Effect of cognitive load on drivers' State and task performance during automated driving: Introducing a novel method for determining stabilisation time following take-over of control

Melnicuk, V., Thompson, S., Jennings, P. & Birrell, S.

Original citation & hyperlink:
Melnicuk, V, Thompson, S, Jennings, P & Birrell, S 2021, 'Effect of cognitive load on drivers' State and task performance during automated driving: Introducing a novel method for determining stabilisation time following take-over of control', Accident Analysis & Prevention, vol. 151, 105967.
https://dx.doi.org/10.1016/j.aap.2020.105967

DOI 10.1016/j.aap.2020.105967
ISSN 0001-4575

Publisher: Elsevier

Effect of Cognitive Load on Drivers' State and Task Performance During Automated Driving: Introducing a Novel Method for Determining Stabilisation Time Following Take-Over of Control

Vadim Melnicuk[1], Simon Thompson[2], Paul Jennings[3] and Stewart Birrell[4]
[1] Lotus Cars, Potash Lane, Hethel, Norwich NR14 8EZ, United Kingdom
[2] Jaguar Land Rover, Banbury Road, Gaydon CV35 0RR, United Kingdom
[3] WMG, University of Warwick, 6 Lord Bhattacharyya Way, Coventry CV4 7AL, United Kingdom
[4] National Transport Design Centre, Coventry University, Swift Road, Coventry CV1 2TT, United Kingdom

1 Abstract

This research paper explores the impact of cognitive load on drivers' physiological state and driving performance during an automated driving to manual control transition scenario, using a driving simulator. Whilst driving in the automated mode, cognitive load was manipulated using the "N-Back" task, which participants engaged with via a visual display. Results suggest that non-optimal levels of workload during the automated driving conditions impair driving performance, especially lateral control of the vehicle, and the magnitude of this impairment varied with increasing cognitive load. In addition to these findings, the present paper introduces a novel method for determining stabilisation times of both driver state and driving performance indicators following a transition of vehicle control. Using this method we demonstrate that mean and standard deviation of lane position impairments were found to take longer to stabilise following transition to manual driving following a higher level of cognitive load during the automated driving period, taking up to 22 seconds for driving performance to normalise after take-over. In addition, heart rate parameters take between 20 and 30 seconds to stabilise following a planned take-over request. Finally, this paper demonstrates how the magnitude of cognitive load can be estimated in context of automated driving using physiological measures, captured by consumer electronic devices. We discuss the impact our findings have on the design of SAE Level 3 systems. Relevant suggestions are provided to the research community and automakers working on future implementation of vehicles capable of conditional automation.

1.1 Keywords

Automotive Engineering, Autonomous Driving, Biometrics, Human Factors, Physiology, Vehicle Safety

2 Introduction

The ongoing and advancing development of adaptive driver assistance systems will likely result in implementation of full vehicle automation. Existing forecasts are that both conditional (i.e., Level 3 in SAE taxonomy, or being able to self-drive in certain situations with the human driver a fall back for regaining control) and fully autonomous passenger vehicles will become an integral part of transportation networks in various counties by 2035 (Worstall, 2015). Moreover, Burghardt, Weig and Choi (2017) are even more optimistic regarding forecasts for automated vehicle uptake, predicting that approximately 35% of newly sold vehicles will be capable of conditional automation, and 15% of vehicles will be embedded with high automation enabling technology (i.e., Level 4 in SAE taxonomy; SAE International, 2015) by 2030. It should be noted that as of now, vehicles capable of SAE Level 4 automation do not exist outside of advanced research concepts (Campbell *et al.*, 2018). However, the pace at which these vehicles become introduced to the market depends on the multitude of factors including the rate of technological advancement, costs, public trust and acceptance of self-driving vehicles, as well as the ability of automakers and the research community to address safety-related concerns (Burghardt, Weig and Choi, 2017). Safety is often deemed as one of the most important concerns, since public perception of 'self-driving vehicles' can be negatively affected by occasional failures of automated vehicle systems. Rare catastrophic events, such as the fatal accident involving Uber self-driving test vehicle, for example, have a powerful effect on people's perceived level of risk of autonomous driving technology (Sage, Bellon and Carey, 2018). It was previously indicated that acceptance of vehicle automation can be linked to assurance regarding the safety of the technology (Khastgir *et al.,* 2018; Lee *et al.*, 2020). Automakers recognize that they must direct their resources towards ensuring high degree of safety of automated vehicles through rigorous testing before these vehicles reach the market.

The focus of the present paper is on aspects of safety during semi-autonomous vehicle control, which can emerge in vehicles capable of SAE Level 3 automation. In such vehicles, a driver may occasionally be required to 'take-over' the control of the vehicle, often in a short period of time (SAE International, 2015). Since conditionally automated vehicles do not require drivers to monitor the situation during the self-driving mode, a driver might be engaged in variety of secondary tasks prior to taking over manual control. For example, a driver may be working, watching a movie, reading, or engage in any other cognitively demanding task. Indeed, previous research found that approximately 45% of US drivers and 32% of UK drivers are willing to engage in cognitively demanding tasks of this kind whilst in a self-driving vehicle (Schoettle and Sivak, 2014). Cunningham and Regan (2017) argued that drivers are likely to exploit the attentional resources freed up by the automation in order to engage in variety of secondary tasks unrelated to the driving situation. However, this poses a potential risk in situation when a driver is required to take control over the vehicle. According to National Highway Traffic Safety Administration (NHTSA) a driver could become so immersed in a secondary task that information prompting them to switch to manual control of the vehicle could result in a substantial time delay or, in the worst case scenario, complete failure to engage in manual driving (Campbell *et al.*, 2018). Alternatively, in the cases when a secondary task is not present, drivers might become subject to passive fatigue or sleepiness due to monotonous nature of automated driving over extended period of time (Saxby *et al.*, 2013). Hence, the lack of cognitive control during driving could cause impairments in driving performance and promote performance-related human errors and ultimately lead to a traffic-related car accident. Both very low and very high levels of arousal are associated with poor performance in difficult tasks (Wickens and Hollands, 2000). Even in situations where a driver does take over manual control of the vehicle, the arousal state at the time of engaging with the secondary task could impact drivers' ability to adequately perform the primary task of driving, particularly after an extensive period of automated driving.

There is no doubt that transition from automated to manual control is not a trivial task. Vogelpohl *et al.* (2018) argued that drivers might require additional time and assistance in order to reach a level of situational awareness necessary to resume manual driving. Interestingly, Merat *et al.* (2014) suggested it could take approximately 40 seconds for drivers to resume an adequate and stable lateral control of driving when they switch from autonomous to manual driving. In the scope of their paper, the signal stabilisation refers to a phenomenon where a driving performance signal enters a visual plateau following a period of significant oscillation. Similarly, Pampel *et al.* (2018) have found a diminishing difference in mean speed, measured using 5-second time bins, after approximately 20 seconds of vehicle control take-over and standard deviation of lane position did not stabilise for the first 10 seconds during a non-distractive short take-over driving scenario. Furthermore, in a study of take-over performance during highly automated driving scenarios, Clark *et al.* (2017) found that younger drivers tend to respond faster to the take-over requests when distractive secondary tasks are present during autonomous driving mode. The gap this present paper addresses is by introducing a novel statistical method for determining this stabilisation time.

It is likely that drivers' ability to take-over control and immediately perform the primary task of driving may in fact be impacted by their state cognitive and emotional state i.e., a level of arousal and stress caused by differences in workload. Drivers who are occupied by a highly demanding secondary task may be distracted (Horberry *et al.*, 2006; Schaap *et al.*, 2013) and, equally, those with low level of workload may become more fatigued (Philip *et al.*, 2005; Matthews *et al.*, 2011; Saxby *et al.*, 2013). For instance, passive fatigue causes reduction of task engagement, focus, and slower responses to the emergency events (Saxby *et al.*, 2013). Moreover, fatigue was found to be associated with impairment of drivers' performance, in both longitudinal and lateral control (Matthews and Desmond, 2002; Gastaldi, Rossi and Gecchele, 2014). This was demonstrated in both simulator and real world conditions (Philip *et al.*, 2005). Similarly, an increase in workload was previously found to be associated with decrease in driving performance (Paxion, Galy and Berthelon, 2014). Of importance to the present paper is the time course over which a driver can return to the level of optimal psychological state, following a heightened level of arousal and/or cognitive load.

How can drivers' psychological state be assessed whilst driving? Some of the existing research has focused on methods of estimating driver psychological state, such as fatigue, using various driving performance indices e.g., steering wheel angles (Krajewski *et al.*, 2009; He, Li and Fan, 2011). Steering entropy, a measure of steering consistency, was also previously linked to driver workload (Nakayama *et al.*, 1999) as well as driving performance impairment (Kersloot, Flint and Parkes, 2003). However, steering input is absent during the period of conditional automation, and as such, cannot be used as an indicator of drivers' state. Instead, "*physiological measures can be sensitive to changes in workload before the appearance of clear decrements in driving performance*" Mehler *et al.* (2009, p. 1). Indeed, in our own work the authors found that physiological measures e.g., heart rate variability, are sensitive to changes in drivers' cognitive workload (Melnicuk, Birrell, Konstantopoulos, *et al.*, 2016; Melnicuk *et al.*, 2017). Consequently, there are clear benefits of objective driver state monitoring in various driving scenarios (Melnicuk, Birrell, Crundall, *et al.*, 2016). This means that in the context of conditional automation, physiological measures might be best suited for assessing driver's psychological state.

A vehicle that is able to estimate drivers' state in real-time, could dynamically adapt its driver assistance systems to address any abnormalities in drivers' state in order to ensure safe and effective transition of manual control. For instance, if such a vehicle detects presence of high cognitive load, it could keep an adaptive cruise control or lane keep assist enabled for extended period of time. Furthermore, full control can be gradually transferred to a driver after workload level is stabilised. This might help to minimise the possibility of human error occurrence and, as the result, ensure efficient and comfortable transition between automated and manual driving. This is echoed by Inagaki (2003), who defined concept of adaptive automation, with the authors advocating for the control of functions between humans and machines to be shifted dynamically, depending on environmental factors, drivers' workload, and performance. Similarly, research by Ulahannan *et al.,* (2020) suggest that using an adaptive in-car interface would also support partially automated driving, as system experience increased over a five day period, so did the number of glances to an in-vehicle display which presented the 'technical competency' of the automated system. Both these adaptive aspects may help to facilitate smoother transitions of vehicle control. The goal of the present work is to evaluate feasibility of a physiology-monitoring system in the context of autonomous to manual transition in conditional automation vehicles.

2.1 Problem identification

A number of measures for quality assessment of control take-over scenarios have been considered in the previous work (Merat *et al.*, 2012, 2014; Zeeb, Buchner and Schrauf, 2016; Clark *et al.*, 2017; Pampel *et al.*, 2018; Vogelpohl *et al.*, 2018). For instance, driving performance was mainly evaluated using changes in mean speed and standard deviation of lane position. A common method is to also monitor eye glance behaviour and reaction times to various on-road events. Occasionally, driver state is evaluated using subjective measures.

In order to further enhance understanding of factors affecting quality of automated to manual control take-over events and subsequent period of manual driving, it could be beneficial to objectively analyse drivers' state during these scenarios. There is limited evidence informing effect of workload on drivers' ability to effectively take-over manual control of a vehicle after a prolonged period of automated driving. It is also essential to understand how workload affects drivers' ability to perform the primary task of driving during the first minute of manual driving.

Previous findings indicate that some driving performance measures can take from 25 to 40 seconds to stabilise from the point of control take-over (Merat *et al.*, 2014). However, it could be that the adopted method was not optimised for deriving stabilisation times. Whilst visually the results suggest a trend for stabilisation of driving performance signals, no statistically significant difference was observed for stabilisation time using the method adopted by the authors. Hence building on the work by Merat *et al.* (2014), a new method for deriving stabilisation time needs be introduced which could yield more sensitive and reproducible results. Reporting this new method is one of the key contributions of this current paper.

It is also unknown whether stabilisation times vary significantly depending on the amount of workload drivers are exposed to prior to transition. Moreover, stabilisation times of objective driver state indicators i.e., through physiology, have not previously been studied in the context of automated driving. The real-time objective driver state assessment could help to better understand precisely how long it might take for a driver to get back in the loop as well as reach an optimal driving performance.

2.2 Hypotheses

To address the knowledge gaps in the literature, the following hypotheses were formulated:

1) The level of cognitive load can be reliably estimated using objective driver state indicators (e.g., through physiology) during the automated driving period;
2) A high level of induced cognitive load will cause significant impairment in driving performance after an automated to manual control take-over;
3) Different levels of induced cognitive load will result in different responses in driver state indicators during the automated driving, during the transition of control period, and during the manual driving period that follows;
4) The time it takes for driving performance to stabilise following a control take-over will be impacted by the cognitive load experienced during the automated driving period;
5) The time it takes for drivers' state to stabilise following a control take-over will be impacted by the amount of cognitive load experienced during the automated driving period;

3 Methodology

3.1 Participants

Forty-two participants were recruited for this study. Participants had to be aged 21 or over, hold a full category "B" driving license, have normal or correct-to-normal vision, and not have any cardiovascular or skin diseases in order to participate. Participants were recruited through direct email contact at University of Warwick and Jaguar Land Rover UK. Ethical approval to run this study was acquired from Biomedical & Scientific Research Ethics Committee (BSREC) at University of Warwick.

3.2 Study design

This experiment was conducted in a simulated driving environment and adopted a repeated measures design. Within a single driving scenario, we varied the environment (i.e., urban and motorway), as well as level of induced workload during the automated driving section using three complexity levels of “N-Back” task (see **Figure 1**).

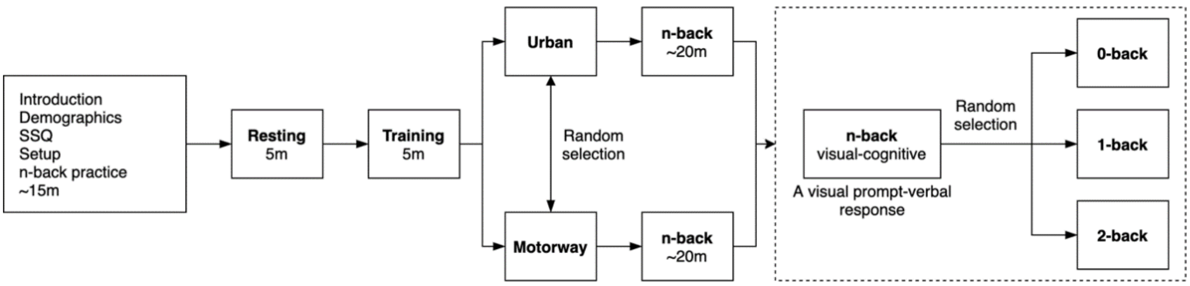


Figure 1. Study design diagram.

3.2.1 Apparatus

The study was performed in the 3xD simulator for intelligent vehicles at University of Warwick (see **Figure 2**). The high-fidelity driving simulator was used to run complex driving scenarios with fixed environmental conditions and simulate transitions of automated to manual control, all in a safe and controlled setting. The driving simulator removed much of the risk associated with the testing of new technology in the driving context including, such as risk of a crash due to high level of workload or distraction due to presence of highly intrusive events. To ensure immersiveness, the 3xD simulator consists of a full-size vehicle and 360 degrees’ cylindrical screen with high definition projection, all enclosed in a soundproof room. The simulator was set to emulate the dynamic model of a Range Rover Evoque with an automatic gear box.

Figure 2. 3xD simulator for intelligent vehicles at University of Warwick.

As part of the study, participants were asked to complete a series of driving scenarios in the simulator, while their driving performance and physiological responses were captured.

To facilitate collection of physiological data using consumer electronic devices (CED), all participants wore a POLAR H10 Heart Monitor (Polar Electro, 2018). Furthermore, a Samsung Galaxy S8 Edge smartphone was mounted to the windscreen inside the driving simulator vehicle. The smartphone also acted as a hub for the data collection and synchronising of all incoming data from the wearable device.

A custom-built Android toolkit was used to collect, synchronise, and store physiological measures from multiple data sources, as per method described in the authors' previous publication (Melnicuk *et al.*, 2017). The toolkit incorporates a mixture of driver state metrics captured from a CED-based sensory network (see **Figure 3**). It facilitates collection, storage, synchronisation, and filtering as well as calculation of some data derivatives e.g., time-domain HRV indicators. As part of this publication the toolkit is released as a citeable, open source application and is accessible at https://github.com/vadimmelnicuk/WMGDSM.

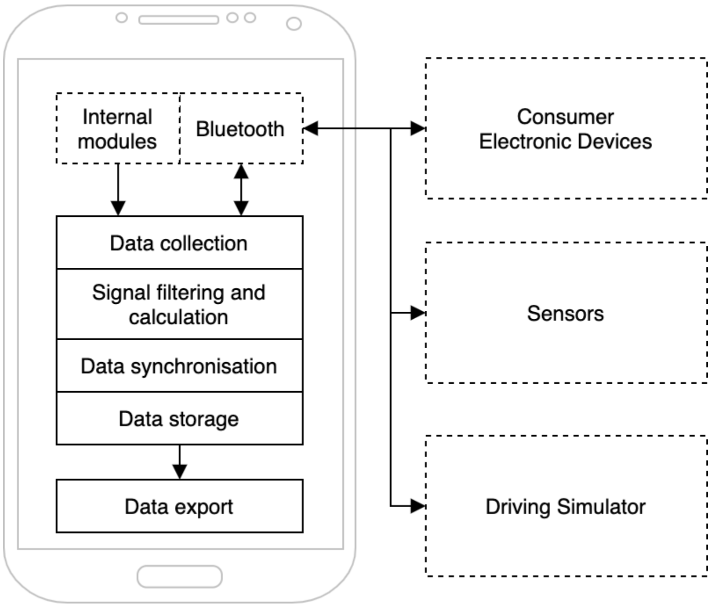


Figure 3. DSM toolkit data flow diagram.

3.2.2 Signals

The summary of signals and their underlying frequencies is provided in **Table 1**. Furthermore, the raw heart rate was used to calculate time-domain Heart Rate Variability (HRV) derivatives following the methods, described in authors' previous publications (Melnicuk, Birrell, Konstantopoulos, *et al.*, 2016; Melnicuk *et al.*, 2017).

Table 1. Summary of measures concerning physiological state estimation.

Measure	Polar H10, Frequency
Heart Beats Per Minute (BPM)	1 Hz
Inter-beat Intervals (RR)	1 Hz

In addition to physiological responses, the standard set of driving performance measures was captured (see **Table 2**). Some additional driving performance derivatives (e.g., lane position and standard deviation of lane position) were later calculated in accordance to SAE's operational definitions of driving performance measures and statistics (SAE International, 2013).

Table 2. Summary of measures concerning driving performance.

Measure	3xD Simulator HIL, Frequency
Position x, y, z (m)	100 Hz
Velocity x, y, z (m/s)	100 Hz
Angular velocity x, y, z (radians per second)	100 Hz
Speed (mph)	100 Hz
Revolutions per minute (RPM).	100 Hz

Steering angle (degrees)	100 Hz

To facilitate collection of driving performance, a communication link between the smartphone and 3xD simulator was established. The link was built using a Raspberry Pi Model 3 hardware unit and some custom-built software written in Python programming language. The wired connection between the Pi and 3xD simulator was established using Ethernet, whereas Bluetooth Low Energy (BLE) was used for communication between Raspberry Pi and the smartphone. This setup allowed to request driving performance measures through 3xD's internal Hardware in the Loop (HIL) API in real-time. The setup also allowed to send various commands to the simulator e.g., a command from an external device to the simulator to initiate transition of vehicle control from automated to manual driving and vice versa.

3.2.3 Secondary task (N-Back)

During each automated driving period, participants were engaged in a secondary task – a visual prompt-verbal response "N-Back" task. The task allowed to induce various levels of visual, auditory, and cognitive load onto participants and measure subsequent effect onto their state during the automated to manual control transition period and fully manual driving afterwards. The "N-Back" is a good reference task for workload and is widely used in the driving context (Mehler, Reimer and Coughlin, 2012). Although, the task procedure was slightly modified to make it more suitable for automated driving context and promote higher attendance to this secondary task. Hence, an auditory cued, visual presentation, verbal response form of the "N-Back" task was adopted. The single-digit numbers, ranging from zero to nine, were displayed on the smartphone's screen for the duration of 2.25 seconds. The appearance of numbers was complemented by an auditory cue (i.e., a short 'beep' tone). Each test contained ten numbers, displayed in a random order. In total, four tests were performed in each scenario with 7 seconds' delay in-between tests.

A description of secondary task procedure was provided to each participant. To assist explanation, paper format visualisation of the "N-Back" interface was used. For the "N-Back 0" task variation participants were asked to verbally recall the number they see on the smartphone's screen i.e., if number "nine" was displayed, participant had to verbally recall "nine". In the "N-Back 1" task variation a number, which was displayed prior to a current number, had to be recalled. Subsequently, in the "N-Back 2" task variation a number, which was displayed two numbers prior to the current one, had to be recalled.

During the "N-Back" practice session participants experienced all three variations of the test, with a chance to practice each variation until were comfortable with it. The practice test was displayed to the participants using the smartphone, and the interface the same as they would later experience during the actual driving experiment. During the driving experiment a study facilitator listened to participants' verbal recalls of the numbers (using the audio feed from an in-vehicle embedded microphone) and noted numbers using pen and paper. These responses were later digitised for further analysis.

3.2.4 Driving scenario

In total, six driving activities were completed by each participant, of which three were in an Urban driving scenario at three different "N-Back" levels (subsequently labelled as U0 (e.g. Urban with "N-Back 0"), U1, and U2) and other three were in Motorway environments (labelled as M0, M1, and M2) (see **Figure 1**). In addition to environmental control variable, the "N-Back" difficulty was randomly selected for each environment (the complexity of the "N-Back" task is illustrated by the numeric value i.e., 0, 1, or 2, after the U or M in the above codes). Therefore, every participant experienced all three difficulties of "N-Back" in each road environment in a random order.

The driving activity design template is summarized in **Figure 4**. Each participant was exposed to both urban and motorway environments. Those environments were designed to mimic real-world driving conditions and differences in driving task complexity. For instance, the urban environment was aiming to replicate complexity and demand of real-world urban driving and, therefore, had high traffic density, 30 mph speed limit, abundance of auditory navigation commands, and large number of junctions and roundabouts. The motorway environment, on the other hand, was designed as to create low complexity and demand. Thus, no traffic was present, the speed limit was set to 70 mph, the auditory navigation commands were absent, and driving had to be performed in a straight line while keeping the vehicle in the nearside lane.

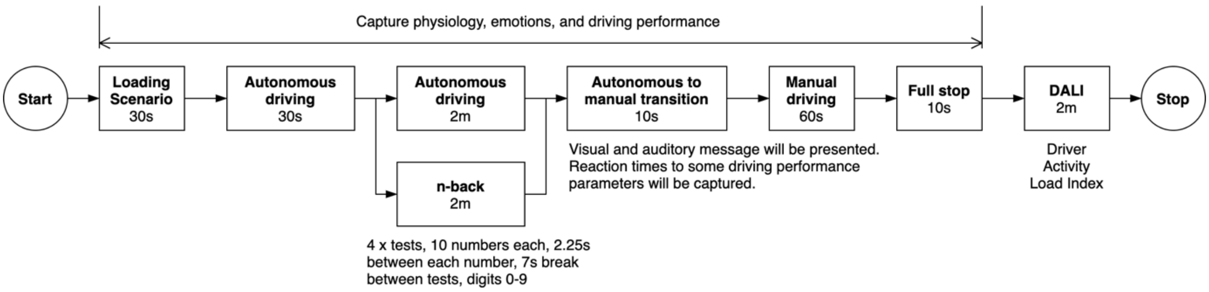


Figure 4. Driving activity design diagram, with the three main driving tasks highlighted (Autonomous, Transition and Manual driving)

As seen in **Figure 4**, each driving activity consisted of three distinct driving tasks, namely Autonomous driving (aka 'AI' in this paper), Transition from autonomous to manual driving (aka Transition), and a period of Manual driving following take-over (aka Manual). Every driving activity began with the vehicle being set to drive autonomously for approximately two and a half minutes. The "N-Back" task was initiated after 30 seconds. The simulated vehicle continued to drive in the automated mode for the next two minutes. During this period, participants were engaged in a secondary task i.e., responding to the "N-Back" task. In total, four "N-Back" sets with seven seconds' delay in-between were displayed. After the task sequence, a "transition of control" message was displayed for the fixed period of 10 seconds. It informed participants about the need to prepare themselves for an automated to manual transition of vehicle control, using a graphical cue of hands on the steering wheel and an embedded auditory message. Participants were instructed to wait until the end of animated countdown message for full manual vehicle control to be automatically released to them. The aim of this study was to evaluate the impact of workload on planned handover events, rather than emergency handover events, hence a time of 10 seconds and interface were decided in collaboration with industrial partners.

After a take-over, a minute-long period of manual driving followed, as previous research demonstrated that stabilisation of drivers' performance does not exceed 60 seconds (Merat *et al.*, 2014; Pampel *et al.*, 2018). During the period of manual driving, participants were asked to drive their vehicle as they would normally do on a real road. They were asked to follow all audio navigation commands, stay behind any vehicle driving in front of them, and remain within the posted speed limits. After the manual driving section, participants were informed that they may come to a stop. Right after, the driving activity was paused, and a study facilitator joined a participant inside the driving simulator in order to record drivers' subjective workload responses.

3.2.5 Subjective workload using DALI

At the end of each driving scenario, participants were asked to rate their subjective workload level by completing the Driving Activity Load Index (DALI) questionnaire (Pauzié, 2008). The questionnaire was supplied in pen-and-paper format, where effort of attention, visual demand, auditory demand, temporal demand, interference, and situational stress had to be rated in the scale from zero (i.e., very low) to twenty (i.e., very high). Participants were required to reflect on their overall experience throughout the driving scenario that is, from the time scenario was loaded, up until the vehicle reached a full stop. All participants were instructed to account for workload they have experienced during the automated mode, transition of control, and subsequent driving in the manual mode.

3.3 Data analysis procedure

Upon completion of all driving simulator experiments, all the data was synchronised and accumulated in a comma separated (.csv) data file. It should be noted that driving performance data was down sampled from original 100 Hz to 64 Hz to allow it to be synchronised to participants' physiological responses. A MATLAB script was composed to facilitate down-sampling. Furthermore, some raw driving performance measures e.g., vehicle position and steering angles, were used to calculate the range (or physical distance) between simulator and lead vehicles, lane position in reference to the centre of a lane, standard deviation of lane position (SDLP), and steering entropy. Those derivatives were calculated in accordance to SAE's operational definitions of driving performance measures and statistics (SAE International, 2013).

Next, a separate MATLAB script was written to facilitate labelling of scenario sections, filtering of heart rate signal, captured by Polar H10, and subsequent calculation of HRV metrics. Specifically, the time-domain HRV measures were calculated including, Root Mean Square of the Successive Differences (RMSSD) as well as percentage of RRs that exceed 50 milliseconds (p50NN) of 10- and 30-seconds' moving windows. Following on previous studies (Melnicuk, Birrell, Konstantopoulos, *et al.*, 2016; Melnicuk *et al.*, 2017), algorithms used to derive HRV were produced in accordance to standards described by Malik *et al.* (1996). As part of this publication the toolkit is released as a citeable, open source application and is accessible at https://github.com/vadimmelnicuk/WMGDSM.

Two separate data files were also produced that included digitised responses from demographics and DALI questionnaires. Once all the data was filtered, synchronised, and labelled, it was imported into IBM SPSS version 24 for an exploratory analysis. Firstly, descriptive statistics concerning demographics data were derived by means of frequency analysis. In addition, mean and standard deviation of participants' age was calculated. Next, "N-Back" responses were analysed i.e., percentages of non-responses and response errors were derived for each scenario type namely, U0, U1, U2, M0, M1, and M2. The median DALI scores were obtained next. Given the non-parametric nature of the DALI scores, they were further analysed using independent-samples Kruskal-Wallis test. This allowed to determine whether the overall DALI score and any specific DALI subcategories were significantly affected by workload differences due to changes in "N-Back" complexity.

An analysis of physiological responses was performed next. For all measures, means and standard deviations were obtained. The measures were also tested for an effect of workload differences due to variation of "N-Back" complexity using one-way analysis of variance (ANOVA).

The manual driving section was analysed further in isolation. All the measures including physiology and driving performance were binned into one-second intervals for the total duration of manual driving i.e., 60 seconds.

3.3.1 Novel method of signal stabilisation

Firstly, the data set was reduced so that it only consisted of time series data from the point of automated to manual control take-over, that is when the vehicle ceded control to the driver following the 10-seconds' warning period. Next, the signals of interest were binned into one-second intervals for the duration of manual driving i.e., 60 bins for 60 seconds of manual driving. Afterwards, the data file was imported into R Studio Version 1.1.383 for further analysis, with a script being produced to perform signal stabilisation analysis.

In summary the stabilisation method looked for a point in time after transition of control when the data for each dependant variable (e.g. HRV or lane position etc.) stopped changing (significantly) in comparison to the previous data points, and reached a stable plateau. This was calculated using normal linear regression, and a linear model that best fit the signal data (i.e., the model that yielded the highest R^2) was deemed to be the time that signal stabilised.

The R script binned all of the data from n+4 onwards into a single average data point. Using Figure 5 as an example to help explain the method, we see that for 'Bin 5' all of the data from 5 to 60 seconds was compressed into a single bin, giving a single mean value for all remaining 56 seconds of manual driving. For 'Bin 6' data from 6 to 60 seconds was again compressed into a single mean value. Finally, 'Bin 59' was the mean data from 59 and 60 seconds combined. It is worth reminding the reader that the driver took over manual control of the vehicle at 0 seconds, and the simulation stopped 60 seconds after transition of control from autonomous to manual driving.

As mentioned previously the point of stabilisation was when the fit of the linear model was at its peak, namely the highest R^2 value. After this point it was deemed that the next binned data point wasn't increasing in a linear manner anymore, but was stabilising or reaching a plateau. Using Figure 5 again to illustrate this, we see that the black circle marks the point where the linear model shows the best fit to the data, this was at Bin 15, or for the first 15 seconds after transition of control. Namely there was no improvement in the model fit when 16^{th} seconds of data was added and analysed, as after this point the data stops changing in a linear manner, i.e. it flattens off.

It should be noted that in order to determine if a signal of interest had stabilised (using this proposed method or any other) it must display one key characteristics. This is that a signal must deviate from, and then return to a baseline of 'normal' driving. For example a signal that continually rises (for example if the distance to the car in front keeps increasing for the remainder of the driving time), is cyclic in nature (if HR keeps going up and down), or does not deviate to start off with (if a driver maintains perfect speed control following handover), in these situations cannot be considered 'stable' at any point.

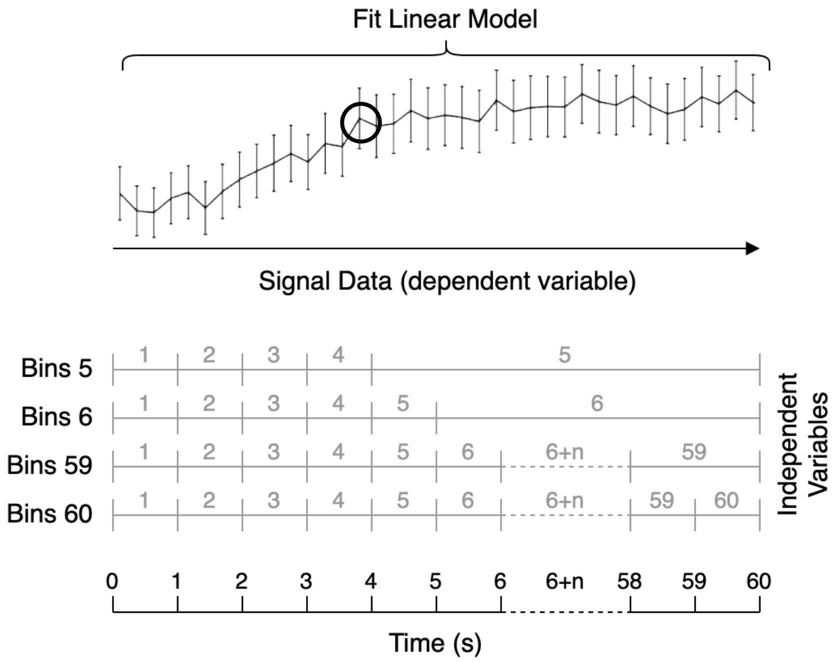


Figure 5. Signal stabilisation algorithm binning diagram.

4 Results

In total, 42 participants took part in this driving simulator experiment. However, four participants' data sets were excluded from the analysis: three exclusions due to technical reasons and one exclusion due to simulator sickness. Therefore, analysis was performed using data sample of 38 participants of mixed age (*Mean = 33.50, SD = 8.37*) and gender (see **Table 3**).

Table 3. Demographics descriptive statistics (N=38).

Character		**Frequency**	**Percent**
Gender	Male	26	68.4
	Female	12	31.6
	Other	0	0
	Prefer not to say	0	0
Age	21-25	5	13.2
	26-30	13	34.2
	31-35	8	21
	36-40	4	10.5
	41-45	2	5.3
	46-50	4	10.5
	50+	2	5.3
Occupation	Professional and managerial	26	68.4
	Clerical and sales	2	5.3
	Skilled and semi-skilled	1	2.6
	Student	9	23.7

4.1 N-Back responses

When it comes to recalling numbers during the "N-Back" task, participants were prone to make more recall errors when the task was more difficult (see **Figure 6**). As expected, percentage of no-responses was the highest during the "N-Back 2" task, in both urban and motorway environments. In the contrast, "N-Back 0" was less challenging for participants, with percentages of errors and no responses were low.

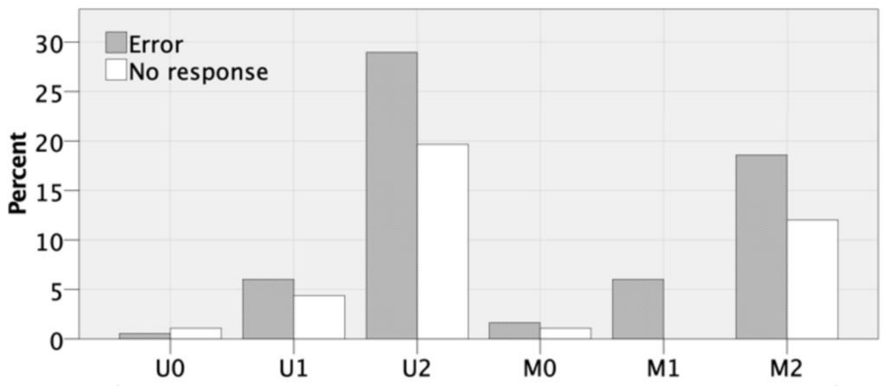


Figure 6. "N-Back" task percentage of errors and no responses from the total number of responses.

4.2 Subjective workload

Mirroring objective performance on the "N-Back" task, participants reported that it was more demanding to recall numbers that were further in the list (see **Figure 7**). The median DALI scores are steadily rising, in both motorway and urban environments as the difficulty of the "N-Back" increases.

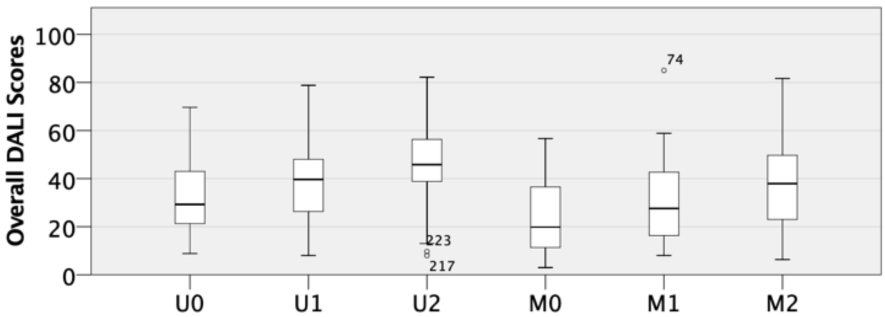


Figure 7. Median DALI overall scores categorised by the environment and difficulty of "N-Back" task.

Given the non-parametric nature of the DALI scores, they were further analysed using independent-samples Kruskal-Wallis test. The test revealed significant statistical score difference across scenarios in almost all DALI subcategories as well as the overall DALI scores in both urban (*H=11.335, p<0.01*) and motorway (*H=12.122, p<0.01*) environments. However, it was found that auditory demand scores were insignificant in both urban and motorway environments. Also, both visual demand and situational stress were scored insignificantly different in the urban environment (see **Table 4**).

Table 4. Kruskal-Wallis H and significance statistics for DALI scores. * = p<0.05, ** = p<0.01, *** = p<0.001, no asterisk = insignificant result.

Environment	Overall	Effort of attention	Visual demand	Auditory demand	Temporal demand	Interference	Situational stress
Urban	11.335**	11.213**	5.122	2.437	13.485**	7.871*	5.926

Motorway	12.122**	17.929***	6.408*	5.351	7.721*	6.772*	9.854**

4.3 Heart rate and its variability

With respect to the Heart Rate (HR) and time-domain Heart Rate Variability (HRV) measures, the differences due to the effect of workload were identified across various scenario sections and driving environments (e.g. Autonomous (AI), Transition and manual driving in **Figure 4**).

For instance, heart rate, represented using inter-beat-intervals (RR), was significantly different across all scenario sections (*One-way ANOVA, $F=42.224$, $p<0.001$*). Some individual statistical differences were also derived by separating data into environment types i.e., urban (U) and motorway (M), and scenario sections i.e., autonomous (AI), transition, and manual driving. During the autonomous mode heart rate was found to be significantly different in urban ($F=48.745$, $p<0.001$) and motorway ($F=46.636$, $p<0.001$) environments. Typically, the heart rate was found to be higher during the high workload induction in both urban and motorway environments. However, during the transition mode heart rate was found to be insignificantly different in both urban and motorway. Finally, during the manual mode heart rate was found to be significantly different in both, urban ($F=39.330$, $p<0.001$) and motorway ($F=40.803$, $p<0.001$) environments. The mean RR intervals are presented in **Figure 8**.

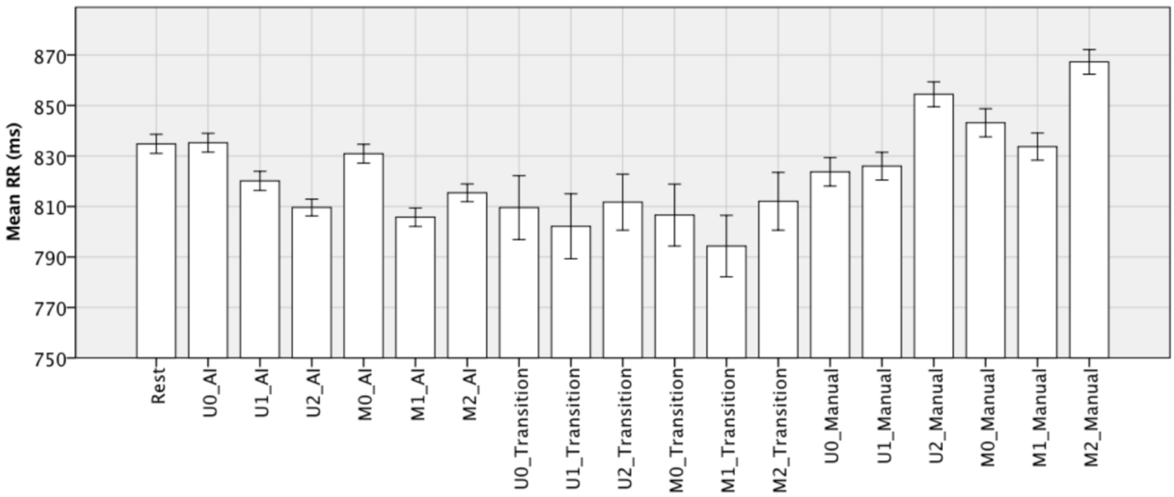


Figure 8. Mean inter-beat-intervals (RR), measured in milliseconds, categorised by various scenario sections. Error bars represent standard errors.

The HRV measures were analysed using one-way ANOVA to identify whether statistically significant effect of scenario sections is present. It should be noted that transition period was excluded from this analysis, since it only lasted for 10 seconds i.e., same as the lowest moving window duration for HRV (see **Figure 4**). As the result, two discrete scenario sections, autonomous and manual, were considered. It was found that RMSSD of 10 seconds' moving window ($F=24.363$, $p<0.001$), p50NN of 10 seconds' moving window (*F=45.033, p<0.001*) are significantly affected by scenario sections (see **Figure 9**).

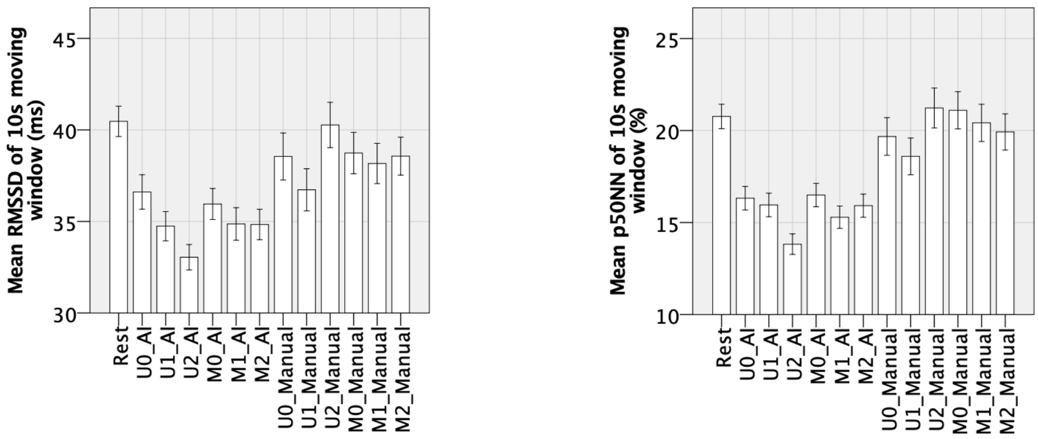


Figure 9. Mean RMSSD (left) and p10NN (right) of 10 seconds' moving window, measured in milliseconds, categorised by various scenario sections. Error bars represent standard errors

4.4 Driver state stabilisation following a transition of control to manual driving

This section presents results for stabilization time of physiological and driving performance that were derived using the new method described previously.

Firstly, the changes of heart rate, averaged across scenarios and conditions, over the period of manual driving were plotted (see **Figure 10**). Clearly, a concavity is present from the point where vehicle control was taken over. Thus, stabilisation times of heart rate were derived for the individual scenario types and for combination of all scenarios. It was found that it takes 20 seconds for the heart rate to stabilise after control of a vehicle was taken over following a period of automated driving (see **Figure 11**).

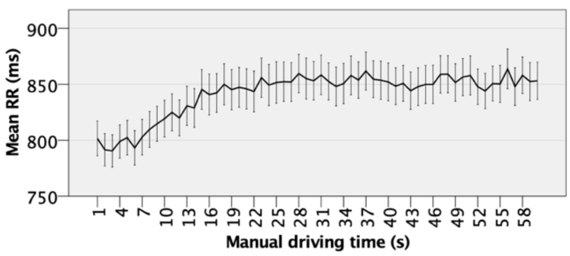


Figure 10. Cumulative mean heart rate binned into one-second periods for the duration of manual driving. Error bars represent standard errors.

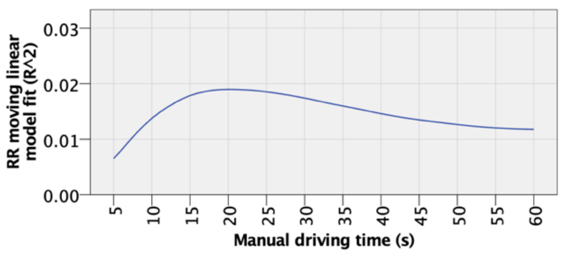


Figure 11. Stabilisation time of cumulative mean heart rate.

Interestingly, it was found that heart rate stabilisation time reduces following an increase in secondary task workload during the automated driving period in the motorway environment (see **Figure 12** and **Figure 13**). However, this pattern is not found in the urban environment where the highest heart rate stabilisation time was found to happen after engaging in "N-Back 1" task variation.

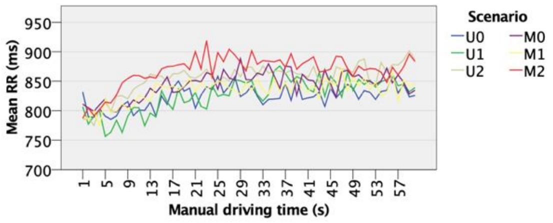
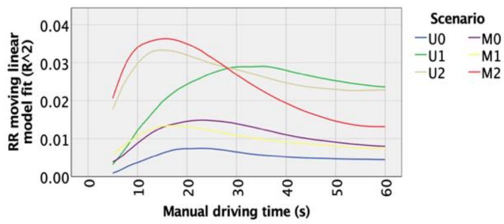


Figure 12. Mean heart rate binned into one-second periods for the duration of manual driving, categorised by the environment and difficulty of "N-Back" task.



U0	U1	U2	M0	M1	M2
23 s	36 s	15 s	23 s	18 s	16 s

Figure 13. Estimated mean stabilisation times of heart rate, categorised by environment and difficulty of "N-Back" task.

Next, the same method was applied to analyse time-domain HRV measures over the period of manual driving. It should be noted that HRV measures of 10 seconds' moving window were used in this analysis. As the result, HRV, which can be attributed solely to the manual driving period, begins from 10 seconds' mark. Hence, when extracting HRV stabilisation times, first 10 seconds of HRV data were trimmed for the analysis but added back onto the total stabilisation time.

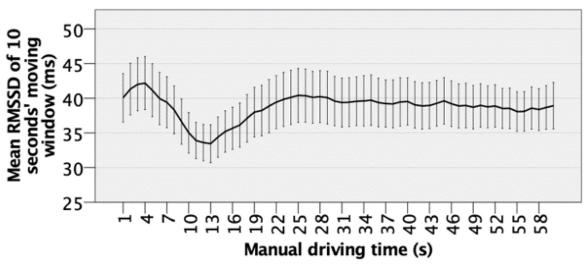


Figure 14. Cumulative mean RMSSD of 10 seconds' moving window binned into one-second periods for the duration of manual driving. Error bars represent standard errors.

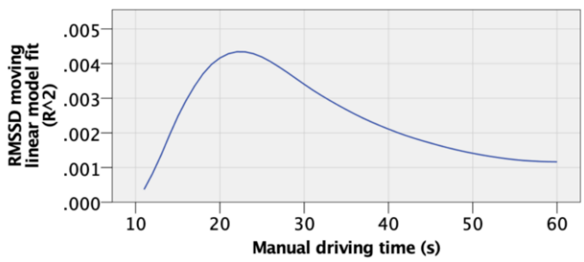


Figure 15. Stabilisation time of RMSSD of 10 seconds' moving window.

An upward linear trend in both RMSSD and p50NN of 10 seconds' moving window was visually identified (see **Figure 14** and **Figure 18**). Thus, stabilisation times for the combination of all scenarios were derived. It was found that it takes 23 seconds for the HRV to stabilise after control of a vehicle was taken over following a period of automated driving (see **Figure 15** and **Figure 19**). This stabilization time was consistent for both time-domain HRV measures i.e., RMSSD and p50NN. Furthermore, stabilisation times for individual scenarios were derived (see **Figure 17** and **Figure 21**).

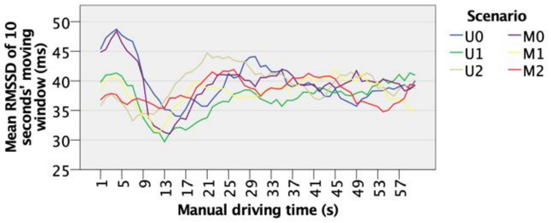
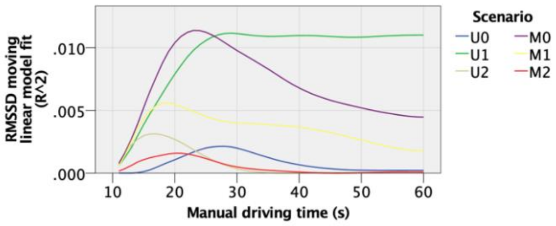


Figure 16. Mean RMSSD of 10 seconds' moving window binned into one-second periods for the duration of manual driving, categorised by environment and difficulty of "N-Back" task.



U0	U1	U2	M0	M1	M2
28 s	29 s	17 s	23 s	19 s	21 s

Figure 17. Stabilisation time of RMSSD of 10 seconds' moving window, categorised by environment and difficulty of "N-Back" task.

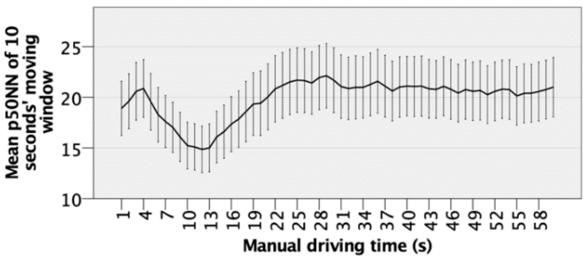


Figure 18. Cumulative mean p50NN of 10 seconds' moving window binned into one-second periods for the duration of manual driving. Error bars represent standard errors.

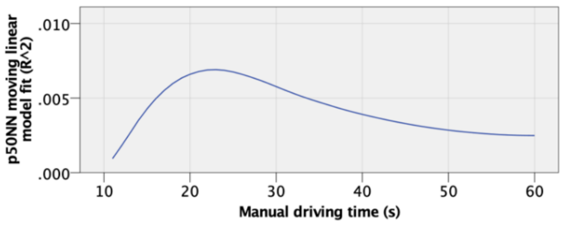


Figure 19. Stabilisation time of p50nn of 10 seconds' moving window.

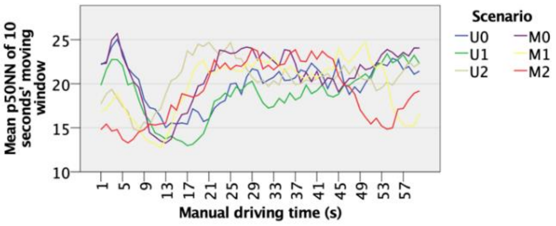
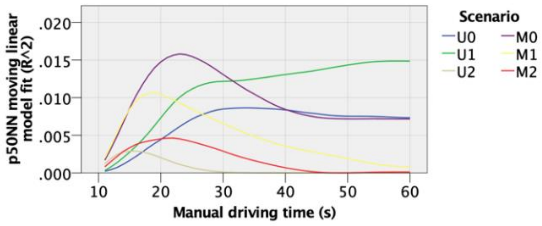


Figure 20. Mean p50NN of 10 seconds' moving window binned into one-second periods for the duration of manual driving, categorised by environment and difficulty of "N-Back" task.



U0	U1	U2	M0	M1	M2
34 s	-	16 s	23 s	19 s	22 s

Figure 21. Stabilisation time of p50NN of 10 seconds' moving window, categorised by environment and difficulty of "N-Back" task.

4.5 Driving performance

The driving performance was also affected by the differences in cognitive load elicited by the "N-Back" during autonomous driving. It should be noted that driving performance was only analysed in the context of motorway driving. The urban environment was excluded from this analysis due to variations in and complexities associated with simulated urban driving. These included speed-associated events within the scenario leading to take-over location differing dependant on driving speed. For these reasons only motorway driving will be discussed.

The mean speed during manual driving in the motorway environment was found to be significantly affected by the variation of workload during the autonomous driving (aka AI) sections (*F=432.695, p<0.001*). After the transition to manual driving speed was found to increase following the higher workload in the "AI" period (see **Figure 22**). Although, the mean difference in speed between M0 and M2 did not exceed 0.4 mph, the difference was significant. Consequently, the range i.e., distance to an in-front vehicle, was also found to be significantly affected by the variation of elicited workload during autonomous driving (*F=716.170, p<0.001*). Contrary to speed, the range after transition was found to decrease following an increase of prior workload in the "AI" period, with driver following 5 meters closer to the car in front between M0 and M2 (see **Figure 22**).

By visually assessing longitudinal performance metrics i.e., speed and range, it became evident that those signals do not stabilise up until the end of manual driving section. Thus, no attempts were made to derive stabilisation times for either speed or range.

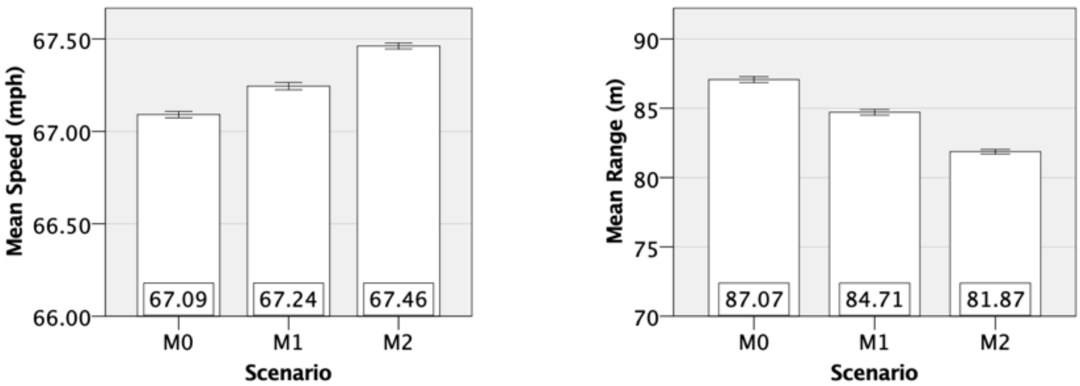


Figure 22. Mean speed (left) and range (right), measured in mph and meters respectively, during manual driving in the motorway environment, categorised by difficulty of "N-Back" task. Error bars represent standard errors.

The measures of lane position were analysed showing that the mean lane position is significantly affected by the variation of workload during autonomous (AI) driving (*F=340.276, p<0.001*). Specifically, it was found to decrease (i.e., driving closer to the centre of the lane) following an increase of prior workload in the "AI" period (approximately 30 millimetres' difference between mean lane positions in M0 and M2) (see **Figure 23**). Similarly, standard deviation of lane position was found to increase following an increase of prior workload in the "AI" period i.e., lane position has deviated by more than 40 millimetres between M2 and M0 scenarios (see **Figure 23**).

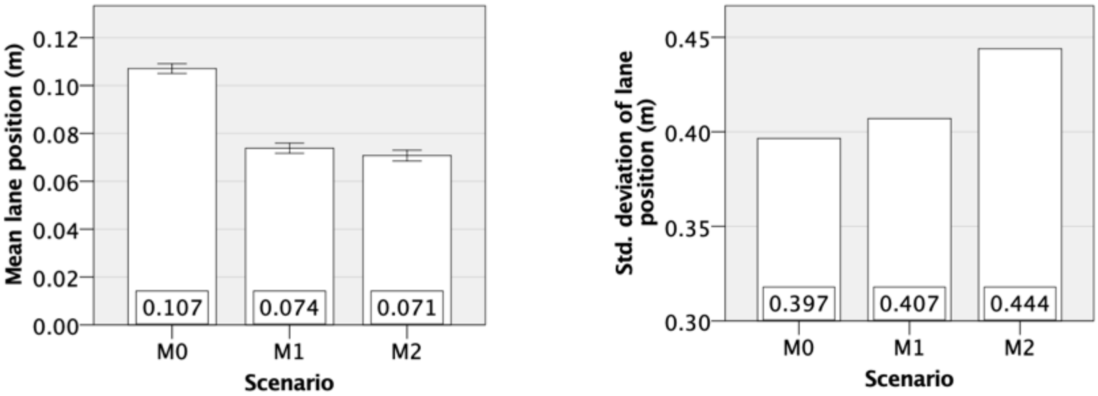


Figure 23. Mean (left) and Std. Deviation (right) of lane position, measured in meters, during manual driving in the motorway environment, categorised by difficulty of "N-Back" task. Error bars represent standard errors.

In addition, stabilisation times for mean lane position and standard deviation of lane position were derived. However, first 10 seconds were trimmed from the mean lane position, and 3 seconds from standard deviation of lane position stabilisation calculations in order to remove an effect of initial steering compensations (caused by gripping the steering wheel and correction for perceived centre of the lane). It was found that the time it takes for lane position and standard deviation of lane position to stabilise increases following an increase of prior workload induced by the "N-Back" during autonomous driving (see **Figure 25** and **Figure 27**). However, it should be noted that the method used to derive stabilisation times for standard deviation of lane position could not determine exact stabilisation times for M0 and M1 scenarios. Perhaps, this is due to stabilisation appearing in the first five seconds of driving in those instances and the initial signal trend could not be established.

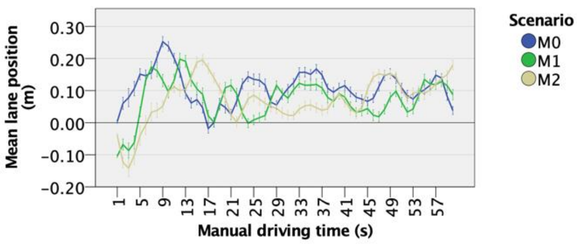
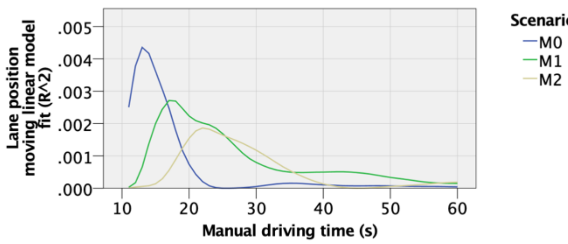


Figure 24. Mean lane position, measured in meters, binned into one-second periods for the duration of manual driving in the motorway environment, categorised by difficulty of "N-Back" task. Error bars represent standard errors.



M0	M1	M2
13 s	17 s	22 s

Figure 25. Stabilisation times of mean lane position, categorised by difficulty of "N-Back" task.

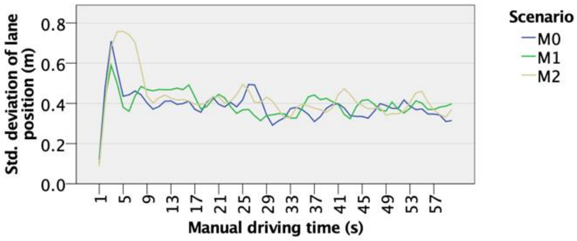
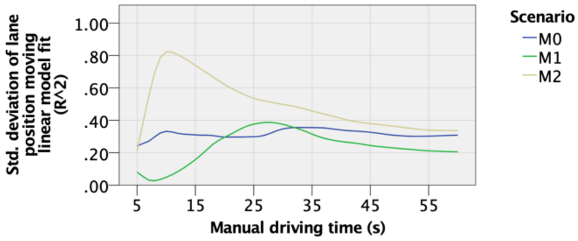


Figure 26. Std. deviation of lane position, measured in meters, binned into one-second periods for the duration of manual driving in the motorway environment, categorised by difficulty of “N-Back” task.



M0	M1	M2
Under 5 s	Under 5 s	10 s

Figure 27. Stabilisation times of std. deviation of lane position, categorised by difficulty of “N-Back” task.

5 Discussion

The goal of the present study was to assess the effect of cognitive load on drivers' ability to transition from autonomous to manual driving. To this end, we used a range of behavioural and physiological measures, all collected during several simulated driving scenarios. We introduce the new stabilization method for identifying a point at which physiological markers of cognitive load and driving indices return to baseline following driver's transition. Key results from this paper suggest that non-optimal levels of workload during the automated driving conditions impair driving performance after the driver takes back control for approximately 20 seconds, especially lateral control of the vehicle, and the magnitude of this impairment varied with increasing cognitive load. In addition, heart rate parameters take between 20 and 30 seconds to stabilise following a take-over request.

5.1 Workload induction

All participants were exposed to the variable level of cognitive load during the automated driving section. An induction of workload was accomplished using visual prompt-verbal response version of the "N-Back" task (Mehler, Reimer and Coughlin, 2012). As expected, the percentage of failures to respond and errors was positively associated with the "N-Back" difficulty. It should be noted that "N-Back" performance in the urban environment deteriorated to a higher extent i.e., 18.6% of false "N-Back 2" responses in the motorway against 29% in the urban environment. Moreover, percentage of no responses during the "N-Back 2" task followed a similar pattern. Thus, it can be concluded that urban, being a more complex environment, required participants to compensate for an increased demand due to presence of mild traffic, curved road layout, and other distractors.

The level of workload during the automated driving section was the primary scenario control. It was tested for a significant effect using DALI, which revealed significant statistical difference across scenarios in almost all DALI subcategories as well as the overall DALI scores. Therefore, it can be concluded that the chosen study design allowed to induce significantly different amount of cognitive load in drivers prior to them taking control of the simulator vehicle.

5.2 Automated driving section

First, it was found that heart rate (represented using RR inter-beat-intervals and captured over the period of two minutes) significantly increases following an increase of workload (measured subjectively) in the urban environment. Previously, an increase of heart rate was found to be associated with elevated cognitive load (Brookhuis and de Waard, 2010). Moreover, this phenomenon is in line with the findings of Mehler *et al.* (2012) and Gable *et al.* (2015). However, in the motorway environment, a typical relationship between heart rate and workload could not be identified. Instead, the higher mean heart rate (hence shorter inter-beat-interval responses) were identified during the period of engagement in the "N-Back 1" task in the automated motorway section.

Furthermore, heart rate variability was found to follow a well-established pattern in the urban automated driving section i.e., inverse relationship between time-domain RMSSD and workload level; it was previously found that increased task complexity causes HRV to decrease (Hoover *et al.*, 2012). However, similar to mean heart rate response in the motorway section, HRV did not impeccably follow an expected pattern in the motorway automated driving section. Despite an overall significant difference for the effect of HRV to decrease as complexity increases, p50NN differences were not observed between "N-Back" 1 and 2 during motorway driving. This lack of effect of increasing workload on HRV during motorway driving, may be as a result of the participants experiencing underload, or low levels of cognitive load during this aspect of the driving task. With the "N-Back" task in turn actually increasing load to 'optimal' levels during motorway driving, this in turn may actually have positive implications for driver state and hence driving safety.

It therefore follows that heart rate and time-domain HRV responses, which are captured using consumer grade electronic devices, can be used to evaluate and quantify the level of workload that drivers are exposed to during the periods of automated driving. Therefore, Hypothesis I, which states that a level of cognitive load can be reliably estimated using driver state indicators (e.g., HR and HRV) during the automated driving period, can be in part supported by these results with the caveat that this was consistently observed for comparisons between low ("N-Back" 0) and high ("N-Back" 2) levels of induced workload.

Moreover, it was demonstrated how some of the physiological indicators i.e., heart rate and HRV, respond to changes in workload. Thus, the first part of Hypothesis III, which states that a varying complexity level of induced cognitive load causes significantly different responses in driver state indicators during the automated driving, can be also supported.

5.3 Automated to manual control transition

It was previously argued that it can be beneficial for vehicles capable of SAE Level 3 automation (meaning they share the responsibility of driving between the vehicle and the driver) to estimate level of drivers' cognitive load in order to ensure safe and comfortable transition of control. However, there is currently limited evidence informing the effect of workload on drivers' ability to effectively take-over manual control of a vehicle after a prolonged period of automated driving.

As part of this study participants had 10 seconds to prepare themselves to take over full manual control of the vehicle – the focus of this paper was not unexpected, emergency handover events. Physiological responses during this planned handover period in both urban and motorway driving were analysed, but it was found that none of the heart rate related physiology responses are reactive to potential changes in drivers' state during this short period of transition. Moreover, time-domain HRV cannot be used as an indicator of workload during such a short measurement period, because it is derived using at least 10 seconds' long moving window. However, measures such as skin conductance (or sweat rate) and executive function (measured using functional near-infrared spectroscopy, fNIRS) have been shown to be sensitive to handover events (Perelló March et al., In Press).

It can be concluded that heart rate derived physiological responses are unlikely to be useful for assessing quality or success of automated to manual control transition periods. Thus, the second part of Hypothesis III, which states that a varying complexity level of induced cognitive load causes significantly different responses in driver state indicators during the transition of control period, should be rejected. Instead, researchers should continue to use reaction times or eye glance behaviour as a measure of quality of automated to manual control transition routines that last for 10 seconds or less. Perhaps, heart rate derived physiological responses can be adopted for studying quality of automated to manual control transition that exceed 10 seconds' time period. These results cannot comment on the efficacy of heart rate derived physiological responses as a measure for emergency, or unexpected handovers, this would be for assessment in future research.

5.4 Manual driving following a take-over of vehicle control

It is also essential to understand how workload may affect drivers' ability to perform the primary task of driving during the first minute. It was hypothesised that non-optimal levels of workload may have a significant effect on drivers' state indicators following a control take-over routine (the third part of Hypothesis III).

Indeed, it was found that mean heart rate significantly differs during the manual driving period in both urban and motorway environments. Surprisingly, mean heart rate was the lowest following a high complexity "N-Back 2" task. This adds to the debate of heart rate relevance in assessing quality of automated to manual control transition routines since, sometimes results can be inconclusive. Conversely it could be implied that drivers will limit their engagement in high workload secondary tasks in order to maintain vehicle monitoring during automated driving. As observed in this study with the increase number of false and missed responses to "N-Back 2" in the Urban driving scenario.

HRV was only significantly affected during driving in the urban environment. According to HRV, participants have experienced significantly more workload during the manual driving after an "N-Back 1" task, compared to estimated workload after "N-Back" 0 and 2. This could be attributed to the fact that drivers have likely compensated for the lower workload (i.e., "N-Back 0") and the higher workload (i.e., "N-Back 2") by reducing their cognitive capacity as described by Brookhuis and de Waard (2010).

To further deepen our understanding of drivers' state during the manual driving period, some of the physiological indicators were binned into second-long intervals for the whole duration of manual driving section i.e., 60 seconds. This allowed to visualise an extent of signal changes and determine their stabilisation times. Previously, attempts were made to determine stabilisation times of driving performance measures in the SAE Level 3 control transition scenarios (Merat *et al.*, 2014; Pampel *et al.*, 2018). Those studies used 5-second-long bins and used ANOVA to determine whether measurement variance get significantly affected by the time factor. Whereas, this study adopted a new method which relies on second-long intervals for better precision in determining stabilisation times. The method allows to determine exact location of signal stabilisation following a prior up or down trend. The point is determined by the best linear model fit.

The mean heart rate responses were analysed first using this new method to determine stabilisation. It was found that a distinct raising linear trend is present at the beginning of manual driving period. Using the proposed method for deriving stabilisation time of a signal, it was found that, on average, it takes 20 seconds for heart rate to stabilise after manual control of a vehicle is taken. Moreover, it was found that stabilisation time tends to decrease following an increase in workload prior to control transition in the motorway environment. However, this does not apply to the urban environment, where mean heart rate stabilisation times could be affected by complexity and inconsistency associated with urban driving, rather than solely by induction of various levels of workload during the automated driving section.

Importantly, time-domain HRV measures were also found to be affected by workload. It took, on average over all six scenarios and workload repeated trial runs, 23 seconds for HRV to stabilise following transition to manual driving. Moreover, stabilisation times during the motorway environment did not differ due to variation of prior workload. Similar to heart rate, HRV stabilisation times in the urban scenario are likely affected by complexity and inconsistency, rather than solely by induction of various levels of workload during the automated driving section. However, a distinct pattern can be seen in the HRV signal for the first 5 seconds of manual driving i.e., the HRV remains higher after engaging in a low complexity task in both urban and motorway driving scenarios. Since HRV was calculated using 10 seconds' moving window, the stabilisation times were derived from 10 seconds onwards, which was in fact a point of the lowest HRV in all scenarios.

Given the discussion points, raised above, the Hypothesis V can be also supported. It states that the time it takes for drivers' state to stabilise following a control take-over is impacted by the amount of cognitive load, experienced during the automated driving period.

5.5 Effect on the driving performance

Next, the driving performance measures in the motorway environment were evaluated. Similar to velocity observations made by Mehler *et al.*, (2009), it was found that mean speed and range in the motorway environment was found to be significantly affected by the variation of prior workload. Although, those differences did not exceed

0.3 mph and 3 meters respectively. Therefore, it could be suggested that these phenomena are not safety critical.

When it comes to the mean lane position analysis, it was found that it is significantly affected by a variation of prior workload. Specifically, it was found that participants tend to drive closer to the "hard shoulder" by approximately 0.03 meters after they take-over vehicle control following a task of low demand i.e., "N-Back 0". Similar to differences in the longitudinal metrics, this difference in not a safety critical impairment. It is in fact below a typical daytime driving offset of 0.05 meters (Sayer *et al.*, 2010). On the other hand, Standard Deviation of Lane Position (SDLP) was found to increase due to a significant increase in workload prior to take-over of manual control. This impairment could be deemed as detrimental to safety, since all SDLP values exceed the typical threshold of normal driving i.e., from 0.2 to 0.3 meters (Green *et al.*, 2004). Therefore, it can be concluded that a varying complexity level of induced cognitive load causes significant impairment in driving performance after an automated to manual control take-over, especially, in the lateral control. Hence, Hypothesis II, can be also supported, but only in regard to the lateral control.

To further deepen our understanding of driving performance during the manual driving period, the stabilisation times of some lateral control measures were derived. Specifically, it was found that time it takes for lane position and SDLP to stabilise increases following an increase of prior workload. However, contrary to findings of Merat *et al.*, (2014), which presented a steady raise of SDLP reaching 0.2 meters in the first 20 seconds, our findings show SDLP peaking in the first 10 seconds (almost reaching 0.8 meters) and stabilise between 0.5 and 0.3 meters thereafter. Furthermore, it took even longer for the mean lane position to stabilise depending on the level of prior workload, from 13 ("N-Back 0") to 22 ("N-Back 2") seconds. Hence, the Hypothesis IV, which state that the time it takes for driving performance to stabilise following a control take-over will be impacted by the amount of workload, experienced during the automated driving period, can be also supported in the context of lateral control, since none of the longitudinal control measures have actually stabilised in the first 60 seconds.

6 Conclusion

The aim of this study was to explore an effect of cognitive workload on drivers' state and driving performance during a planned automated to manual control transition scenarios. In the past, a small number of studies have attempted to evaluate quality of control take-over routines, predominantly, using driver behaviour and driving performance indicators e.g., reaction times and lane position. However, there is a lack of evidence on how various levels of workload, induced during the automated driving period, may affect drivers' state during the period of control transition and the period of manual driving that follows. This knowledge gap is of high importance, given the safety critical nature of automated to manual control take-over routines. To address this knowledge gap, the study adopted repeated measure design and was performed in a highly immersive driving simulator. The results suggest that non-optimal levels of workload during the automated driving conditions could impair driving performance, especially, the lateral control. Furthermore, it was demonstrated that level of workload can impact severity and duration of driving performance impairment. For instance, mean and standard deviation of lane position impairments were found to last longer following a higher level of workload during over the automated driving period. Finally, it was demonstrated how workload level can be estimated in context of automated driving using the physiological measures, captured by means consumer grade electronic devices. The study also discussed an impact the key finding may have on the design of SAE Level 3 systems. Relevant suggestions were provided for the research community and automakers that are working on implementing future vehicles that are capable of SAE Level 3 automation.

7 Funding

The financial support for this work was provided by Engineering & Physical Sciences Research Council (EPSRC) and Jaguar Land Rover (JLR). We would like to thanks Dr Lukasz Walasek of University of Warwick for his support with the development of the stabilisation methodology, and the blind reviewers whose feedback enhanced the clarity and impact of the paper.

8 References

Brookhuis, K. A. and de Waard, D. (2010) 'Monitoring drivers' mental workload in driving simulators using physiological measures', *Accident Analysis and Prevention*. Elsevier Ltd, 42(3), pp. 898–903. doi: 10.1016/j.aap.2009.06.001.

Burghardt, S., Weig, F. and Choi, S. (2017) *Mobility trends: What's ahead for automotive semiconductors*, *McKinsey & Company*. Available at: https://www.mckinsey.com/industries/semiconductors/our-insights/mobility-trends-whats-ahead-for-automotive-semiconductors (Accessed: 1 July 2018).

Campbell, J. L. *et al.* (2018) *Human factors design principles for level 2 and level 3 automated driving concepts (Report No. DOT HS 812 555)*. Washington, DC: National Highway Traffic Safety Administration.

Clark, H. *et al.* (2017) 'Performance in takeover and characteristics of non-driving related tasks during highly automated driving in younger and older drivers', *Proceedings of the Human Factors and Ergonomics Society*, 2017-Octob(September), pp. 37–41. doi: 10.1177/1541931213601504.

Cunningham, M. and Regan, M. (2017) 'Driver Distraction and Inattention in the Realm of Automated Driving', *IET Intelligent Transport Systems*, (December 2017). doi: 10.1049/iet-its.2017.0232.

Gable, T. M. *et al.* (2015) 'Comparing heart rate and pupil size as objective measures of workload in the driving context', in *Adjunct Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '15*. New York, New York, USA: ACM Press, pp. 20–25. doi: 10.1145/2809730.2809745.

Gastaldi, M., Rossi, R. and Gecchele, G. (2014) 'Effects of driver task-related fatigue on driving performance', *Procedia - Social and Behavioral Sciences*. Elsevier B.V., 111, pp. 955–964. doi: 10.1016/j.sbspro.2014.01.130.

Green, P. *et al.* (2004) *SAVE-IT Typical Values for Driving Performance of Lane Position : Prepared by*.

He, Q., Li, W. and Fan, X. (2011) 'Estimation of Driver's Fatigue Based on Steering Wheel Angle', in *Engineering Psychology and Cognitive Ergonomics*, pp. 145–155. doi: 10.1007/978-3-642-21741-

8_17.

Hoover, A. *et al.* (2012) 'Real-time detection of workload changes using heart rate variability', *Biomedical Signal Processing and Control*. Elsevier Ltd, 7(4), pp. 333–341. doi: 10.1016/j.bspc.2011.07.004.

Horberry, T. *et al.* (2006) 'Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance', *Accident Analysis and Prevention*, 38(1), pp. 185–191. doi: 10.1016/j.aap.2005.09.007.

Inagaki, T. (2003) 'Adaptive Automation: Sharing and Trading of Control', in *Handbook of Cognitive Task Design*, pp. 147–169. doi: 10.1201/9781410607775.

Kersloot, T., Flint, A. and Parkes, A. (2003) 'Steering Entropy as a Measure of Impairment', *Presented during the Young Researchers …*, 44(0). Available at: http://www.ectri.org/YRS03/Session-6/Flint.pdf.

Khastgir, S., Birrell, S., Dhadyalla, G. & Jennings, P. (2018) 'Calibrating trust through knowledge: Introducing the concept of informed safety for automation in vehicles', Transportation Research Part C: Emerging Technologies. 96, 290-303.

Krajewski, J. *et al.* (2009) 'Steering wheel behavior based estimation of fatigue', *Proceedings of the Fifth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, pp. 118–124. doi: 10.17077/drivingassessment.1311.

Lee, C. *et al.* (2020) 'Consumer Comfort with Vehicle Automation: Changes Over Time', pp. 412–418. doi: 10.17077/drivingassessment.1726.

Malik, M. *et al.* (1996) 'Heart rate variability. Standards of measurement, physiological interpretation, and clinical use', *European Heart Journal*, pp. 354–381.

Matthews, G. *et al.* (2011) 'Driving in States of Fatigue or Stress', in *Handbook of Driving Simulation for Engineering, Medicine, and Psychology*, pp. 389–394. doi: 10.1201/b10836-30.

Matthews, G. and Desmond, P. A. (2002) 'Task-induced fatigue states and simultated driving performance', *The Quarterly Journal of Experimental Psychology*, 55A(2), pp. 659–686. doi: 10.1080/0272498014300050.

Mehler, B. *et al.* (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), pp. 6–12. doi: 10.3141/2138-02.

Mehler, B., Reimer, B. and Coughlin, J. F. (2012) 'Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task: An on-road study across three age groups', *Human Factors*, 54(3), pp. 396–412. doi: 10.1177/0018720812442086.

Melnicuk, V., Birrell, S., Konstantopoulos, P., *et al.* (2016) 'JLR heart: Employing wearable technology in non-intrusive driver state monitoring. Preliminary study', in *2016 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, pp. 55–60. doi: 10.1109/IVS.2016.7535364.

Melnicuk, V., Birrell, S., Crundall, E., *et al.* (2016) 'Towards hybrid driver state monitoring: Review, future perspectives and the role of consumer electronics', in *2016 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, pp. 1392–1397. doi: 10.1109/IVS.2016.7535572.

Melnicuk, V. *et al.* (2017) 'Employing consumer electronic devices in physiological and emotional evaluation of common driving activities', in *2017 IEEE Intelligent Vehicles Symposium (IV)*. Los Angeles: IEEE, pp. 1529–1534. doi: 10.1109/IVS.2017.7995926.

Merat, N. *et al.* (2012) 'Highly automated driving, secondary task performance, and driver state', *Human Factors*, 54(5), pp. 762–771. doi: 10.1177/0018720812442087.

Merat, N. *et al.* (2014) 'Transition to manual: Driver behaviour when resuming control from a highly automated vehicle', *Transportation Research Part F: Traffic Psychology and Behaviour*, 27(PB), pp. 274–282. doi: 10.1016/j.trf.2014.09.005.

Nakayama, O. *et al.* (1999) 'Development of a steering entropy method for evaluating driver workload', *Proceedings of JSAE Annual Congress*, (724), pp. 5–8. doi: 10.4271/1999-01-0892.

Pampel, S. M. *et al.* (2018) 'Getting the Driver back into the Loop : The Quality of manual Vehicle Control following long and short non-critical Transfer-of-control Requests : TI : NS', *Theoretical Issues in Ergonomics Science*, (In Press)(April).

Pauzié, A. (2008) ‘A method to assess the driver mental workload: The driving activity load index (DALI)’, *IET Intelligent Transport Systems*, 2(4), p. 315. doi: 10.1049/iet-its:20080023.

Paxion, J., Galy, E. and Berthelon, C. (2014) ‘Mental workload and driving’, *Frontiers in Psychology*, 5(December), pp. 1–11. doi: 10.3389/fpsyg.2014.01344.

Perelló March, J., Burns, C., Woodman, R., Elliott, M. and Birrell, S. (In Press) 'Driver State Monitoring: Manipulating Reliability Expectations in Simulated Automated Driving Scenarios', *IEEE Transactions on Intelligent Transportation Systems*. In Press

Philip, P. *et al.* (2005) ‘Fatigue, Sleepiness, and Performance in Simulated Versus Real Driving Conditions’, *Sleep*, 28(12), pp. 1511–1516. doi: 10.1093/sleep/28.12.1511.

Polar Electro (2018) *Polar H10 heart rate sensor*. Available at: https://www.polar.com/uk-en/products/accessories//h10_heart_rate_sensor.

SAE International (2013) *Operational Definitions of Driving Performance Measures and Statistics*.

SAE International (2015) ‘Guidelines for Safe On-Road Testing of SAE Level 3, 4, and 5 Prototype Automated Driving Systems (ADS)’.

Sage, A., Bellon, T. and Carey, N. (2018) *Self-driving car industry confronts trust issues after Uber crash*, *Reuters*. Available at: https://uk.reuters.com/article/uk-autos-selfdriving-uber-trust/self-driving-car-industry-confronts-trust-issues-after-uber-crash-idUKKBN1GY26A (Accessed: 1 July 2018).

Saxby, D. J. *et al.* (2013) ‘Active and passive fatigue in simulated driving: Discriminating styles of workload regulation and their safety impacts.’, *Journal of Experimental Psychology: Applied*, 19(4), pp. 287–300. doi: 10.1037/a0034386.

Sayer, J. *et al.* (2010) *Integrated Vehicle-Based Safety Systems Heavy-Truck Field Operational Test Key Findings Report*, *Nhthsalas201.Nhtsa.Dot.Gov*. Available at: http://nhthsalas201.nhtsa.dot.gov/DOT/NHTSA/NVS/Crash Avoidance/Technical Publications/2011/811464.pdf.

Schaap, N. *et al.* (2013) ‘The relationship between driver distraction and mental workload’, in *Driver distraction and inattention: advances in research and countermeasures*. CRC Press, pp. 63–82. Available at: https://research.utwente.nl/en/publications/the-relationship-between-driver-distraction-and-mental-workload-2.

Schoettle, B. and Sivak, M. (2014) *Public Opinion About Self-Driving Vehicles in China, India, Japan, The U.S., The U.K. and Australia*. doi: UMTRI-2014-30.

Ulahannan, A., Jennings, P., Oliveira, L. & Birrell, S. (2020) 'Designing an Adaptive Interface: Using Eye Tracking to Classify How Information Usage Changes Over Time in Partially Automated Vehicles', *IEEE Access*. 8, 16865-16875.

Vogelpohl, T. *et al.* (2018) ‘Transitioning to manual driving requires additional time after automation deactivation’, *Transportation Research Part F: Traffic Psychology and Behaviour*. Elsevier Ltd, 55, pp. 464–482. doi: 10.1016/j.trf.2018.03.019.

Wickens, C. D. and Hollands, J. G. (2000) *Engineering Psychology and Human Performance*, *Engineering psychology and human performance*. doi: 10.1146/annurev.ps.27.020176.001513.

Worstall, T. (2015) *9% Of Cars To Be Driverless Cars By 2035 - Forbes*, *Reuters*. Available at: goo.gl/NXfnvO (Accessed: 23 December 2015).

Zeeb, K., Buchner, A. and 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*. Elsevier Ltd, 92, pp. 230–239. doi: 10.1016/j.aap.2016.04.002.