument) must be an integer value between 1 and 8');
end

% check slope-constraint
if (strcmp(slope_constraint,'Yes')==0 && strcmp(slope_constraint,'No')==0)
    error('myfuns:edafilter:IncorrectValue', ...
        'slope_constraint (3rd arugment) must be a char of value Yes or No');
end

% check percentile
if (strcmp(percentile,'Yes')==0 && strcmp(percentile,'No')==0)
    error('myfuns:edafilter:IncorrectValue', ...
        'slope_constraint (3rd arugment) must be a char of value Yes or No');
end

% check for max_slope
if ( max_slope>1 || max_slope<0)
    error('myfuns:edafilter:IncorrectValue', ...
        'max_slope(5th arugment) must be a value between 0 and 1');
end

% check for min_slope
if ( min_slope>0 || min_slope<-1)
    error('myfuns:edafilter:IncorrectValue', ...
        'min_slope (6th arugment) must be a value between -1 and 0');
end

% check linear filter value
if (strcmp(linear_value,'Yes')==0 && strcmp(linear_value,'No')==0)
    error('myfuns:edafilter:IncorrectValue', ...
        'linear_value (8th arugment) must be a char of value Yes or No');
end

% check linear filter value
if (strcmp(smoothing,'Yes')==0 && strcmp(smoothing,'No')==0)
    error('myfuns:edafilter:IncorrectValue', ...
        'smoothing (10th arugment) must be a char of value Yes or No');
end

sampling_rate = 1/(eda_data(4,1)-eda_data(3,1));
sampling_rate = round(sampling_rate);

j = size(eda_data,1);

% Check if the slope-based filter is applied, and if the percentile values
% need to be calculated for max and min slope.

if strcmp (slope_constraint,'Yes') && strcmp(percentile,'Yes') == 1
    eda_prct = eda_data(:,:);

    for m1 = 1:j
```

```
        var2 = isnan(eda_data(m1,2));
        if var2 == 1
            continue;
        else
            for k1=1:sampling_rate
                n0 =m1+k1;
                if (n0<=j)
                    if(eda_prct(n0,2)> 1.2*eda_prct(m1,2))
                        eda_prct(n0,2)= NaN;
                    elseif (eda_prct(n0,2)<0.9*eda_prct(m1,2))
                        eda_prct(n0,2)= NaN;
                    end
                end
            end
        end
    end

    % Determine max_slope and min_slope values using linear fitting for 1s
    % windows

    for m2 = 1:j

        var1 = isnan(eda_prct(m2,2));
        if var1 == 1
            continue;
        else
            k2 = m2+sampling_rate;
            if k2<=j
                eda1 = eda_prct(m2:k2,2);
                t1 = eda_prct(m2:k2,1);
                idx1 = isnan(eda1);
                p1 = polyfit(t1(~idx1),eda1(~idx1),1);
                s1(m2) = p1(:,1);
            end
            clearvars eda1 t1
        end
    end
    slope(1,1) = max(s1);
    slope(1,2) = min(s1);
    slope(1,3) = mean(s1(s1>0));
    slope(1,4) = mean(s1(s1<0));
    slope(1,5) = prctile(s1(s1>0),95);
    slope(1,6) = prctile(s1(s1<0),5);

    if slope(1,5)>0.4 && slope(1,6)<-0.2
        max_slope = slope(1,5);
        min_slope = slope(1,6);
    end

else
    slope(1,1) = NaN;
    slope(1,2) = NaN;
    slope(1,3) = NaN;
    slope(1,4) = NaN;
    slope(1,5) = NaN;
    slope(1,6) = NaN;

end
```

Shape-based algorithm to remove noise

where an EDA signal does not increase by more than 20% or decrease by more than 10%, within a 1 second time window. max_slope and min_slope based on this calculation, with a 3x buffer

```
d1 = designfilt('lowpassiir','FilterOrder',filt_order, ...

'HalfPowerFrequency',cutoff_freq,'DesignMethod','butter','SampleRate',sampling_rate
);

if strcmp(smoothing,'Yes')==1
    eda_data(:,2) = filtfilt(d1,eda_data(:,2));
end

switch seconds  %switch-case statement to apply the algorithms
    % from 1:n seconds, where n is given by input
    % argument seconds

    case {1}
        for m = 1:j
            var = isnan(eda_data(m,2));
            if var == 1
                continue;
            else
                for k=1:sampling_rate
                    n =m+k;

                    if (n<=j)
                        z = (eda_data(n,2)-eda_data(m,2))/(eda_data(n,1)-
eda_data(m,1));

                        if strcmp (slope_constraint,'Yes')==1 && ~isnan(z)
                            if(eda_data(n,2)> 1.2*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif (eda_data(n,2)<0.9*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif z>max_slope
                                eda_data(n,2)= NaN;
                            elseif z<min_slope
                                eda_data(n,2)= NaN;
                            end

                        else
                            if(eda_data(n,2)> 1.2*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif (eda_data(n,2)<0.9*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            end
                        end
                    end
                end
            end
        end


    case {2}
        for m = 1:j
            var = isnan(eda_data(m,2));
            if var == 1
                continue;
            else
                for k=1:sampling_rate
                    n =m+k;
                    n2 = n+sampling_rate;

                    if (n<=j && n2<=j)
                        z = (eda_data(n,2)-eda_data(m,2))/(eda_data(n,1)-
eda_data(m,1));

                        if strcmp (slope_constraint,'Yes')==1 && ~isnan(z)
                            if(eda_data(n,2)> 1.2*eda_data(m,2))
```

```
                                eda_data(n,2)= NaN;
                            elseif (eda_data(n,2)<0.9*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif z>max_slope
                                eda_data(n,2)= NaN;
                            elseif z<min_slope
                                eda_data(n,2)= NaN;
                            elseif(eda_data(n2,2)> 1.44*eda_data(m,2))
                                eda_data(n2,2)= NaN;
                            elseif (eda_data(n2,2)<0.81*eda_data(m,2))
                                eda_data(n2,2)= NaN;
                            end

                        else
                            if(eda_data(n,2)> 1.2*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif (eda_data(n,2)<0.9*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif(eda_data(n2,2)> 1.44*eda_data(m,2))
                                eda_data(n2,2)= NaN;
                            elseif (eda_data(n2,2)<0.81*eda_data(m,2))
                                eda_data(n2,2)= NaN;
                            end
                        end
                    end
                end
            end
        end


    case {3}
        for m = 1:j
            var = isnan(eda_data(m,2));
            if var == 1
                continue;
            else
                for k=1:sampling_rate
                n =m+k;
                n2 = n+sampling_rate;
                n3 = n2+sampling_rate;

                if (n<=j && n2<=j && n3<=j )
                    z = (eda_data(n,2)-eda_data(m,2))/(eda_data(n,1)-
eda_data(m,1));

                    if strcmp (slope_constraint,'Yes')==1 && ~isnan(z)
                        if(eda_data(n,2)> 1.2*eda_data(m,2))
                            eda_data(n,2)= NaN;
                        elseif (eda_data(n,2)<0.9*eda_data(m,2))
                            eda_data(n,2)= NaN;
                        elseif z>max_slope
                            eda_data(n,2)= NaN;
                        elseif z<min_slope
                            eda_data(n,2)= NaN;
                        elseif(eda_data(n2,2)> 1.44*eda_data(m,2))
                            eda_data(n2,2)= NaN;
                        elseif (eda_data(n2,2)<0.81*eda_data(m,2))
                            eda_data(n2,2)= NaN;
                        elseif(eda_data(n3,2)> 1.728*eda_data(m,2))
                            eda_data(n3,2)= NaN;
                        elseif(eda_data(n3,2)<0.729*eda_data(m,2))
                            eda_data(n3,2)= NaN;
                        end

                    else
                        if(eda_data(n,2)> 1.2*eda_data(m,2))
                            eda_data(n,2)= NaN;
```

```
                            elseif (eda_data(n,2)<0.9*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif(eda_data(n2,2)> 1.44*eda_data(m,2))
                                eda_data(n2,2)= NaN;
                            elseif (eda_data(n2,2)<0.81*eda_data(m,2))
                                eda_data(n2,2)= NaN;
                            elseif(eda_data(n3,2)> 1.728*eda_data(m,2))
                                eda_data(n3,2)= NaN;
                            elseif(eda_data(n3,2)<0.729*eda_data(m,2))
                                eda_data(n3,2)= NaN;
                            end
                        end
                    end
                end
            end
        end


    case {4}
        for m = 1:j
            var = isnan(eda_data(m,2));
            if var == 1
                continue;
            else
                for k=1:sampling_rate

                    n =m+k;
                    n2 = n+sampling_rate;
                    n3 = n2+sampling_rate;
                    n4 = n3+sampling_rate;

                    if (n<=j && n2<=j && n3<=j && n4<=j )
                        z = (eda_data(n,2)-eda_data(m,2))/(eda_data(n,1)-
eda_data(m,1));

                        if strcmp (slope_constraint,'Yes')==1 && ~isnan(z)
                            if(eda_data(n,2)> 1.2*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif (eda_data(n,2)<0.9*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif z>max_slope
                                eda_data(n,2)= NaN;
                            elseif z<min_slope
                                eda_data(n,2)= NaN;
                            elseif(eda_data(n2,2)> 1.44*eda_data(m,2))
                                eda_data(n2,2)= NaN;
                            elseif (eda_data(n2,2)<0.81*eda_data(m,2))
                                eda_data(n2,2)= NaN;
                            elseif(eda_data(n3,2)> 1.728*eda_data(m,2))
                                eda_data(n3,2)= NaN;
                            elseif(eda_data(n3,2)<0.729*eda_data(m,2))
                                eda_data(n3,2)= NaN;
                            elseif(eda_data(n4,2)> 2.0736*eda_data(m,2))
                                eda_data(n4,2)= NaN;
                            elseif (eda_data(n4,2)<0.6561*eda_data(m,2))
                                eda_data(n4,2)= NaN;
                            end

                        else
                            if(eda_data(n,2)> 1.2*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif (eda_data(n,2)<0.9*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif(eda_data(n2,2)> 1.44*eda_data(m,2))
                                eda_data(n2,2)= NaN;
                            elseif (eda_data(n2,2)<0.81*eda_data(m,2))
                                eda_data(n2,2)= NaN;
```

```
                            elseif(eda_data(n3,2)> 1.728*eda_data(m,2))
                                eda_data(n3,2)= NaN;
                            elseif(eda_data(n3,2)<0.729*eda_data(m,2))
                                eda_data(n3,2)= NaN;
                            elseif(eda_data(n4,2)> 2.0736*eda_data(m,2))
                                eda_data(n4,2)= NaN;
                            elseif (eda_data(n4,2)<0.6561*eda_data(m,2))
                                eda_data(n4,2)= NaN;
                            end
                        end
                    end
                end
            end
        end


    case {5}
        for m = 1:j
            var = isnan(eda_data(m,2));
            if var == 1
                continue;
            else
                for k=1:sampling_rate

                    n =m+k;
                    n2 = n+sampling_rate;
                    n3 = n2+sampling_rate;
                    n4 = n3+sampling_rate;
                    n5 = n4+sampling_rate;

                    if (n<=j && n2<=j && n3<=j && n4<=j && n5<=j)
                        z = (eda_data(n,2)-eda_data(m,2))/(eda_data(n,1)-
eda_data(m,1));

                        if strcmp (slope_constraint,'Yes')==1 && ~isnan(z)
                            if(eda_data(n,2)> 1.2*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif (eda_data(n,2)<0.9*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif z>max_slope
                                eda_data(n,2)= NaN;
                            elseif z<min_slope
                                eda_data(n,2)= NaN;
                            elseif(eda_data(n2,2)> 1.44*eda_data(m,2))
                                eda_data(n2,2)= NaN;
                            elseif (eda_data(n2,2)<0.81*eda_data(m,2))
                                eda_data(n2,2)= NaN;
                            elseif(eda_data(n3,2)> 1.728*eda_data(m,2))
                                eda_data(n3,2)= NaN;
                            elseif(eda_data(n3,2)<0.729*eda_data(m,2))
                                eda_data(n3,2)= NaN;
                            elseif(eda_data(n4,2)> 2.0736*eda_data(m,2))
                                eda_data(n4,2)= NaN;
                            elseif (eda_data(n4,2)<0.6561*eda_data(m,2))
                                eda_data(n4,2)= NaN;
                            elseif(eda_data(n5,2)> 2.48832*eda_data(m,2))
                                eda_data(n5,2)= NaN;
                            elseif (eda_data(n5,2)<0.59049*eda_data(m,2))
                                eda_data(n5,2)= NaN;
                            end

                        else
                            if(eda_data(n,2)> 1.2*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif (eda_data(n,2)<0.9*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif(eda_data(n2,2)> 1.44*eda_data(m,2))
```

```
                                eda_data(n2,2)= NaN;
                            elseif (eda_data(n2,2)<0.81*eda_data(m,2))
                                eda_data(n2,2)= NaN;
                            elseif(eda_data(n3,2)> 1.728*eda_data(m,2))
                                eda_data(n3,2)= NaN;
                            elseif(eda_data(n3,2)<0.729*eda_data(m,2))
                                eda_data(n3,2)= NaN;
                            elseif(eda_data(n4,2)> 2.0736*eda_data(m,2))
                                eda_data(n4,2)= NaN;
                            elseif (eda_data(n4,2)<0.6561*eda_data(m,2))
                                eda_data(n4,2)= NaN;
                            elseif(eda_data(n5,2)> 2.48832*eda_data(m,2))
                                eda_data(n5,2)= NaN;
                            elseif (eda_data(n5,2)<0.59049*eda_data(m,2))
                                eda_data(n5,2)= NaN;
                            end
                        end
                    end
                end
            end
        end


    case {6}
        for m = 1:j
            var = isnan(eda_data(m,2));
            if var == 1
                continue;
            else
                for k=1:sampling_rate

                    n =m+k;
                    n2 = n+sampling_rate;
                    n3 = n2+sampling_rate;
                    n4 = n3+sampling_rate;
                    n5 = n4+sampling_rate;
                    n6 = n5+sampling_rate;

                    if (n<=j && n2<=j && n3<=j && n4<=j && n5<=j && n6<=j)
                        z = (eda_data(n,2)-eda_data(m,2))/(eda_data(n,1)-
eda_data(m,1));

                        if strcmp (slope_constraint,'Yes')==1 && ~isnan(z)
                            if(eda_data(n,2)> 1.2*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif (eda_data(n,2)<0.9*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif z>max_slope
                                eda_data(n,2)= NaN;
                            elseif z<min_slope
                                eda_data(n,2)= NaN;
                            elseif(eda_data(n2,2)> 1.44*eda_data(m,2))
                                eda_data(n2,2)= NaN;
                            elseif (eda_data(n2,2)<0.81*eda_data(m,2))
                                eda_data(n2,2)= NaN;
                            elseif(eda_data(n3,2)> 1.728*eda_data(m,2))
                                eda_data(n3,2)= NaN;
                            elseif(eda_data(n3,2)<0.729*eda_data(m,2))
                                eda_data(n3,2)= NaN;
                            elseif(eda_data(n4,2)> 2.0736*eda_data(m,2))
                                eda_data(n4,2)= NaN;
                            elseif (eda_data(n4,2)<0.6561*eda_data(m,2))
                                eda_data(n4,2)= NaN;
                            elseif(eda_data(n5,2)> 2.48832*eda_data(m,2))
                                eda_data(n5,2)= NaN;
                            elseif (eda_data(n5,2)<0.59049*eda_data(m,2))
                                eda_data(n5,2)= NaN;
```

```
                            elseif(eda_data(n6,2)> 2.985984*eda_data(m,2))
                                eda_data(n6,2)= NaN;
                            elseif(eda_data(n6,2)<0.531441*eda_data(m,2))
                                eda_data(n6,2)= NaN;
                            end

                        else
                            if(eda_data(n,2)> 1.2*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif (eda_data(n,2)<0.9*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif(eda_data(n2,2)> 1.44*eda_data(m,2))
                                eda_data(n2,2)= NaN;
                            elseif (eda_data(n2,2)<0.81*eda_data(m,2))
                                eda_data(n2,2)= NaN;
                            elseif(eda_data(n3,2)> 1.728*eda_data(m,2))
                                eda_data(n3,2)= NaN;
                            elseif(eda_data(n3,2)<0.729*eda_data(m,2))
                                eda_data(n3,2)= NaN;
                            elseif(eda_data(n4,2)> 2.0736*eda_data(m,2))
                                eda_data(n4,2)= NaN;
                            elseif (eda_data(n4,2)<0.6561*eda_data(m,2))
                                eda_data(n4,2)= NaN;
                            elseif(eda_data(n5,2)> 2.48832*eda_data(m,2))
                                eda_data(n5,2)= NaN;
                            elseif (eda_data(n5,2)<0.59049*eda_data(m,2))
                                eda_data(n5,2)= NaN;
                            elseif(eda_data(n6,2)> 2.985984*eda_data(m,2))
                                eda_data(n6,2)= NaN;
                            elseif(eda_data(n6,2)<0.531441*eda_data(m,2))
                                eda_data(n6,2)= NaN;
                            end
                        end
                    end
                end
            end
        end


    case {7}
        for m = 1:j
            var = isnan(eda_data(m,2));
            if var == 1
                continue;
            else
                for k=1:sampling_rate

                    n =m+k;
                    n2 = n+sampling_rate;
                    n3 = n2+sampling_rate;
                    n4 = n3+sampling_rate;
                    n5 = n4+sampling_rate;
                    n6 = n5+sampling_rate;
                    n7 = n6+sampling_rate;

                    if (n<=j && n2<=j && n3<=j && n4<=j && n5<=j && n6<=j && n7<=j)
                        z = (eda_data(n,2)-eda_data(m,2))/(eda_data(n,1)-
eda_data(m,1));

                        if strcmp (slope_constraint,'Yes')==1 && ~isnan(z)
                            if(eda_data(n,2)> 1.2*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif (eda_data(n,2)<0.9*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif z>max_slope
                                eda_data(n,2)= NaN;
                            elseif z<min_slope
```

```
                            eda_data(n,2)= NaN;
                        elseif(eda_data(n2,2)> 1.44*eda_data(m,2))
                            eda_data(n2,2)= NaN;
                        elseif (eda_data(n2,2)<0.81*eda_data(m,2))
                            eda_data(n2,2)= NaN;
                        elseif(eda_data(n3,2)> 1.728*eda_data(m,2))
                            eda_data(n3,2)= NaN;
                        elseif(eda_data(n3,2)<0.729*eda_data(m,2))
                            eda_data(n3,2)= NaN;
                        elseif(eda_data(n4,2)> 2.0736*eda_data(m,2))
                            eda_data(n4,2)= NaN;
                        elseif (eda_data(n4,2)<0.6561*eda_data(m,2))
                            eda_data(n4,2)= NaN;
                        elseif(eda_data(n5,2)> 2.48832*eda_data(m,2))
                            eda_data(n5,2)= NaN;
                        elseif (eda_data(n5,2)<0.59049*eda_data(m,2))
                            eda_data(n5,2)= NaN;
                        elseif(eda_data(n6,2)> 2.985984*eda_data(m,2))
                            eda_data(n6,2)= NaN;
                        elseif(eda_data(n6,2)<0.531441*eda_data(m,2))
                            eda_data(n6,2)= NaN;
                        elseif(eda_data(n7,2)> 3.5831808*eda_data(m,2))
                            eda_data(n7,2)= NaN;
                        end

                    else
                        if(eda_data(n,2)> 1.2*eda_data(m,2))
                            eda_data(n,2)= NaN;
                        elseif (eda_data(n,2)<0.9*eda_data(m,2))
                            eda_data(n,2)= NaN;
                        elseif(eda_data(n2,2)> 1.44*eda_data(m,2))
                            eda_data(n2,2)= NaN;
                        elseif (eda_data(n2,2)<0.81*eda_data(m,2))
                            eda_data(n2,2)= NaN;
                        elseif(eda_data(n3,2)> 1.728*eda_data(m,2))
                            eda_data(n3,2)= NaN;
                        elseif(eda_data(n3,2)<0.729*eda_data(m,2))
                            eda_data(n3,2)= NaN;
                        elseif(eda_data(n4,2)> 2.0736*eda_data(m,2))
                            eda_data(n4,2)= NaN;
                        elseif (eda_data(n4,2)<0.6561*eda_data(m,2))
                            eda_data(n4,2)= NaN;
                        elseif(eda_data(n5,2)> 2.48832*eda_data(m,2))
                            eda_data(n5,2)= NaN;
                        elseif (eda_data(n5,2)<0.59049*eda_data(m,2))
                            eda_data(n5,2)= NaN;
                        elseif(eda_data(n6,2)> 2.985984*eda_data(m,2))
                            eda_data(n6,2)= NaN;
                        elseif(eda_data(n6,2)<0.531441*eda_data(m,2))
                            eda_data(n6,2)= NaN;
                        elseif(eda_data(n7,2)> 3.5831808*eda_data(m,2))
                            eda_data(n7,2)= NaN;
                        end
                    end
                end
            end
        end
    end


case {8}
    for m = 1:j
        var = isnan(eda_data(m,2));
        if var == 1
            continue;
        else
            for k=1:sampling_rate
```

```
                    n =m+k;
                    n2 = n+sampling_rate;
                    n3 = n2+sampling_rate;
                    n4 = n3+sampling_rate;
                    n5 = n4+sampling_rate;
                    n6 = n5+sampling_rate;
                    n7 = n6+sampling_rate;
                    n8 = n7+sampling_rate;

                    if (n<=j && n2<=j && n3<=j && n4<=j && n5<=j && n6<=j && n7<=j
&& n8<=j)
                        z = (eda_data(n,2)-eda_data(m,2))/(eda_data(n,1)-
eda_data(m,1));

                        if strcmp (slope_constraint,'Yes')==1 && ~isnan(z)
                            if(eda_data(n,2)> 1.2*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif (eda_data(n,2)<0.9*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif z>max_slope
                                eda_data(n,2)= NaN;
                            elseif z<min_slope
                                eda_data(n,2)= NaN;
                            elseif(eda_data(n2,2)> 1.44*eda_data(m,2))
                                eda_data(n2,2)= NaN;
                            elseif (eda_data(n2,2)<0.81*eda_data(m,2))
                                eda_data(n2,2)= NaN;
                            elseif(eda_data(n3,2)> 1.728*eda_data(m,2))
                                eda_data(n3,2)= NaN;
                            elseif(eda_data(n3,2)<0.729*eda_data(m,2))
                                eda_data(n3,2)= NaN;
                            elseif(eda_data(n4,2)> 2.0736*eda_data(m,2))
                                eda_data(n4,2)= NaN;
                            elseif (eda_data(n4,2)<0.6561*eda_data(m,2))
                                eda_data(n4,2)= NaN;
                            elseif(eda_data(n5,2)> 2.48832*eda_data(m,2))
                                eda_data(n5,2)= NaN;
                            elseif (eda_data(n5,2)<0.59049*eda_data(m,2))
                                eda_data(n5,2)= NaN;
                            elseif(eda_data(n6,2)> 2.985984*eda_data(m,2))
                                eda_data(n6,2)= NaN;
                            elseif(eda_data(n6,2)<0.531441*eda_data(m,2))
                                eda_data(n6,2)= NaN;
                            elseif(eda_data(n7,2)> 3.5831808*eda_data(m,2))
                                eda_data(n7,2)= NaN;
                            elseif(eda_data(n8,2)<0.43046721*eda_data(m,2))
                                eda_data(n8,2)= NaN;
                            end

                        else
                            if(eda_data(n,2)> 1.2*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif (eda_data(n,2)<0.9*eda_data(m,2))
                                eda_data(n,2)= NaN;
                            elseif(eda_data(n2,2)> 1.44*eda_data(m,2))
                                eda_data(n2,2)= NaN;
                            elseif (eda_data(n2,2)<0.81*eda_data(m,2))
                                eda_data(n2,2)= NaN;
                            elseif(eda_data(n3,2)> 1.728*eda_data(m,2))
                                eda_data(n3,2)= NaN;
                            elseif(eda_data(n3,2)<0.729*eda_data(m,2))
                                eda_data(n3,2)= NaN;
                            elseif(eda_data(n4,2)> 2.0736*eda_data(m,2))
                                eda_data(n4,2)= NaN;
                            elseif (eda_data(n4,2)<0.6561*eda_data(m,2))
                                eda_data(n4,2)= NaN;
```

```
                            elseif(eda_data(n5,2)> 2.48832*eda_data(m,2))
                                eda_data(n5,2)= NaN;
                            elseif (eda_data(n5,2)<0.59049*eda_data(m,2))
                                eda_data(n5,2)= NaN;
                            elseif(eda_data(n6,2)> 2.985984*eda_data(m,2))
                                eda_data(n6,2)= NaN;
                            elseif(eda_data(n6,2)<0.531441*eda_data(m,2))
                                eda_data(n6,2)= NaN;
                            elseif(eda_data(n7,2)> 3.5831808*eda_data(m,2))
                                eda_data(n7,2)= NaN;
                            elseif(eda_data(n8,2)<0.43046721*eda_data(m,2))
                                eda_data(n8,2)= NaN;
                            end
                        end
                    end
                end
            end
        end
end
```

Algorithm to detect loss of contact between skin and electrodes

```
for m = 1:j
    if (eda_data(m,2)<=contact_loss)
        eda_data(m,2)= NaN;
    end
end


counter = 0;
counter_array = [];
```

Algorithm to do detect missing data of more than "time_window" seconds

```
for c=1:j-1
    if (~isnan(eda_data(c,2)) && isnan(eda_data(c+1,2)))
        b = c+1;
        while (isnan(eda_data(b,2)) && b+c<j)
            if((b-c >= sampling_rate*time_window) && ~isnan(eda_data(b+1,2)))
                counter = counter+1;
                counter_array(counter,:) = [c+1,b];
            end
            b = b+1;
        end
    end
end


if isempty(counter_array==1)
    eda_filtered1 =fillmissing(eda_data(:,2),'makima');

else

    switch linear_value    % check whether fill missing data of
morecounter_array(i-1,2)+1:counter_array(i,1)-1
        % than time-window seconds with linear
        % interpolation (Yes) or not (No).
```

```
        case 'Yes'

            for i=1:size(counter_array,1)

                if (size(counter_array,1) == 1)
                    eda_filtered1(1:counter_array(i,1)-1) =
fillmissing(eda_data(1:counter_array(i,1)-1,2),'makima');
                    eda_filtered1(counter_array(i,1):counter_array(i,2)) =
fillmissing(eda_data(counter_array(i,1):counter_array(i,2),2),'linear','EndValues',
'nearest');
                    eda_filtered1(counter_array(i,2)+1:j) =
fillmissing(eda_data(counter_array(i,2)+1:j,2),'makima');


                else

                    if(i==1)
                        eda_filtered1(1:counter_array(i,1)-1) =
fillmissing(eda_data(1:counter_array(i,1)-1,2),'makima');
                        eda_filtered1(counter_array(i,1)-1:counter_array(i,2)+1) =
fillmissing(eda_data(counter_array(i,1)-
1:counter_array(i,2)+1,2),'linear','EndValues','nearest');


                    elseif (i~=1 && i~=size(counter_array,1))

                        eda_filtered1(counter_array(i-1,2)+1:counter_array(i,1)-1)
= fillmissing(eda_data(counter_array(i-1,2)+1:counter_array(i,1)-1,2),'makima');
                        eda_filtered1(counter_array(i,1)-1:counter_array(i,2)+1) =
fillmissing(eda_data(counter_array(i,1)-
1:counter_array(i,2)+1,2),'linear','EndValues','nearest');


                    elseif (i==size(counter_array,1))

                        eda_filtered1(counter_array(i,1)-1:counter_array(i,2)+1) =
fillmissing(eda_data(counter_array(i,1)-
1:counter_array(i,2)+1,2),'linear','EndValues','nearest');
                        eda_filtered1(counter_array(i,2)+1:j) =
fillmissing(eda_data(counter_array(i,2)+1:j,2),'makima');


                    end

                end

            end

        case 'No'

            for i=1:size(counter_array,1)

                if (size(counter_array,1) == 1)
                    eda_filtered1(1:counter_array(i,1)-1) =
fillmissing(eda_data(1:counter_array(i,1)-1,2),'makima');
                    eda_filtered1(counter_array(i,1):counter_array(i,2)) = NaN;
                    eda_filtered1(counter_array(i,2)+1:j) =
fillmissing(eda_data(counter_array(i,2)+1:j,2),'makima');

                else

                    if(i==1)
                        eda_filtered1(1:counter_array(i,1)-1) =
fillmissing(eda_data(1:counter_array(i,1)-1,2),'makima');
                        eda_filtered1(counter_array(i,1):counter_array(i,2)) = NaN;
```

```
                    elseif (i~=1 && i~=size(counter_array,1))

                        eda_filtered1(counter_array(i-1,2)+1:counter_array(i,1)-1)
= fillmissing(eda_data(counter_array(i-1,2)+1:counter_array(i,1)-1,2),'makima');
                        eda_filtered1(counter_array(i,1):counter_array(i,2)) = NaN;


                    elseif (i==size(counter_array,1))

                        eda_filtered1(counter_array(i,1):counter_array(i,2)) = NaN;
                        eda_filtered1(counter_array(i,2)+1:j) =
fillmissing(eda_data(counter_array(i,2)+1:j,2),'makima');


                    end
                end
            end
    end
end

eda_filtered(:,1) = eda_data(:,1);
eda_filtered(:,2) = eda_data(:,2);
eda_filtered(:,3) = eda_filtered1';
end
```

Published with MATLAB® R2021b

APPENDIX B

Other related work

I was awarded the UKRI – MITACs research award which funded my placement in the Human Factors and Statistics Lab (HFASt), at the school of Mechanical and Industrial Engineering (MIE), University of Toronto. For this project, I investigated the use of sensor-fusion of physiological sensors and eye-tracking, to develop machine learning models that can predict drivers' future takeover performance, for if and when they resume manual control of the vehicle. The outcome of this research is currently in preparation for submission as a journal manuscript by August 2022.