



In a heart beat: Using driver's physiological changes to determine the quality of a takeover in highly automated vehicles



Mohamed Taher Alrefaie, Stever Summerskill, Thomas W Jackon

Loughborough University, Epinal Way, Loughborough, United Kingdom

ARTICLE INFO

Keywords:

Physiological data
Highly automated driving
Take-over
Driver performance
Heart rate
Pupil diameter
Autonomous vehicle
Accident prevention

ABSTRACT

Developing conditionally automated driving systems is on the rise. Vehicles with full longitudinal and latitudinal control will allow drivers to engage in secondary tasks without monitoring the roadway, but users may be required to resume vehicle control to handle critical hazards. The loss of driver's situational awareness increases the potential for accidents. Thus, the automated systems need to estimate the driver's ability to resume control of the driving task.

The aim of this study was to assess the physiological behaviour (heart rate and pupil diameter) of drivers. The assessment was performed during two naturalistic secondary tasks. The tasks were the email and the twenty questions task in addition to a control group that did not perform any tasks. The study aimed at finding possible correlations between the driver's physiological data and their responses to a takeover request. A driving simulator study was used to collect data from a total of 33 participants in a repeated measures design to examine the physiological changes during driving and to measure their takeover quality and response time.

Secondary tasks induced changes on physiological measures and a small influence on response time. However, there was a strong observed correlation between the physiological measures and response time. Takeover quality in this study was assessed using two new performance measures called PerSpeed and PerAngle. They are identified as the mean percentage change of vehicle's speed and heading angle starting from a take-over request time. Using linear mixed models, there was a strong interaction between task, heart rate and pupil diameter and PerSpeed, PerAngle and response time. This, in turn, provided a measurable understanding of a driver's future responses to the automated system based on the driver's physiological changes to allow better decision making. The present findings of this study emphasised the possibility of building a driver mental state model and prediction system to determine the quality of the driver's responses in a highly automated vehicle. Such results will reduce accidents and enhance the driver's experience in highly automated vehicles.

1. Introduction

In the past few decades, vehicle automation has gained substantial traction in both industry and automotive research (e.g., Department of Transport, 2015; Meinschmidt et al., 2018). Several manufacturers introduced simple automation systems to handle several driving tasks. As these systems evolve, different levels of automation have been characterised based on the system's ability to intervene in longitudinal and lateral control of the vehicle (Flemisch et al., 2011; SAE International, 2018).

Level 3 of automation, identified as the Conditional Automation System, will have full control over all safety-critical functions with occasional cases for the driver to intervene (SAE International, 2018). While full automation is in action, drivers may direct their attention away from driving to engage in non-driving related (NDR) tasks (SAE

International, 2018). Nevertheless, Level-3 systems are limited and will require the driver to re-engage immediately to handle a critical latent hazard (SAE International, 2018). This mandatory transition could be due to either a sensory or decision-making limitations (NHTSA, 2013).

The theoretical background of this paper was based on the following. In summary, there are several factors affecting driver's responses during an emergency takeover in highly automated driving studies, as reviewed in Sections 2.1 and 2.2. Though, a very limited research has been given to the driver's physiological data in highly automated driving environment, as reviewed in Section 2.3. This research gap was the main motivation of this research and has been covered in the Results and Discussion sections. Moreover, the study identified a dispersion in the literature regarding the performance measures used to assess the driver's ability to manoeuvre the vehicle, as reviewed in Section 2.4. Thus, the study introduced new performance

E-mail address: m.t.a.alrefaie@lboro.ac.uk (M.T. Alrefaie).

<https://doi.org/10.1016/j.aap.2019.06.011>

Received 7 December 2018; Received in revised form 11 June 2019; Accepted 13 June 2019

Available online 11 July 2019

0001-4575/ © 2019 Elsevier Ltd. All rights reserved.

measures assessing driver's performance in handling the vehicle. Finally, the study examined the validity of the driver's physiological data as valid predictors to driver's responses. More details are laid out in the following sections.

1.1. Driver's take-over

Level-3 automated driving systems have to identify its limitations using system boundaries (SAE International, 2018). For example, the system may not be able to handle construction sites or heavy weather conditions due to sensory limitations. When a system boundary is detected, a take-over request is issued to the driver to take over the vehicle's control. The take-over request (TOR) will have to be prompted in a timely manner allowing the driver to perform a safe transition before a potential collision is expected (NHTSA, 2013).

The time given to drivers before they respond has been extensively studied in the literature. Several studies reported seven (Gold et al., 2013), eight (Wandtner et al., 2018), ten (Melcher et al., 2015) and 12 s (Zeeb et al., 2015) as a safe time budget to be given to drivers to respond. Gold et al.'s (2013) study concluded that a 7 s budget is adequate for drivers to respond to critical hazards in highly automated driving. Since then, Gold et al. (2013) study was recommended by NHTSA for manufacturers to design their automated system's transitions (Campbell et al., 2018).

The takeover time is spent by drivers in regaining physical and mental control over the driving task (Zeeb et al., 2015). Motor readiness is identified as the time it takes a driver to regain mechanical control i.e., hands-on wheel and feet on pedals (Zeeb et al., 2015) while response time is the time between a TOR and the driver applying a significant change on braking or steering wheel (Zeeb et al., 2015). Several studies showed that motor readiness is consistent, ranging from 1.2 to 1.8 s (Gold et al., 2013; Zeeb et al., 2015). However, response time varied among drivers based on several factors which are discussed in the next subsection.

1.2. Factors Affecting Drivers Responses

Driver's response time and quality are reviewed in this section. To start, the response time differed significantly among studies (e.g., Gold et al., 2016, 2013; Zeeb et al., 2015). Literature showed that response time was influenced by NDR tasks (Merat et al., 2012), drivers background measures which are the factors unique to each driver such as age (Körber et al., 2016), years of driving experience (Wright et al., 2016b), and road conditions such as traffic situation (Gold et al., 2016; Radlmayr et al., 2014). In addition, it was reported that the take-over time budget influenced driver's response time. For example, when given a longer time budget, drivers had a longer response time (Gold et al., 2013). An explanation to this could be due to drivers investing time in restoring situational awareness before taking an action. Hence, drivers who spent a longer time had a safer response as measured by the objective measures in comparison to the group who were given a shorter time to respond (Gold et al., 2013; Radlmayr et al., 2014; Zeeb et al., 2016).

Similarly, response quality is impacted by the NDR tasks due to their visual or visuo-cognitive distraction from the driving environment (Zeeb et al., 2016). To respond to a take-over situation, drivers require a cognitive processing time to restore situational awareness and then respond accordingly (Endsley, 1995; Endsley et al., 1997).

In automated driving literature, studies investigated secondary tasks that induced cognitive workload such as n-back task (Radlmayr et al., 2014), Twenty Questions Task (Merat et al., 2012), visuo-cognitive tasks such as reading news (Zeeb et al., 2016), internet search (Zeeb et al., 2016), vehicle's multimedia systems (Zeeb et al., 2015) and IQ questions (Louw et al., 2016). Gold et al. (2015a, b) reported a significant decrease in performance for tasks including manual versus cognitive workload.

Such results concur with Petermann-Stock et al. (2013) findings that the worst performance decrements were caused by quizzes requiring a combined visual, cognitive and manual workload in comparison to quizzes requiring one or two of those workloads. Such findings contradict Gold et al.'s (2015b) findings that reported cognitive and manual tasks had the same detrimental effect. The contradiction among those results provides the necessary motivation to study the physiological differences caused by cognitive and visuo-cognitive tasks in order to provide a better understanding of the differences among those studies.

1.3. Physiological analysis of drivers in highly automated driving environment

Many studies examined the influence of NDR tasks on gaze related measures during highly automated and manual driving (Marquart et al., 2015). In previous studies, takeover time and performance have been correlated with eye blinking (Merat et al., 2012), gaze behaviour (Gold et al., 2016; Ko and Ji, 2018; Louw et al., 2016; Wright et al., 2016a; Zeeb et al., 2016, 2015), eye movements and PERCLOS (Jamson et al., 2013). Two manual driving studies induced mental workload using verbal and spatial imagery NDR tasks (Recarte and Nunes, 2003, 2000). Their results indicated that the used NDR tasks caused pupil dilation which indicated a higher mental workload. Finally, the literature had limited to no studies that investigated pupil diameter changes in a highly automated driving environment, to the best of our knowledge.

However, few studies examined the effect of NDR tasks on heart rate (Carsten et al., 2012; de Waard et al., 1999; Wille et al., 2007). The heart rate measured in highly automated driving is lower than manual driving and ACC driving (Carsten et al., 2012) which matches with findings of de Waard et al. (1999) that reported a slight decrease (73.2 vs 74.0 beats/min) in driver's heart rate during automated driving. This difference is an indication of mental workload reduction (De Winter et al., 2014). In a manual driving study, heart rate increased incrementally with increasing mental workload, and a plateau of the physiological measures was observed (Mehler et al., 2009). Moreover, the study reports a decrease in driving performance as the mental workload increases. Additionally, wearable technologies at a consumer level achieved a reasonable accuracy in detecting the cognitive workload using collected physiological data of the driver (i.e., heart rate). In the study by Mehler et al (2009) the experiment took place on a simulator with 14 participants driving in rural, motorways and other roads and with a simulated accident. Results showed reliability in detecting cognitive workload based on heart rates (Melnicuk et al., 2016).

1.4. Performance measures of driver's responses

Highly automated driving studies have used several performance measures for vehicle handling during take-over scenarios (Radlmayr et al., 2019). Several studies identified time to collision (Gold et al., 2016; Körber et al., 2016; Radlmayr et al., 2014), longitudinal acceleration (Gold et al., 2016; Radlmayr et al., 2014), braking (Körber et al., 2016; Larsson et al., 2014), minimum speed (Larsson et al., 2014) and occurred collisions (Radlmayr et al., 2014; Wandtner et al., 2018). Such scarcity makes it challenging to provide a cross-comparison among studies. Therefore Radlmayr et al. (2019) reported new take-over performance measures named TOPS that aggregates vehicle, mental and subjective ratings of the take-over to provide a single metric assessing the takeover. Though, the study has not provided any correlation between driver's physiological changes and the TOPS results for each participant.

While minimum time to collision is a popular performance measure in driving studies generally (Gold et al., 2016; Körber et al., 2016; Radlmayr et al., 2014), the performance measure has some flaws in highly automated driving scenarios. The main flaw of min-TTC is its tendency to reward hard braking which is confirmed by Radlmayr et al.

(2019) who stated that “drivers [...] staying longer in the ego lane (lower TTC’s) tend to brake harder before executing the lane change”. Finally, the min-TTC ignores the changes in speed or steering angle of the vehicle even though they are considered poor performance indicators in motorway driving scenarios (Kass et al., 2007; Zeeb et al., 2016); especially in this study’s scenario design, please see the Methodology section.

The poor depth of min-TTC motivated the authors to identify new performance measures that could quantify a smooth take-over manoeuvre. Starting off, no rapid change in speed or heading angle was required in the design of the scenario because it allowed participants to perform a smooth transition from their lane to the next lane. Therefore, hard braking or steering were penalised in this study because it correlates with a lack of situational awareness according to Kass et al. (2007). Thus, the study examined new performance measures based on the changes of steering and braking changes to assess the driver’s performance. More details are laid out in Section 3.7.

1.5. The purpose of the study

The main purpose of the study reported here was to examine physiological changes caused by cognitive and visuo-cognitive secondary tasks and how they influence response time and quality during take-over scenarios. Reading and responding to an email and twenty questions tasks are known to degrade driver’s response time and quality (Merat et al., 2012; Zeeb et al., 2016). Thus, the study investigated the influence of secondary tasks on the driver’s physiological behaviour. Furthermore, the study was designed to examine whether the reported learning curve (Körber et al., 2016) of take-over handling could have any effect on drivers physiological changes before and during the take-over, how it correlates with response time and quality after a TOR. Specifically, the literature used objective and subjective measures to quantify the quality of the takeover. Few studies identified a convergence between reported subjective and objective measures (Zeeb et al., 2016). Thus, this study was designed to assess whether physiological changes at TOR could provide more information to explain the aforementioned convergence. A further aim was to propose new performance measures to assess the quality of drivers handling vehicles.

The remainder of this paper is organised as follows. The Methodology section discusses the experiment’s procedure including participants, driving simulator, secondary tasks and driving scenario used in this study. Results section lays out the data analysis approach and provides the results of this study which is then discussed in the Discussion section. The future work and limitations are provided in Future Work section. Finally, conclusions are drawn in the Conclusion chapter.

2. Methodology

2.1. Participants

Data represented in this study were collected in an experiment run at Loughborough University Design School. There was a total of thirty-six participants recruited for this study (53% females) and between 20 and 30 years of age ($M = 25.8$, $SD = 5.7$). Participants were invited to the lab for 90 min. Participants were asked to fill in a demographics survey and a questionnaire regarding their driving experience.

Drivers had a minimum of two years of driving experience ($M = 7.01$, $SD = 3.6$), 95% of them driving once a day with an average reported mileage of 7340 a year ($SD = 5350$). In addition, they were required to have a normal or corrected-to-normal vision. Of all participants, 84.6% had no experience with adaptive cruise control. An informed consent was obtained after the experimenter explained the procedure which was approved by the university ethics panel. All experimental procedures were conducted in accordance with the ethical guidelines of the hosting university. Three participants were excluded



Fig. 1. Photo of the driving simulator during an experiment.

from the study. Two of these were due to some missing data during the data collection process and another participant who informed the researcher they were profoundly fatigued during the experiment.

2.2. Experiment equipment

The study used the STISIM driving simulator (STISIM, 2018). The simulator provided 135° with graphics projection serving as a test environment. The rig consists of an SUV seat and steering wheel with automatic transmission, see Fig. 1. The cockpit included a tablet for multimedia use. On the right side of the cabin, a camera was placed to record the participant’s posture and behaviour.

To communicate the transition from normal to automated driving, a visuo-auditory message was played informing drivers that the system is taking over control when a certain point in the simulated environment is reached. When the ego vehicles get close to a predefined hazard for the scenario, an intermittent beep (based on NHTSA guidelines (Campbell et al., 2007)) was played to instruct the driver to takeover. When automated driving is activated, the vehicle speed is set to 70mph and is placed to provide a seven seconds gap from the leading vehicle. A seven-second gap was chosen based on the study by (Gold et al., 2013). The automated system provided lateral and longitudinal control with no overtaking manoeuvres, changing lanes or changing speed. This was done to reduce the number of dependent variables and ensure drivers will engage in secondary tasks. Finally, the driving simulator collected any changes to steering wheels and pedals which were the variables used to calculate response time, PerSpeed and PerAngle.

Before the training phase starts, participants were required to put on an eye-tracking system (Tobii Pro Glasses) to track their eye movements and heart rate monitor (Polar H7 chest strap). The eye tracking system calculated the pupil diameter and blinking rate of both eyes. The heart rate monitor has a 99% accuracy in comparison to ECG devices (Wang et al., 2016).

2.3. Non-driving related (NDR) tasks

To understand how visuo-cognitive and cognitive distractions may affect physiological behaviour of drivers, non-driving related tasks were selected based upon two previous studies (Merat et al., 2012; Zeeb et al., 2015); namely email and twenty questions tasks. Also, participants were requested to pay attention to the road without engaging in any tasks (control group). The sequence of those tasks was picked randomly for each participant to alleviate the order effect.

The email task included reading an email on a tablet on the vehicle’s dashboard then writing a reply. The emailer asked participants to pick a close friend and describe their perfect birthday party. The Twenty Questions Task (TQT) was chosen because it causes a cognitive distraction and has been selected in similar studies (Gold et al., 2016; Merat et al., 2012). During the TQT, participants were asked to guess an animal by asking the researcher a maximum of 20 polar questions via a simulated hands-free phone call (Jamson et al., 2004). These two tasks were designed to ensure participants engagement until the takeover

request (similar to (Zeeb et al., 2015)) to maintain the same effect on mental workload among all participants. The NDR task engagement lasted for 10 min which is in line with Zeeb's et al., (2016) to ensure driver's full involvement in the task.

2.4. Training procedure

At the start of the participant's visit, they were given an information sheet explaining the experiment, what to expect, how data is stored and manipulated and the importance of the experiment. The information sheets included an explanation of how the system works, its limitations and their role in the vehicle. Participants were informed orally of how the experiment was run and encouraged to ask questions. They were also informed they could stop at any point. After answering all their questions, participants were asked to sign an informed consent and fill in the demographics survey. Before showing them the simulator, participants were asked of their familiarity with Android or iOS. Based on that answer, the multimedia tablet placed in the vehicle rig was chosen accordingly. This was done to alleviate any learning curve of handling the tablet during the experiment.

Participants were given manual and highly automated driving training that included a TOR. They started with 20 min of manual driving to get them familiar with the driving simulator in manual driving. Then, they were given 15 min of highly automated driving practice that involved the activation of the automated system and the handling of a critical incident.

2.5. The experiment procedure

The main experiment was 40 min long. It included three takeovers and three slots for secondary tasks with no stop in between. The main scenario of this experiment was a repetition of the same test scenario where participants started driving manually and then placed the vehicle on automated mode at the second kilometre. This was done by asking the drivers to place the ego-vehicle in a specific lane. Lane choice was alternated every time to cover the three lanes for every participant; this was done to reduce expectancy effects and to consider takeover behavioural changes in different lanes. After a minute, drivers were asked via an audio message to engage in one of the secondary tasks or pay attention to the road.

After 10 min of NDR tasks engagement, an intermittent beep was sounded accompanied by a visual cue to a signal a takeover request. Participants were expected to stop the NDR task and engage immediately in handling the vehicle. This entire scenario is then repeated twice to cover a total of three takeovers three slots for secondary tasks. The main difference among these phases are 1) the secondary task; participants are expected to do a different secondary task in each phase and 2) the lane; the vehicle changes lane right after the automation starts. This repeated measure approach allowed more results for each participant and provided data to study the learning effect of drivers and how it correlates with their physiological measures.

2.6. Data analysis procedure

Pupil diameter (PD) and heart rate (HR) data were normalised according to Eq. (1) where V is a window of the collected data ending at each takeover request. A window is a number of recordings that span over a period of time, e.g., a window size of 30 s includes 30 recordings of HR data over a period of 30 s at a sampling rate of 1 Hz or 300 recordings of pupil diameter at 10 Hz. Normalisation was performed to alleviate the individuality of the collected data by putting minimum HR or PD value at 0.0 and a maximum value at 1.0 per each participant for the entire duration of the experiment.

Out of six participants, 21 HR readings were missing out of an average of 2000 readings. To resolve those missing points, linear interpolation was applied to fill in those gaps. When calculating the mean

value of HR or PD, several window sizes were tested to assess their effect on the accuracy of the statistical models.

$$\alpha = \frac{1}{n} * \sum_{i=k}^n \frac{V_i - \min(V)}{\max(V) - \min(V)} \quad (1)$$

Normalisation of heart rate and pupil diameters data were performed using the method, where V is a vector containing readings from k is time0 and n is window size.

2.7. Performance measures of the study

To assess the performance of driver's handling the vehicle, response time, PerSpeed and PerAngle were used. Response Time analysis examined the driver's readiness to respond but not necessarily the quality of drivers' responses. It is collected using the driver's simulator data and was defined in section 0. Response time is not the only measure of performance in takeover situations; other studies take the quality of takeover as another important performance measure (Gold et al., 2016; Radlmayr et al., 2014; Zeeb et al., 2016).

To quantify the driver's manoeuvre, PerSpeed and PerAngle measures were introduced in this study. PerSpeed is the mean percentage change of vehicle's speed for a period before the TOR (e.g., 30 s), see Eq. (3). A higher percentage indicates a sharper change in speed which could be either braking or acceleration. PerAngle, similarly, is the mean percentage change of vehicle's heading angle. The usage of the accumulation of the mean percentage change provides a comprehensive judgement of the driver's manoeuvre; unlike using maximum or minimum values which focus on a specific moment in time. In addition, the standard deviation was rejected because it provides a reading of the overall deviation from the mean steering or braking whilst mean percentage change provides an accumulative reading of the steering or braking changes made by the driver throughout the transition.

$$y = \frac{100}{n} \times \sum_{i=k}^n \frac{x_i - x_{i-1}}{|x_{i-1}|} \quad (2)$$

Mean percentage change formula, where y is PerSpeed or PerAngle, x is a vector containing readings (speed or heading angle values), k is time₀, and n is window size.

3. Results

3.1. Physiological behaviour at the start of the scenario

At the beginning of the experiment, a peak in normalised HR and PD were observed in all participants. This was probably due to the stress caused by the anticipation of the experiment. After three seconds of manual driving, it dropped by an average of 20% ($M = .6$, $SD = 0.15$), see Fig. 3, and stabilised ($M = .29$, $SD = .14$) as the vehicle switched to automation. To simplify the plot, the drop in heart rate after the start is not plotted in Fig. 3.

3.2. Effect of the NDR tasks on physiological behaviour

3.2.1. Pupil diameter and secondary task

Repeated measures ANOVA analysis among the three groups of secondary tasks was conducted to explore their impact on the right pupil diameter. The data were collected whilst drivers were performing a secondary task prior to a takeover request. There was a statistically significant difference at $F(1.22, 21.9) = 60.741$, $p < .0001$ with Greenhouse-Geisser correction. Post-hoc tests using Bonferroni correction demonstrated an increase in pupil diameter by an average of 0.314 during email task ($p < 0.0001$) and 0.06 during TQT ($p = .02$) in comparison to the control group. A five-second window length demonstrated the highest p values of repeated measures ANOVA test on periods from 1 to 150 s. The overall mean values of PD of all

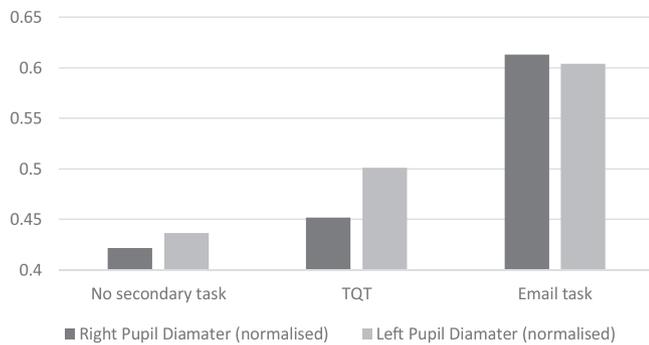


Fig. 2. the values of driver’s pupil diameters under the conditions of different non-driving related tasks.

participants (after normalisation) are plotted in Fig. 2.

3.2.2. Heart rate and secondary tasks

The relationship between normalised heart rate and response time was investigated using repeated measures ANOVA to investigate whether there was a significant difference among different groups. Normalised HR, with an average of 90 s window, had the strongest significance, $F(2,36) = 7.75, p < .02$, among window sizes. Pairwise comparison using Bonferroni correction showed a definite difference between the control group and the and email task ($p = .03$) and the control group and the TQT task ($p = .013$). Differences showed that heart rate increases significantly when drivers are engaged in secondary tasks which align with the results of Carsten et al. (2012). No statistical difference was reported between TQT and email tasks ($p = .95$). Therefore, normalised HR could be considered a valid physiological measure to identify engagement in secondary tasks; however, it cannot distinguish secondary task type for the tasks used in the study reported here.

3.3. Effect of the handover process on the driver’s physiological behaviour

Takeover requests sparked a peak in HR ($M = .43, SD = .2$) that lasted for few seconds, see Fig. 3. Then, HR gradually drops to the mean of the control group ($M = .36, SD = .14$) among the three groups within a mean of 10 s ($SD = 6.2$) of a successful takeover. In takeovers that ended with an accident, the significant increase of HR remained for a longer period ($M = 20, SD = 6.2$).

Fig. 3 plots a limited view of the heart rate changes across the experiment. The mean value of the heart rate is limited and doesn’t allow a better understanding of the fluctuations of the heart rate. Thus, the study used the Markov Transition Field’s visualisation to plot the heart rate as seen in Fig. 4.

The Markov Transition Fields (MTF) is one of the recent approaches to encode time series to an image. MTF images “represent the first order Markov transition probability along one dimension and temporal dependency along the other”, (Wang and Oates, 2014). MTF images were

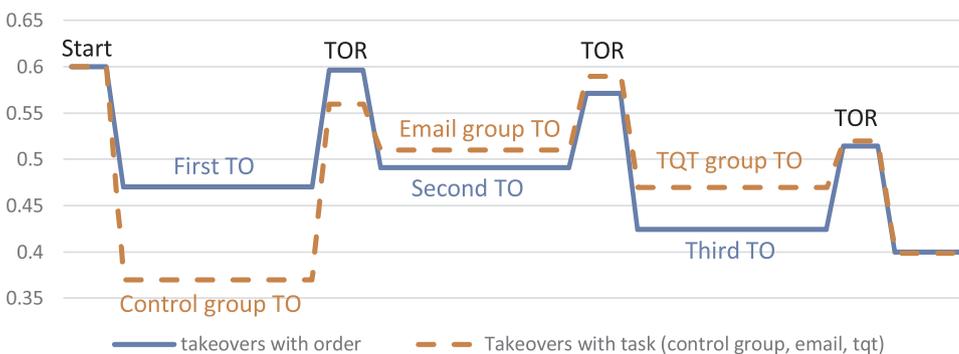


Fig. 3. Average normalised HR fluctuations throughout the scenario of the experiment based on 60-second window average. In blue, are the average heart rate values of participants grouped by order of the takeover. In orange (dashes), are the average heart rate values of participants grouped by the secondary task type they performed. Standard deviation of reported variables was between 0.13 and 0.14 among all groups; therefore, was omitted from the graph for simplicity (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

used to visualise HR signals and their temporal dependencies in an image. So, the more squares and more colours in an image, the more transitions and temporal dependencies are among the signals. It is essential to understand that MTF provides a subjective comparison between signals and is used in this study for this purpose.

3.3.1. Effect of the order of the NDR tasks on driver’s heart rate

Design of repeated measure studies could cause a severe ambiguity and bias when learning or practice effects are not taken into account; especially when observations of learning effects are reported in the handover process studies (Larsson et al., 2014). The repeated measure approach was used in this study to assess whether the reported learning curve may affect the physiological behaviour of drivers; specifically, their heart rate changes.

Consequently, one-way analysis between takeover groups based on the analysis of variance was performed to assess the correlation between heart rate and the order of takeover requests. Average heart rate was calculated of all participants grouped by the order in which a takeover is performed. For example, the second takeover group means the values collected at the second takeover among all participants regardless of the secondary task type they performed.

Results showed that the second takeover group had a higher HR mean ($M = .49, SD = .13$) than the first ($M = .462, SD = 0.13$) and third ($M = 0.42, SD = 0.14$) groups. There was a slight difference, $F = 3.1, p < .05$, even though the difference in mean (as identified earlier) between the three groups was quite small. Therefore, the hypothesis that HR could decrease over time because of drivers experience in handling the takeover was rejected. Accordingly, the study suggests that the order of the secondary tasks has no significant influence on the physiological changes happening during the secondary task engagement.

3.3.2. Response time estimation using driver’s physiological data

Response time of drivers was collected on the three different takeovers of drivers. A histogram of the response time is plotted on Fig. 5. The next sections assessed the correlation between the response time and the driver’s physiological data.

HR and PD were analysed using linear mixed models. LLMs assessed their correlation with response time. Since LMMs assumes no correlation among their covariate variables, left pupil diameter was excluded from this analysis due to its strong dependence with right pupil diameter, $r = .98, p = .002$.

On the main effects (at window size = 30 s), right pupil diameter was significant, $F(1, 30.5) = 12.2, p = .001$, and HR was not significant $F(1, 27.2) = 3.8, p = .06$. When using secondary task type as a fixed effect in the model, HR demonstrated a strong significance, $F(1, 30.3) = 11, p = .002$. This means HR window length may not have been long enough to provide enough significance among groups. Moreover, HR and right PD had a strong interaction term, $F(1, 24.7) = 4.2, p = .049$. Table 1 shows the interaction effect among variables and intercepts value of the analysis. This confirms that

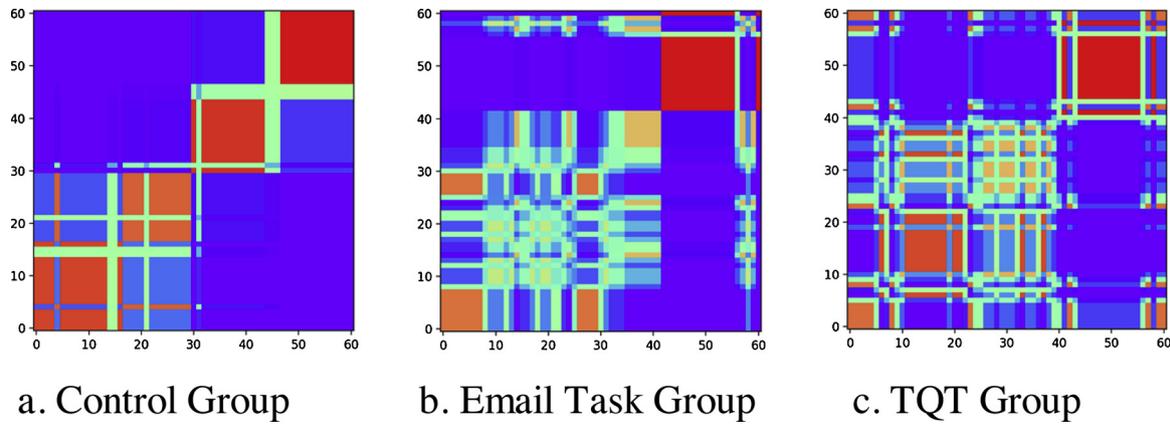


Fig. 4. Markov Transition Field of heart rate of drivers' groups. Colours and squares represent a visual approach to subjectively visualise the shapes and transitions of the signal. A clear difference is seen between the control group (a) and secondary task groups (b and c). Group B and group C show a slight difference among their signals. Those subjective observations match with repeated measures ANOVA results reported earlier in the study.

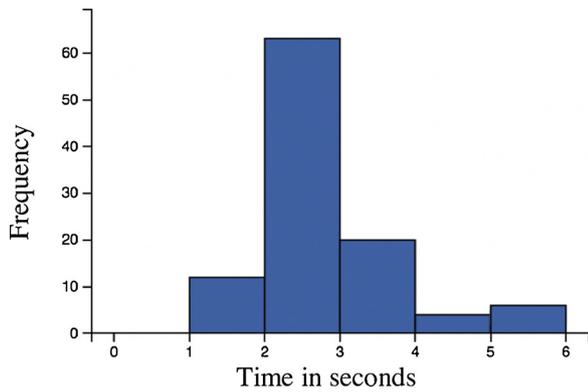


Fig. 5. A histogram of the response time values.

Table 1
Estimates of Fixed Effects on response time using 30-sec window size.

Parameter	Estimate	Std. Error	Sig.
Intercept	1.100432	.361619	.005
hr	1.077343	.552368	.062
pdr	6.786712	1.935462	.001
hr × pdr	-4.645041	2.247214	.049

normalised HR has a predictability potential for response time assuming the right window size is chosen.

In order to understand if window size influences the analysis, the same statistical methods were applied to the data extracted from a 60 s window size. HR was statistically significant $F(1,24.3) = 6.2, p = .019$ with no secondary tasks type added as a fixed effect. When added, significance increased, $p < .0001$. Reversely, right pupil diameter was not significant, $F(1, 19.2) = 1, p = .317$. This indicated that window size has a strong effect on the correlation between physiological changes and response time since pupil diameter was significant at 5-second window size.

In order to find the optimal window size in which heart rate and pupil diameters performed at, a simple optimisation algorithm was run on window sizes of 1–120 seconds. Results, as indicated above, showed that a 30 s window was the best performing window size for PD correlation with response time and 60 s for heart rate. This could be correlated to HR responding slower than PD to external changes. Hence, a longer HR window captures long-term physiological changes, and PD captures short-term physiological changes; hence window size values reacted accordingly. Those findings align with Solovey et al. (2014) that reported that the best window size for their physiological data was 30 s.

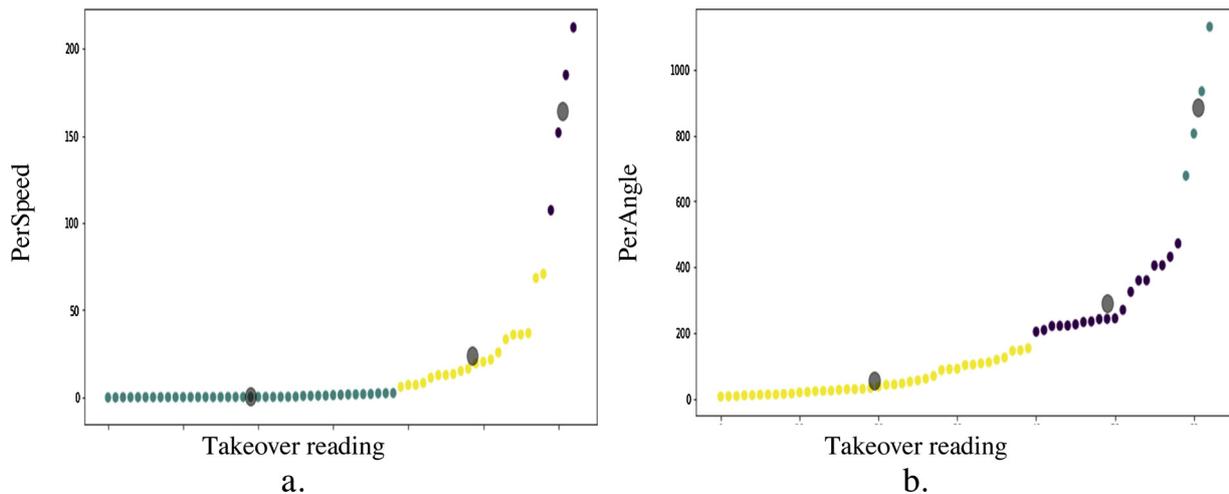


Fig. 6. Mean Percentage Change of a.) Speeding and b.) Heading angle of the vehicle. Participants are split independently into three clusters representing low, medium and high-risk groups. Black circles represent the centroid of each cluster.

Table 2
Definition of clusters of PerSpeed and PerAngle.

	PerSpeed				PerAngle			
	Start Range (%)	End Range (%)	HR mean	HR std.	Start Range (%)	End Range (%)	PDR mean	PDR std.
Group 1	0	20	0.47	0.13	0	195	0.5	0.2
Group 2	> 20	40	0.56	0.10	> 195	453	0.7	0.08
Group 3	> 40	220	0.61	0.11	> 453	1200	0.57	0.19

3.4. Quality of the takeover

3.4.1. PerSpeed and PerAngle objective measures

PerSpeed and PerAngle at each takeover are plotted in Fig. 6 by sorting takeover incidents by their corresponding PerSpeed (a) and PerAngle (b) values. The K-Means clustering algorithm (Lloyd, 1982) was used to cluster PerSpeed and PerAngle readings independently into three groups. Since Zeeb et al.'s, (2016) study split participants into three groups based on the quality of their response, this study followed the same approach by defining clusters 1, 2 and 3 as low, medium and high-risk groups. Clusters start/end ranges identified by the K-Means algorithms, and their corresponding mean HR and PD are defined in Table 2. The table includes start and end ranges of both HR and PD of each cluster, their mean and standard deviation values. As indicated in Fig. 6, the high-risk group had four incidents; three of them ended with an accident, and the fourth one was an anomaly where a participant decided to stop the vehicle for six seconds before deciding to move and change lane.

$$y = \frac{100}{n} \times \sum_{i=k}^n \frac{x_i - x_{i-1}}{|x_{i-1}|} \tag{3}$$

Mean percentage change formula, where y is PerSpeed or PerAngle, x is a vector containing readings (speed or heading angle values), k is time₀, and n is window size.

Statistical analysis of K-Means groups based on their corresponding HR/PD data yielded interesting results. A one-way ANOVA showed that the mean normalised HR had a significant difference, F = 4.2, p = .01, among PerSpeed groups. These findings indicate that higher HR means a higher probability of strong braking; which is considered a bad performance measure. No significant difference was reported based on drivers' PD mean values.

When analysing PerAngle K-Means groups, no correlation, F = 2.4, p = 0.09 was identified between PD and mean of low (m = .5, SD = .2), medium (m = .7, SD = .08) and high (m = .57, SD = .19) risk groups. Though, higher risk groups had higher pupil dilation than the low one. Based on (Batmaz and Ozturk, 2008) findings, pupil diameter dilates with the mental workload which explains the increase of PD mean value from low to high-risk groups. However, the medium risk group had a significantly higher mean than the high risk one.

3.4.2. Correlation of PerSpeed and PerAngle with driver's physiological measures

To measure whether the study's independent variables (task type, HR and PD) have any influence on PerAngle and PerSpeed, a linear mixed model using Toeplitz covariance type with repeated measures test was performed. The task was used as a fixed effect; HR and PD were used as a covariate to understand whether they significantly impacted the predefined quality measures.

For PerSpeed performance metric, task type was significant F(2, 43) = 4.3, p = .019, as was HR, F(1,44) = 5.5, p = .01, and PD, F(1,46) = 2.5, p = .01. All other higher-order interactions were significant, specifically PD × HR, F(1,46) = 8.2, p = .006 that had the highest significance. Bonferroni test showed no significant differences in PerSpeed values among task groups. Additionally, estimates of fixed effects demonstrated that each one per cent increase in HR corresponds to 4.6% decrease in PerSpeed (p = .002) and for each one per cent increase in PD corresponds to 9.1% decrease in PerSpeed (p = .004).

Similarly, for the PerAngle metric, the secondary task type, F(2,18), p = .001, was significant, so was PD, F(1,26) = 4.4, p = .04, and HR, F(1,26) = 5.1, p = .003. Higher level interactions were not more significant, and Bonferroni pairwise comparison showed no difference among task groups. Estimates of fixed effects indicated that one per cent change in HR corresponds to approximately 54% change in PerAngle and for each one per cent in PD, 71% change in expected. Such results indicate that physiological measures are valid predictors for PerSpeed and PerAngle performance measure.

3.4.3. Subjective measures of driver's performance

Drivers reported the difficulty of each takeover as seen in Fig. 7 based on their order. In addition, drivers reported that the email task was ranked more difficult than the TQT.

4. Discussion

The study identified normalised mean HR and PD as valid physiological measures to predict response time. The study introduced PerSpeed and PerAngle as new measures of takeover quality assessment. The findings identified a correlation between normalised mean HR/PD and PerAngle/PerSpeed. In addition, window size, in which mean values were calculated, had a substantial effect on the identified correlations.

In general, predicting response time and quality are crucial for

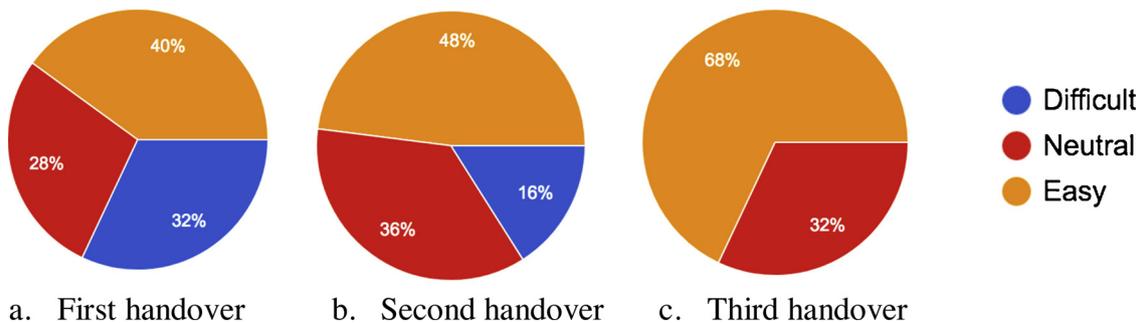


Fig. 7. Subjective ratings of the difficulty of first, second and third handovers.

highly automated driving to plan for transition times. There have been several papers studying different fixed transition times between 2–12 seconds (Gold et al., 2016, 2013; Körber et al., 2016; Zeeb et al., 2015). With our findings, automation systems could identify a dynamic transition time based on the driver's physiological state to allow for the safest handover process possible. In addition, based upon PerSpeed and PerAngle's correlation with driver's physiological measures, the automation systems could predict poorly performing drivers during handovers before a TOR is initiated. With such knowledge, the automation system may choose to resort to an emergency manoeuvre or give drivers a longer reaction time to regain situational awareness.

4.1. Effect of the window size on the physiological measures

The window size played a crucial role in the stability of the linear mixed models for response time, PerSpeed and PerAngle. Finding an optimal window size is usually an essential task in similar studies (Grimes et al., 2008; Tapia et al., 2007). In addition, several studies reported a trade-off between window size and accuracy in other studies (Solovey et al., 2014). Specifically, the findings of Solovey et al. (2014) showed that their longest heart rate window performed better than smaller ones which match with the findings in this study. Essentially, a large window size means a lag in understanding HR state (in comparison to instantaneous HR), but it provides an overall understanding of how driver's mental state affected the heart rate values.

Conversely, the findings of this study suggested that a smaller window size of PD performed better than larger window sizes. This could be because pupil diameter can change rapidly according to the changes in driver's cognitive workload (Klingner et al., 2011; Kramer et al., 2013). Our findings indicated that a 30 s window size performed well for the secondary task, response time and performance measures. Those findings match with outcomes of related studies (e.g., Klingner et al., 2011; Son et al., 2012) that concluded best performing window size was 30 s.

Surveyed literature used the mean value to analyse HR data (ROSCOE, 1993; Solovey et al., 2014). However, mean values negate both shape and transition of the HR signal that could provide better predictability than just the mean. MTF images, see Fig. 4, were generated to plot the average transitions of the HR signals of each group. Images demonstrate significant signal transitions among TQT and email tasks in comparison to the control group. A clear difference could be identified among the three images. Looking specifically at email and TQT, the signals had similar probabilities of transitions; however, temporal dependencies were significantly different. In contrast, the control group had fewer transitions probabilities and less temporal dependencies. This means that during secondary tasks, HR transitions were much more frequent than the control group. In addition, TQT transitions were much higher than the email group. The same analysis of PD will be reported in a future study.

4.2. Subjective ratings of the takeover's difficulty

The subjective rating reported by drivers indicated some interesting insight to understand some of the previously reported subjective measures of the NDR tasks. Zeeb et al. (2016) reported that the email task was ranked the most challenging task in a subjective rating which conflicted with their study's objective measures (e.g., deviation from lane centre). Their justifications were "email task was simply less demanding... drivers had difficulties rating their workload", (Zeeb et al., 2016).

On the contrary, this study reported that the email task engagement caused a significant increase in HR during the email task followed by a significant HR peak at TOR, see Fig. 3. Thus, an email task cannot be assumed a less demanding task as Zeeb et al. (2016) reported. In fact, Salvucci and Bogunovich (2010) reported that "Interruptions occurring at points of higher mental workload are more disruptive and lead to

larger resumption lags than those occurring at points of lower mental workload". Also, writing emails had a strong correlation with stress (Marulanda-Carter and Jackson, 2012). Reflecting that on our analysis, we could assume that the email task-induced a high mental workload which matches with the reported subjective measures of our study and Zeeb's et al., (2016) study. Accordingly, switching from the demanding email task to another demanding task (i.e., takeover) justified the subjective measures choices by the drivers and reflected on their physiological behaviour as reported earlier. Another explanation could be due to a significantly degraded situational awareness (due to a higher time-off-road (Kass et al., 2007) in comparison to the TQT group that maintained eyes-on-road throughout the task.

On the other hand, subjective measures showed that TQT was considered the least demanding task in our study by 82% of drivers. HR peak at TOR was significantly lower than the email task and so was the average HR as reported in Fig. 3. During TQT, drivers spent an average of 3.3 s (SD = 4.2) coming up with a new question. When drivers noticed a critical incident of a vehicle tailgating another on the neighbouring lane, a significant delay was reported (M = 4.2, SD = 1.2) among drivers to ask a new question. Such delayed responses showed that drivers performed multitasking between road monitoring and TQT.

According to Young and Stanton (2002), active engagement in tasks makes participants more engaged and more alert which makes it easier to takeover during TQT. Interestingly, the TQT group had a lower HR peak at TOR in comparison to the control group. Considering Young and Stanton (2002) findings again, the control group had no active task engagement before the TOR which turned them into a passive state causing a reduction in their level of alertness. Such an assumption could justify how the control group reported that it was difficult to engage in the takeover and why the control group had a higher HR peak than TQT group during the TOR phase. This indicates that a certain level of mental workload is preferential in the context of improved TOR quality. This should be explored in further research.

In addition to the takeover type, the order of handover processes played another role in the perception of the difficulty of the handover phase due to a previously reported learning curve of the automated system (Körber and Bengler, 2014; Larsson et al., 2014; Wright et al., 2016a). Reported subjective ratings (see Fig. 7) of the takeover difficulty sorted by order showed that 40% of drivers perceived the first takeover as easy in comparison to 68% in the third takeover. Such results are taken with caution because of previously reported divergence of driving performance and subjective estimates of performance (Horrey et al., 2009). Reported subjective measures align with the average heart rate peaks observed after TOR in the three takeovers as seen in Fig. 3.

To assess whether this assumption is valid, HR peak at TOR was recalculated by subtracting the mean of 'a 60 s normalised HR window before TOR' from the HR peak at TOR, referred to as relative HR peak in Table 3. Relative HR peak on Table 3 of the second (M = .08) and third (M = .09) takeovers were significantly smaller than the first takeover (M = .126). Relative peak HR may be considered a proxy measure of the difficulty of the takeover; especially that it matches with the reported subjective measures.

Table 3
Comparing normalised HR at each takeover.

		Normalised HR mean before TOR (60-sec window)	Relative HR Peak at the TOR
Takeover 1	Mean	0.47	0.126
	SD	0.125	0.22
Takeover 2	Mean	0.491	0.08
	SD	0.135	0.187
Takeover 3	Mean	0.424	0.093
	SD	0.130	0.127

4.3. Objective ratings of the takeover's difficulty

This study introduced PerSpeed and PerAngle as two new performance measures for take-over quality in Level-3 automated driving. Minimum time to collision is an established performance measure in the field (Radlmayr et al., 2019) but it does not provide any measure of how drivers handled their vehicle. For example, a driver could employ sharp braking; this may maximise the minimum time to the collision but introduces a significant hazard to other vehicles, as would sharp transitions to a neighbouring lane. PerSpeed and PerAngle provide the missing measures that will support minimum time to collision analysis.

PerSpeed and PerAngle metrics are a further development of previously introduced performance measures such as maximum lateral and longitudinal accelerations (Gold et al., 2016), deviation from lane centre (Zeeb et al., 2016), speed reduction (Larsson et al., 2014), percent road centre (Jamson et al., 2013). The introduction of PerSpeed and PerAngle could be a standard for assessing performance quality to provide comparability among studies. The literature review of this study identified no open access datasets that we could have been used to cross-validate those measures on previous studies.

4.3.1. The correlation between the physiological data and performance measures

PerSpeed showed some interesting correlations with the physiological measures. Even though there was no statistical difference among secondary task groups, they correlated with physiological measures. This could be due to some drivers who can handle secondary tasks better than others; thus, no observed increase in HR/PD. For example, a driver who has more experience would respond better than others under the same secondary task condition (Ko and Ji, 2018) causing less stress and probably less increase in their physiological behaviour (Adler et al., 2000). This could explain why PerSpeed or PerAngle have no statistical significance among secondary task groups.

Though, PerSpeed and PerAngle showed a strong correlation with physiological measures which, in turn, has a strong correlation with mental workload (Brookhuis and De Waard, 2010; Marquart et al., 2015). Previous studies reported braking as a link to the lack of situational awareness (Zeeb et al., 2015), which shows the possibility that situational awareness could also be correlated to PerSpeed and potentially PerAngle. Hence, future work is required for further exploration.

The results demonstrated further evidence that HR and PD correlate with braking behaviour of drivers. According to (Gold et al., 2013), braking is associated with out-of-loop drivers allowing themselves a longer time to restore situational awareness. Due to an increase in mental workload is associated with an increase in HR (Wilson, 2002) and PD (Batmaz and Ozturk, 2008), the reported results indicated that PerSpeed, PerAngle and driver's mental workload have an indirect negative correlation; assuming drivers are out-of-loop.

4.3.2. Assessment of the physiological data as valid predictors of the performance measures

The study analysed heart rate and pupil diameter to assess their predictability of the performance measures. The study identified that pupil diameter was unstable measure throughout the experiment. For example, the main secondary task performed by the medium risk group was the email task. Reflecting that on the reported results in Section 4.4.2. The increase in pupil diameter in the second task group could have been due to the difference in lighting between the tablet screen and the simulator screens. This could explain the significant difference in pupil diameter in comparison to other groups; even though, the experiment setup insured a minimal change in lighting throughout all screens. Another explanation could be due to the change in pupil diameter as participants transit from one screen to another since the lab was significantly darker than the two screens.

The results highlight that the established knowledge that the pupil diameter measure is valid only in highly controllable measures

(Marquart et al., 2015) where the environment of the experiment is highly controlled. This means that pupil diameter estimation in real-world driving may not be accurate; however, there has been significant research in assessing mental workload under different lighting conditions (Pfleging et al., 2016). Hence, results presented in this study are potentially useful to be used in real-world applications.

In conclusion, the study provided the experimental evidence that physiological measures prior to a takeover are capable of predicting drivers' response time and quality as predicted by (Chan and Singhal, 2015). It also aligns with the vision introduced by Rakotonirainy et al. (2014) that predicting driver's behaviour could enhance their experience. Based on previous suggestions, Heger et al. (2010) a mental workload recognition system using EEG and machine learning techniques is extendable to highly automated driving scenarios. Findings of this study have enabled a machine learning model to be built to accurately predict drivers' mental workload and extend that to predict the response quality of the takeover.

5. Future work

As identified through the literature review of this study, there are some factors affecting response time and quality such as fatigue (Driver, 2014), age (Körber et al., 2016), traffic density (Christian Gold et al., 2016), weather conditions (Louw et al., 2016) and driving experience (Larsson et al., 2014). Those variables were not taken into consideration due to the limitations of the study; however, it should be considered for future work. Additionally, future work will need to recruit different age groups which was identified by others as a critical variable in driver's performance in highly automated driving

Finally, the study acknowledges the limited effort placed on comparing PerSpeed and PerAngle with other measures that were surveyed and introduced in Radlmayr's et al., (2019) comprehensive study. The comparison between TOPS model and their affiliated vehicle-based performance measure would be a great contribution to the research of highly automated vehicles.

6. Conclusion

There is a substantial interest in predicting the driver's response time and the quality of takeovers in highly automated driving. Not only do they affect the safety of the system but also provide a great opportunity for the automated system to assess the reliability of the person behind the wheel throughout the journey and specifically before starting the handover process.

As automated systems will change the dynamics of driving, it is essential that the systems have a deeper understanding of the driver's mental state and their ability to perform well in emergencies that could cause serious accidents. This research was performed to explore the prospect of identifying the necessary measure to define the driver's mental state and link their physiological behaviour before a takeover request with their response time and takeover quality.

The most significant findings of the study are, heart rate and pupil diameters of drivers are valid predictors for both response time and determining the quality of takeovers in highly automated driving environments. Interestingly, these results are similar to findings in air traffic control and aviation systems in addition to manual driving studies that were performed previously. The findings of this experiment provide the path to building 1) generic driver state model and 2) a driver's response time and the quality of takeover prediction models. The models could be applied to all drivers based on the physiological behaviours without necessarily accounting for individual differences or relying on identifying the secondary tasks drivers were performing.

Moreover, two new quality measures were introduced and examined in this study to provide an estimate of braking and steering, and they were linked to drivers' physiological measures. Those measures could be used by the automation system to assess the driver's future

responses. Those quality measures, PerSpeed and PerAngle, could be generalised as valid performance measures for takeover scenarios. Those proposed measures provide insight into driver's readiness and therefore, allow automated systems to adopt the right driving strategy and plan to enhance their experience and make the transition phase safer for everyone.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research is sponsored by the UK's Engineering and Physical Sciences Research Council through a grant from the Embedded Intelligence Centre of Doctoral Training. All opinions, findings and conclusions are of the authors and do not necessarily reflect on the views of the sponsors.

References

- Adler, N.E., Epel, E.S., Castellazzo, G., Ickovics, J.R., 2000. Relationship of subjective and objective social status with psychological and physiological functioning: preliminary data in healthy white women. *Health Psychol.* 19 (6), 586–592. <https://doi.org/10.1037/0278-6133.19.6.586>.
- Batmaz, I., Ozturk, M., 2008. Using pupil diameter changes for measuring mental workload under mental processing. *J. Appl. Sci.* 8 (1), 68–76. <https://doi.org/10.3923/jas.2008.68.76>.
- Brookhuis, K.A., De Waard, D., 2010. Monitoring drivers' mental workload in driving simulators using physiological measures. *Accid. Anal. Prev.* 42, 898–903. <https://doi.org/10.1016/j.aap.2009.06.001>.
- Campbell, J.L., Brown, J.L., Graving, J.S., Richard, C.M., Lichty, M.G., Bacon, L.P., Morgan, J.F., Sanquist, T., 2018. *Human Factors Design Principles for Level 2 and Level 3 Automated Driving Concepts*. August, 122.
- Campbell, J.L., Richard, C.M., Brown, J.L., McCallum, M., 2007. *Crash Warning System Interfaces: Human Factors Insight and Lessons Learned*. US Dep. Transp. Natl. Highw. Traffic Saf. Adm. No. HS-810 January, 1–179. doi:HS 810 697.
- Carsten, N., Lai, F.C.H., Barnard, Y., Jamson, A.H., Merat, N., 2012. Control task substitution in semiautomated driving. *Hum. Factors J. Hum. Factors Ergon. Soc.* 54 (5), 747–761. <https://doi.org/10.1177/0018720812460246>.
- Chan, M., Singhal, A., 2015. Emotion matters : implications for distracted driving. *Saf. Sci.* 72, 302–309. <https://doi.org/10.1016/j.ssci.2014.10.002>.
- de Waard, D., van der Hulst, M., Hoedemaeker, M., Brookhuis, K.A., 1999. Driver behavior in an emergency situation in the automated highway system. *Transp. Hum. Factors* 1 (1), 67–82. https://doi.org/10.1207/sthf0101_7.
- De Winter, J.C.F., Happee, R., Martens, M.H., Stanton, N.A., 2014. Effects of adaptive cruise control and highly automated driving on workload and situation awareness: a review of the empirical evidence. *Transp. Res. Part F: Traffic Psychol. Behav.* <https://doi.org/10.1016/j.trf.2014.06.016>.
- Department of Transport, 2015. *The Pathway to Driverless Cars Summary Report and Action Plan*.
- Driver, E.T.H.E., 2014. *Fatigue in the Automated Vehicle: Do Games and Conversation Distract or Energize the Driver?* 2009. pp. 2053–2057.
- Endsley, M.R., 1995. Toward a theory of situation awareness in dynamic systems. *Hum. Factors J. Hum. Factors Ergon. Soc.* 37 (1), 32–64.
- Endsley, M.R., Onal, E., Problems, A.A., 1997. *The Impact of Intermediate Levels of Automation on Situation Awareness and Performance in Dynamic Control Systems*.
- Flemisch, F., Schieben, A., Schoemig, N., Strauss, M., Lueke, S., Heyden, A., 2011. *Design of Human Computer Interfaces for Highly Automated Vehicles in the EU-Project HAVEit*. Springer, Berlin, Heidelberg, pp. 270–279. https://doi.org/10.1007/978-3-642-21666-4_30.
- Gold, C., Berisha, I., Bengler, K., 2015a. Utilization of drivetime - Performing non-driving related tasks while driving highly automated. *Proc. Hum. Factors Ergon. Soc.* 1666–1670. <https://doi.org/10.1177/1541931215591360>. 2015–Janua 2013.
- Gold, C., Berisha, I., Bengler, K., 2015b. Utilization of Drivetime – Performing Non-driving Related Tasks While Driving Highly Automated. 2013. pp. 1666–1670.
- 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 Annual Meeting*. pp. 1938–1942.
- Gold, C., Körber, M., Lechner, D., Bengler, K., Gold, Christian, Körber, Moritz, David Lechner, K.B., 2016. Taking over control from highly automated vehicles in complex traffic situations the role of traffic density. *Hum. Factors J. Hum. Factors Ergon. Soc.* <https://doi.org/10.1177/0018720816634226>.
- Grimes, D., Tan, D.S., Hudson, S.E., Shenoy, P., Rao, R.P.N., 2008. Feasibility and pragmatics of classifying working memory load with an electroencephalograph. *Proceeding Twenty-Sixth Annu. CHI Conf. Hum. Factors Comput. Syst. - CHI' 08* 835. <https://doi.org/10.1145/1357054.1357187>.
- Heger, D., Putze, F., Schultz, T., 2010. Online workload recognition from EEG data during cognitive tests and human-machine interaction. *Ki* 2010. *Adv. Artif. Intell.* 6359, 410–417.
- Horrey, W.J., Lesch, M.F., Garabet, A., 2009. Dissociation between driving performance and drivers' subjective estimates of performance and workload in dual-task conditions. *J. Saf. Res.* 40 (1), 7–12. <https://doi.org/10.1016/J.JSR.2008.10.011>.
- Jamson, A.H., Merat, N., Carsten, O.M.J., Lai, F.C.H., 2013. Behavioural changes in drivers experiencing highly-automated vehicle control in varying traffic conditions. *Transp. Res. Part C Emerg. Technol.* 30, 116–125. <https://doi.org/10.1016/j.trc.2013.02.008>.
- Jamson, A.H., Westerman, S.J., Hockey, G.R., Carsten, O.M., 2004. Speech-based E-mail and driver behavior: effects of an in-vehicle message system interface. *Hum. Factors* 46 (4), 625–639. <https://doi.org/10.1518/hfes.46.4.625.56814>.
- Kass, S.J., Cole, K.S., Stanny, C.J., 2007. Effects of distraction and experience on situation awareness and simulated driving. *Transp. Res. Part F Traffic Psychol. Behav.* 10 (4), 321–329. <https://doi.org/10.1016/j.trf.2006.12.002>.
- Klingner, J., Tversky, B., Hanrahan, P., 2011. Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks. *Psychophysiology* 48 (3), 323–332. <https://doi.org/10.1111/j.1469-8986.2010.01069.x>.
- Ko, S.M., Ji, Y.G., 2018. How we can measure the non-driving-task engagement in automated driving: comparing flow experience and workload. *Appl. Ergon.* 67, 237–245. <https://doi.org/10.1016/j.apergo.2017.10.009>.
- Körber, M., Bengler, K., 2014. Potential individual differences regarding automation effects in automated driving. *Proceedings of the XV International Conference on Human Computer Interaction*. pp. 22.
- Körber, M., Gold, C., Lechner, D., Bengler, K., 2016. The influence of age on the take-over of vehicle control in highly automated driving. *Transp. Res. Part F Traffic Psychol. Behav.* 39, 19–32.
- Kramer, S.E., Lorens, A., Coninx, F., Zekveld, A.A., Piotrowska, A., Skarzynski, H., 2013. Processing load during listening: the influence of task characteristics on the pupil response. *Lang. Cogn. Process.* 28 (4), 426–442. <https://doi.org/10.1080/01690965.2011.642267>.
- Larsson, A.F.L., Kircher, K., Andersson, J., Hultgren, J.A., 2014. Learning from experience: familiarity with ACC and responding to a cut-in situation in automated driving. *Transp. Res. Part F Traffic Psychol. Behav.* 27, 229–237. <https://doi.org/10.1016/j.trf.2014.05.008>.
- Lloyd, S.P., 1982. Least squares quantization in PCM. *IEEE Transactions On Information Theory*.
- Louw, T., Madigan, R., Carsten, O., Merat, N., 2016. Were they in the loop during automated driving? Links between visual attention and crash potential. *Inj. Prev.* 2016.
- Marquart, G., Cabrall, C., de Winter, J., 2015. Review of eye-related measures of drivers' mental workload. *Procedia Manuf.* 3, 2854–2861. <https://doi.org/10.1016/J.PROMFG.2015.07.783>.
- Marulanda-Carter, L., Jackson, T.W., 2012. Effects of e-mail addiction and interruptions on employees. *J. Syst. Inf. Technol.* 14 (1), 82–94. <https://doi.org/10.1108/13287261211221146>.
- Mehler, B., Reimer, B., Coughlin, J.F., Dusek, J.A., 2009. Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transp. Res. Rec. J. Transp. Res. Board* 2138 (1), 6–12. <https://doi.org/10.3141/2138-02>.
- Meinlschmidt, G., Stalujan, E., Tegethoff, M., 2018. The psychobiology of using automated driving systems: a systematic review and integrative model. *Psychoneuroendocrinology* (September), 1–13. <https://doi.org/10.1016/J.PSYNEUEN.2018.09.029>.
- Melcher, V., Rauh, S., Diederichs, F., Bauer, W., 2015. *ScienceDirect Take-Over Requests for Automated Driving*. 00. pp. 4219–4225.
- Melnicuk, V., Birrell, S.A., Konstantopoulos, P., Crundall, E., Jennings, P.A., 2016. *JLR Heart: Employing Wearable Technology in Non-intrusive Driver State Monitoring. Preliminary Study*.
- Merat, N., Jamson, A.H., Lai, F.C.H., Carsten, O., 2012. Highly automated driving, secondary task performance, and driver state. *Hum. Factors J. Hum. Factors Ergon. Soc.* 54 (5), 762–771. <https://doi.org/10.1177/0018720812442087>.
- NHTSA, 2013. *U.S. Department of Transportation Releases Policy on Automated Vehicle Development*. U.S. Dep. Transp. Releases Policy Autom. Veh. Dev.
- Petermann-Stock, I., Hackenberg, L., Muhr, T., Mergl, C., 2013. Wie lange braucht der Fahrer—eine Analyse zu {Ü}bernahmezeiten aus verschiedenen Nebent{ä}tigkeiten w {ä}hrend einer hochautomatisierten Staufahrt. 6. Tagung Fahrerassistenzsysteme. Der Weg zum Autom. Fahr. Wie lange braucht der Fahrer—eine Analyse zu {Ü}bernahmezeiten aus verschiedenen Nebent{ä}tigkeiten w {ä}hrend einer hochautomatisierten Staufahrt. 6. Tagung Fahrerassistenzsysteme. Der Weg zum Autom. Fahr.
- Pfleging, B., Fekety, D.K., Schmidt, A., Kun, A.L., 2016. A model relating pupil diameter to mental workload and lighting conditions. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI' 16* 5776–5788. <https://doi.org/10.1145/2858036.2858117>.
- 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 Annual Meeting*. pp. 2063–2067.
- Radlmayr, J., Ratter, M., Feldhütter, A., Körber, M., 2019. Take-overs in Level 3 Automated Driving – Proposal of the Take-over Performance Score (TOPS). 821 January. pp. 1–10. <https://doi.org/10.1007/978-3-319-96080-7>.
- Rakotonirainy, A., Schroeter, R., Soro, A., 2014. Three social car visions to improve driver behaviour. *Pervasive Mob. Comput.* 14, 147–160. <https://doi.org/10.1016/j.pmcj.2014.06.004>.
- Recarte, M.A., Nunes, L.M., 2003. Mental workload while driving: effects on visual search, discrimination, and decision making. *J. Exp. Psychol. Appl.* 9 (2), 119.

- Recarte, M.A., Nunes, L.M., 2000. Effects of verbal and spatial-imagery tasks on eye fixations while driving. *J. Exp. Psychol. Appl.* 6 (1), 31–43. <https://doi.org/10.1037/1076-898X.6.1.31>.
- ROSCOE, A.H., 1993. Heart rate as a psychophysiological measure for in-flight workload assessment. *Ergonomics* 36 (9), 1055–1062. <https://doi.org/10.1080/00140139308967977>.
- SAE International, 2018. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. <https://doi.org/10.4271/981198>. Dingus.
- Salvucci, D.D., Bogunovich, P., 2010. Multitasking and monotasking. *Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI' 10* 85. <https://doi.org/10.1145/1753326.1753340>.
- Solovey, E.T., Zec, M., Garcia Perez, E.A., Reimer, B., Mehler, B., 2014. Classifying driver workload using physiological and driving performance data. *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems - CHI' 14* 4057–4066. <https://doi.org/10.1145/2556288.2557068>.
- Son, J., Park, M., Oh, H., 2012. Detecting cognitive workload using driving performance and eye movement in a driving simulator. *PROceeding 11th Int. Symp. Adv. Veh. Control.* pp. 1–4.
- STISIM, 2018. Car Driving Simulator and Software for Occupational Therapy, Research, and Training [WWW Document]. 2017. URL <http://stisimdrive.com/> (Accessed 12 February 2018).
- Tapia, E.M., Intille, S.S., Haskell, W., Larson, K.W.J., King, A., Friedman, R., 2007. Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart monitor. *Int. Symp. Wearable Comput.* 37–40. <https://doi.org/10.1109/ISWC.2007.4373774>.
- Wandtner, B., Schmidt, G., Schoemig, N., Kunde, W., 2018. Non-driving related tasks in highly automated driving – effects of task modalities and cognitive workload on take-over performance Participants. *Automot. Meets Electron. Conf.* 8–13.
- Wang, R., Blackburn, G., Desai, M., Phelan, D., Gillinov, L., Houghtaling, P., Gillinov, M., 2016. Accuracy of wrist-worn heart rate monitors. *JAMA Cardiol.*
- Wang, Z., Oates, T., 2014. Encoding Time Series as Images for Visual Inspection and Classification Using Tiled Convolutional Neural Networks.
- Wille, M., Röwenstrunk, M., assistance, G.D.-H.F. for, 2007, undefined, n.d. KONVOI: Electronically coupled truck convoys. hfes-europe.org.
- Wilson, G.F., 2002. An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *Int. J. Aviat. Psychol.* 12 (1), 3–18. https://doi.org/10.1207/S15327108IJAP1201_2.
- Wright, T.J., Samuel, S., Borowsky, A., Zilberstein, S., Fisher, D.L., 2016a. Experienced drivers are quicker to achieve situation awareness than inexperienced drivers in situations of transfer of control within a level 3 autonomous environment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.* pp. 270–273.
- Wright, T.J., Samuel, S., Borowsky, A., Zilberstein, S., Fisher, D.L., Swedish National, R., Transport Research, I., 2016b. Are Experienced Drivers Quicker to Regain Full Situation Awareness in Scenarios Involving Transfer of Control From the Automation to the Driver? p. 3p..
- Young, M.S., Stanton, N.A., 2002. Malleable attentional resources theory: a new explanation for the effects of mental underload on performance. *Hum. Factors J. Hum. Factors Ergon. Soc.* 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. *Accid. Anal. Prev.* 92, 230–239.
- Zeeb, K., Buchner, A., Schrauf, M., 2015. What determines the take-over time? An integrated model approach of driver take-over after automated driving. *Accid. Anal. Prev.* 78, 212–221. <https://doi.org/10.1016/j.aap.2015.02.023>.