Employing Consumer Electronic Devices in Physiological and Emotional Evaluation of Common Driving Activities*

Vadim Melnicuk, Stewart Birrell, Elizabeth Crundall and Paul Jennings

Abstract— **It is important to equip future vehicles with an on-board system capable of tracking and analysing driver state in real-time in order to mitigate the risk of human error occurrence in manual or semi-autonomous driving. This study aims to provide some supporting evidence for adoption of consumer grade electronic devices in driver state monitoring. The study adopted repeated measure design and was performed in high-fidelity driving simulator. Total of 39 participants of mixed age and gender have taken part in the user trials. The mobile application was developed to demonstrate how a mobile device can act as a host for a driver state monitoring system, support connectivity, synchronisation, and storage of driver state related measures from multiple devices. The results of this study showed that multiple physiological measures, sourced from consumer grade electronic devices, can be used to successfully distinguish task complexities across common driving activities. For instance, galvanic skin response and some heart rate derivatives were found to be correlated to overall subjective workload ratings. Furthermore, emotions were captured and showed to be affected by extreme driving situations.**

I. Introduction

Despite recent advances in driver assistance and in-car safety systems the risk of serious accidents on the roads is still present, with between four and five people per day being killed on UK roads [1]. According to NHTSA [2], up to 94% of their observed accidents have occurred due to presence of human error that is, a car crash critical reason attributed to a driver. The cause of these errors is often an impaired mental and physical state of a driver, including fatigue [3], high level of workload [4], and distraction [5]. Similarly, presence of extensive emotions, especially anger, can make a driver prone to errors and could lead to dangerous driving [6]. Joy or happiness, on the other hand, could be beneficial to driving [7]. Therefore, it is important to equip future vehicles with an on-board system capable of tracking and analysing driver state to be able to mitigate the risk of incorrect vehicle control prior to its actual occurrence, which subsequently has a potential to reduce the number of human error related road accidents. Such feature is often referred to as Driver State Monitoring (DSM) system. Transportation consultancy firm Frost & Sullivan have forecasted that DSM is expected to become a standard passenger car feature by 2025 [8].

This study aimed to demonstrate how several consumer grade electronic devices can be employed in physiological and emotional evaluation of common driving activities. The mobile DSM system, consisting of a wristband, chest heart monitor, and a smartphone, was developed. It was used to capture several physiological indicators and evaluate driver's emotional state in highly immersive simulated driving environment. It was hypothesised that significant differences in drivers' state can be detected on road sections of distinct task complexities, such as urban, rural, and motorway. In addition to common road types, fully autonomous driving and racing sections were simulated. Furthermore, several extreme traffic situations were present including rapidly decelerating an in-front vehicle and unexpected crossing pedestrians. These situations were expected to produce distinct physiological and emotional responses.

II. Background

DSM systems mostly rely on objective driver state evaluation that is, estimation of drivers' cognitive distraction, workload, or fatigue through physiological assessment of a driver. It was found that continuous assessment of various physiological responses of human body such as, Heart Rate (HR), Heart Rate Variability (HRV), Galvanic Skin Response (GSR), and Peripheral Skin Temperature (PST), can provide a good indication of distraction, fatigue and workload levels in real-time [9], [10]. Furthermore, facial expressions, captured in accordance to Facial Action Coding System (FACS), were found to be a good predictor of emotional state, including seven basic emotions namely, happiness, sadness, surprise, fear, anger, disgust, and contempt [11].

The majority of previous studies have adopted expensive and highly complex medical grade devices to capture drivers' physiological responses, whereas emotions were derived by manually coding videos of participants' facial expressions. These studies helped to establish state of the art as well as understand the links between psychophysiological patterns and drivers' behaviour. Additionally, subjective methods were used to validate specific physiological patterns for example, decrease of heart rate variability due to increase of cognitive workload [12]. However, it is unlikely that medical equipment or subjective assessment will become part of commercial DSM solutions, because of specialised nature and cost implication of former and potential invasiveness of later. Therefore, the cost-effective and user-acceptable DSM systems need to be designed to ensure initial uptake of these systems by potential customers. A solution could incorporate recent advances of sensory technology embedded into wearable devices, such as fitness wristbands and smartwatches. These sensors can offer reliable and mobile physiological tracking. Furthermore, consumer grade

*Research supported by Engineering and Physical Sciences Research Council (EPSRC) and Jaguar Land Rover (JLR)

V. Melnicuk is with WMG, University of Warwick, Coventry, CV4 7AL, UK (phone: +44 24 765 28028; e-mail: v.melnicuk@warwick.ac.uk).

S. Birrell is an Assistant Professor at WMG, University of Warwick, Coventry, CV4 7AL, UK (e-mail: s.birrell@warwick.ac.uk).

E. Crundall is a Human Factors Specialist at Jaguar Land Rover, Banbury Road, Gaydon, CV35 0RR, UK (e-mail: ecrundal@jaguarlandrover.com).

P. Jennings is a Professor at WMG, University of Warwick, Coventry, CV4 7AL, UK (e-mail: paul.jennings@warwick.ac.uk).

wearables can be worn for extensive periods of time outside of driving activities, whereas in-vehicle-embedded sensors can only evaluate drivers' state during actual driving. Authors of this current paper have advocated for such approach in their previous publication, titled "Towards Hybrid Driver State Monitoring: Review, Future Perspectives and the Role of Consumer Electronics" [13]. The hybrid approach to driver state monitoring argues that Consumer Electronic Devices (CEDs) should be used to source driver state related data not as a substitute to in-vehicle-embedded systems, but as a complementing source thus, potentially contribute to enchased validity and reliability of driver state related measurements. Moreover, CEDs could offer more flexible updatability of DSM due to shorter product life cycles compared to in-vehicle embedded sensors. Also, some commercially available devices offer patented solutions to physiological evaluation. For instance, Empatica E4 wristband is equipped with multiple sensors capable of capturing Blood Volume Pulse (BVP), Galvanic Skin Response (GSR), and Peripheral Skin Temperature (PST) readings from human's wrist [14]. Such device can be worn during real world driving activities and provide real-time assessment of drivers' state in non-intrusive manner. In addition to wearable devices, a smartphone can be used to track drivers' emotions. A mobile device can be equipped with commercially available software solutions, such as Affectiva Affdex SDK, which derives emotions from humans' facial expressions in real-time by means of image processing [15].

In summary, consumer grade electronic devices could offer a great potential for state monitoring in driving context in non-intrusive and cost-effective manner. However, research community and automakers are reluctant to adopt DSM though CEDs due to lack of evidence on data validity and reliability. To promote CEDs uptake in DSM, this study presents an attempt to employ several consumer grade devices for the purpose of DSM.

III. Methodology

The study adopted repeated measure design and was performed in WMG's 3xD Simulator for Intelligent Vehicles at University of Warwick (see Figure 1). The driving simulator removed much of the risk associated with the testing of new technology in driving context including, risk of crash due to distraction. It also allowed to replicate complex driving scenario, fixed environmental conditions, and several unexpected events, all in a safe and repeatable setting [16]. The simulator consists of a full-size vehicle and 360 degrees' cylindrical screen with high definition projection. The vehicle was set to emulate road motion through its embedded hydraulics and preprogramed dynamic model of Range Rover Evoque. The simulator was also set to emulate automatic gear box. The simulated driving activities consisted of various road types i.e., training (mixture of various road layouts), urban (high traffic load, 30 mph, 7-minute duration), rural (moderate traffic load, variable speed limit of 40-60 mph, 5-minute duration), motorway (absence of traffic, straight road, constant 70 mph, 5-minute duration), fully autonomous (moderate traffic, variable speed limit of 40-70 mph, 2-minute duration), and racing (moderate traffic, unrestricted speed limit). Every road section was separated by a one-minute transition period. During this period participants were asked to evaluate their subjective workload level in a scale from zero (very low overall workload) to 100 (extremely high level of overall workload) in accordance to Overall Workload Scale (OWS) [17]. This method allowed to avoid continuous driving interruption, which could potentially affect physiological responses. Therefore, total duration of continuous driving lasted for approximately 33 minutes. Road sections were designed to replicate differences in driving task complexity and mimic real-world driving conditions. The scenario also included several events e.g., rapidly decelerating in-front vehicles, unexpectedly crossing pedestrians, forced overtaking of stationary vehicles, avoidance of stationary vehicles in autonomous mode, and extremely dense and slow moving traffic in urban area. These events were designed to encourage unusual physiological and emotional responses under highly distressful situations. Throughout the experiment participants were exposed to auditory navigation commands which were embedded into the scenario at fixed locations. Before driving, each participant was asked to rest for five minutes inside the simulator in seated position while avoiding any extensive body movements. Physiological measurements captured while resting were later used as a base line reference.

Figure 1: WMG, University of Warwick, 3xD Simulator for Intelligent Vehicles.

A. Apparatus

Throughout the experiment participants wore a wristband (Empatica E4) and a chest heart monitor (Polar H7) to capture physiological responses to various driving scenario modes and events. In addition to wearable devices, a smartphone (Samsung S7 Edge), equipped with a front-facing camera, was used to capture drivers' facial expressions for emotional state estimation. The summary of all signals that were captured is provided in the table below (see Table 1).

Table 1: Summary of signals used.

Signal	Empatica E4	Polar H7	Affectiva Affdex
Blood Volume Pulse (BVP)	64 Hz	-	-
Interbeat Intervals (RR)	4 Hz	1 Hz	-
Galvanic Skin Response (GSR)	4 Hz	-	-
Peripheral Skin Temperature (PST)	4 Hz	-	-
Facial Expressions	-	-	6 Hz
Emotions	-	-	6 Hz

To facilitate wireless connectivity of multiple devices through Bluetooth, an Android application was developed (see Figure 2). The application allowed to collect, synchronise, and store physiology and emotions measures from multiple data sources. All the signals were synchronised to the highest frequency measurement that is, BVP, which was captured at 64Hz. All measurements were later extracted into a single comma-separated file for post-analysis.

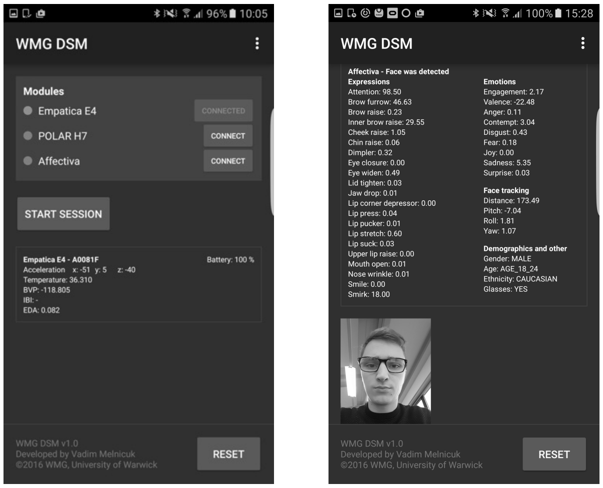

Figure 2: Mobile Driver State Monitoring application.

B. Data preparation

The data exported from WMG DSM application was processed in MATLAB R2016b. Physiology and emotions data was synchronised with precise timings of all road section changes for later categorisation; additional road section labels were generated to account for data delay required for derived signals with pre-fixed moving window. Heart rate metrics from Polar H7 were processed by custom built software to derive Heart Rate Variability (HRV) in both time and frequency domains. Algorithms used to derive HRV were produced in accordance to standards described by Malik et al. [12]. Time domain derivative was represented by widely used Root Mean Square of the Successive Differences (RMSSD), which was continuously calculated every second over 10, 30, 60, and 120-second-long moving windows. Frequency domain metrics consisted of standard parameters including, Very Low (VLF, 0.0033-0.04 Hz), Low (LF, 0.04-0.15 Hz), High (HF, 0.15-0.4) frequencies, and a ratio between low and high frequencies (LF/HF). The HRV spectrum was derived using Lomb-Scargle periodogram as it allows direct use of unevenly sampled interbeat intervals and is robust to missed heart beats [18]. Once all participants' data was processed, it was imported into SPSS Version 27 for further analysis.

A. Recruitment of participants

The recruitment of participants was conducted internally at WMG, University of Warwick; some Jaguar Land Rover employees have also taken part in the trials. Participants were required to meet the following criteria: be over 21 years of age, hold full category "B" driving license, have normal or correct-to-normal vision, and do not have any cardiovascular diseases. In total, 39 participants have volunteered, of which 28 were males and 11 were females. The sample consisted of mixed age, M = 32.23, SD = 9.45. All relevant ethical considerations were described in a protocol guidance document, which was approved by Warwick's Biomedical & Scientific Research Ethics Committee (BSREC).

IV. Results

In total, 25 out of 39 participants successfully completed driving experiment. Ten participants have experienced simulator sickness and four participants could not complete the experiment due to various technical reasons. Additionally, heart rate data of eight participants was rejected after visual inspection due to abundance of artefacts.

A. Subjective evaluation

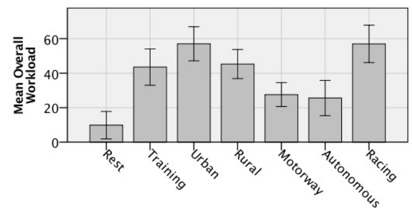

Figure 3: Mean Overall Workload categorised by scenario modes.

The impact of task complexity in different scenario modes on subjective evaluation of overall workload (see Figure 3) was studied using univariate analysis of variance (ANOVA) and post-hoc pairwise comparison with Bonferroni correction (see Table 2). Results suggest that scenario modes had a statistically significant effect on subjective evaluations of OWS, $F(5, 94) = 19.900$, $p<0.001$.

B. Objective evaluation

Several objective measures were studied using ANOVA to identify statistically significant effect of task complexity in various scenario modes. Moreover, each metric was further analysed using post-hoc pairwise comparison with Bonferroni correction (see Table 2). Besides that, unrestricted driving areas including, Training and Transition, were excluded from this analysis as these modes were deemed undistinguishable in terms of experimental design.

It was found that GSR (see Figure 4a) was significantly affected by changes in scenario modes, $F(5, 28844) = 57.630$, $p<0.001$. Overall, PST measures (see Figure 4b) were also found to be significantly affected by various scenario modes, $F(5, 21658) = 195.844$, $p<0.001$. The impact of scenario modes on interbeat intervals (RR) measured by Polar H7 (see Figure 4e) and Empatica E4 (see Figure 4g) was studied next. It was found that RRs from both measurement methods were significantly affected by various scenario modes (Polar H7, $F(5, 38672) = 129.843$, $p<0.001$; Empatica E4, $F(5, 16225) = 51.144$, $p<0.001$). Even though Empatica E4 did not detect approximately 62% of RR data points in comparison to Polar H7, strong correlation was present between measurement methods, Pearson's correlation $r = 0.876$, $p<0.01$.

The response to several emergency situations was also studied across GSR, temperature, and RR measures (see Figure 4b, d, f). Each measure contained 15 seconds' window prior to an event and 15 seconds' post-event response. In total, five emergency situations in Urban area were subject to evaluation. Neither GSR, $F(1, 1839) = 1.721$, $p = 0.190$, nor temperature, $F(1, 1382) = 0.381$, $p = 0.537$, measures were found to be significantly different Before and After emergency situations. However, RRs were found to be affected by stimuli $F(1, 2578) = 49.120$, $p<0.001$.

Time domain HRV that is, RMSSD (see Figure 5) was also studied using ANOVA. RMSSD across all moving windows was found to be significantly affected due to impact on various scenario modes, $F(5, 18674) = 137.069$, $p<0.001$ (results reported for 120s moving window). Furthermore, gradual decrease of variance was observed in RMSSD across all moving windows that is, $(SD_{RMSSD10} = 17.083) > (SD_{RMSSD30} = 15.286) > (SD_{RMSSD60} = 14.603) > (SD_{RMSSD120} = 14.026)$.

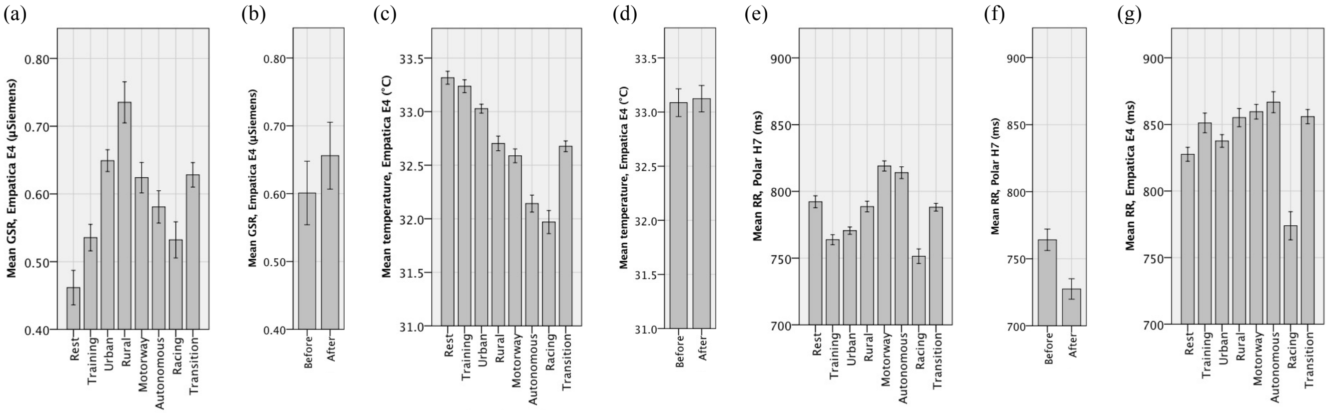

Figure 4: Plots (a), (c), (e), and (g) represent physiological responses categorised by scenario modes. Plots (b), (d), and (f) represent physiological responses before and after emergency situations in urban scenario mode.

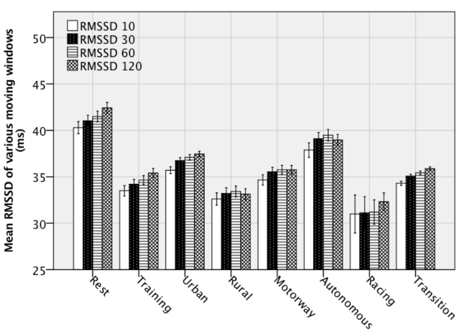

Figure 5: Mean RMSSD of 10, 30, 60, and 120 seconds moving window categorised by scenario modes.

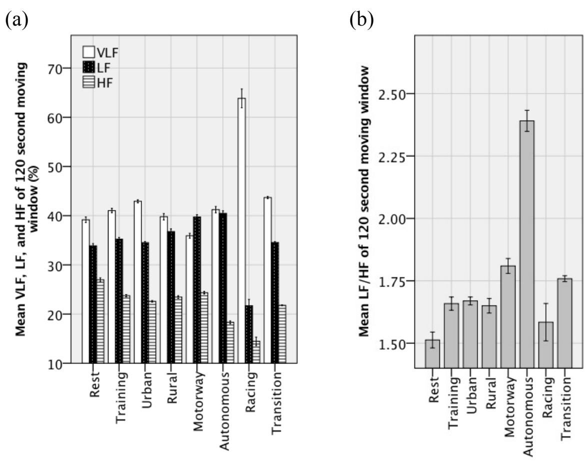

Figure 6: Plot (a) represents mean VLF, LF, and HF, plot (b) represents mean LF to HF ratio of 120 second moving windows also categorised by scenario modes.

Frequency domain HRV metrics (see Figure 6a, b) were also evaluated using ANOVA. Significant differences between scenario modes were found in all frequency spectrum ranges that is, VLF, $F(5, 18674) = 201.692$, $p<0.001$, LF, $F(5, 18674) = 204.335$, $p<0.001$, and HF, $F(5, 18674) = 319.836$, $p<0.001$. LF to HF ratio was also studied using ANOVA and was found to be significantly affected by scenario mode changes, $F(5, 18674) = 292.011$, $p<0.001$.

Table 2: Post-hoc pairwise comparisons with Bonferroni correction.

Measure	Rest	Urban	Rural	Motorway	Autonomous	Racing
OWS	2,3,6	1,4,5	1,5	2,6	2,3,6	1,4,5
GSR	2,3,4,5	1,3,5,6	1,2,4,5,6	1,3,6	1,2,3	1,2,3,4
PST	2,3,4,5,6	1,3,4,5,6	1,2,5,6	1,2,5,6	1,2,3,4	1,2,3,4
RR, Polar	2,4,5,6	1,3,4,5,6	2,4,5,6	1,2,3,6	1,2,3,6	1,2,3,4,5
RR, Empatica	3,4,5,6	3,4,5,6	1,2,6	1,2,6	1,2,6	1,2,3,4,5
RMSSD10	2,3,4,5,6	1,3,5,6	1,2,4,5	1,3,5	1,2,3,4,6	1,2,5
RMSSD120	2,3,4,5,6	1,3,4,5,6	1,2,4,5	1,2,3,5,6	1,2,3,4,6	1,2,4,5
VLF	2,4,5,6	1,3,4,5,6	2,4,5,6	1,2,3,5,6	1,2,3,4,6	1,2,3,4,5
LF	3,4,5,6	3,4,5,6	1,2,4,5,6	1,2,3,6	1,2,3,6	1,2,3,4,5
HF	2,3,4,5,6	1,3,4,5,6	1,2,4,5,6	1,2,3,5,6	1,2,3,4,6	1,2,3,4,5
LF/HF	2,3,4,5	1,4,5	1,4,5	1,2,3,5,6	1,2,3,4,6	4,5
Anger	3,4,5	3,4,5	1,2,4,5	1,3,6	1,2,3,6	2,4,5
Contempt	2,4,5,6	1,3,4,5,6	2,5,6	2,5,6	1,2,3,4,6	1,2,3,4,5
Disgust	2,4,5,6	1,3,4,6	2,4,5,6	1,2,3,5,6	1,3,4,6	1,2,3,4,5
Fear	2,5	1,3,4,5,6	2,5	2,5	1,2,3,4,6	2,5
Joy	2,5,6	1,3,4,5,6	2,5,6	2,5,6	1,2,3,4,6	1,2,3,4,5
Sadness	2,4,5	1,4,5,6	4,5,6	1,2,3,5,6	1,2,3,4,6	2,3,4,5
Surprise	2,3,4,5,6	1,3,4,6	1,2,5,6	1,2,5,6	1,3,4,6	1,2,3,4,5
Eye closure	2,3,4,5,6	1,3,4,5	1,2,4,5	1,2,3,5	1,2,3,4,6	1,5
Eye widen	2,3,4,5,6	1,3,4,5,6	1,2,4,5,6	1,2,3,5,6	1,2,3,4	1,2,3,4
Mouth open	2,3,4,5,6	1,3,4,5,6	1,2,4,5,6	1,2,3,5,6	1,2,3,4,6	1,2,3,4,5
Attention	2,3,4,6	1,3,4,5,6	1,2,4,5,6	1,2,3,5,6	2,3,4,6	1,2,3,4,5

1 – significantly different from Rest ($p<0.05$)
2 – significantly different from Urban ($p<0.05$)
3 – significantly different from Rural ($p<0.05$)
4 – significantly different from Motorway ($p<0.05$)
5 – significantly different from Autonomous ($p<0.05$)
6 – significantly different from Racing ($p<0.05$)

Variance of seven basic emotions (see Figure 7a) was studied using ANOVA. It was found that all emotions including anger, $F(5, 212524) = 26.863$, $p<0.001$, contempt, $F(5, 212524) = 290.024$, $p<0.001$, disgust, $F(5, 212524) = 300.968$, $p<0.001$, fear, $F(5, 212524) = 78.382$, $p<0.001$, joy, $F(5, 212524) = 1294.228$, $p<0.001$, sadness, $F(5, 212524) = 99.428$, $p<0.001$, and surprise, $F(5, 212524) = 469.598$, $p<0.001$, are significantly different across various scenario modes. Similarly, all emotions demonstrated significant difference in response to emergency situations (see Figure 7b). Furthermore, some individual facial expressions (see Figure 7c) were studied using ANOVA. Results showed that eye closure patterns, $F(5, 212524) = 1096.692$, $p<0.001$, eye widen patterns, $F(5, 212524) = 653.318$, $p<0.001$, and mouth open patterns, $F(5, 212524) = 819.127$, $p<0.001$, were significantly affected by difference in task complexity in various scenario modes.

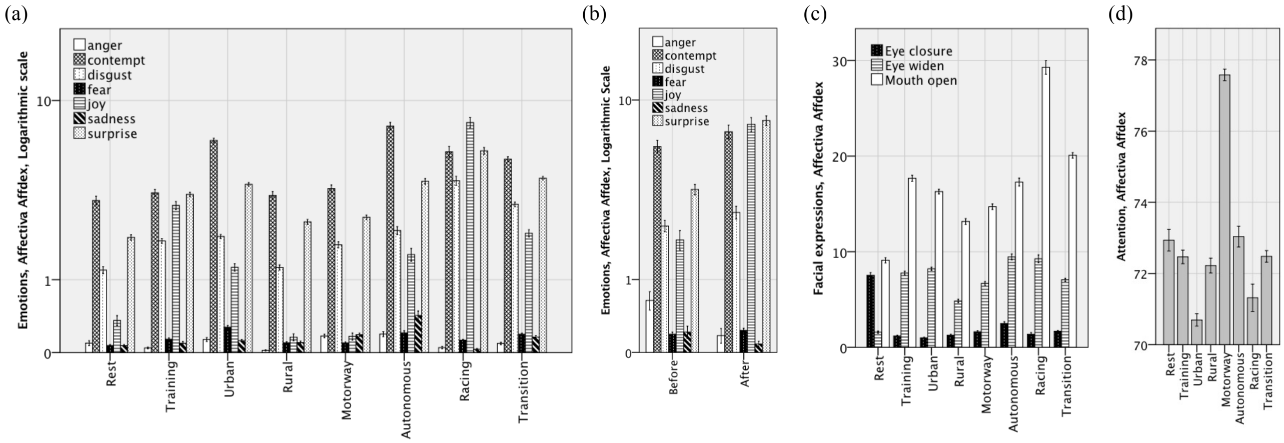

Figure 7: Plot (a) represents mean emotions categorised by scenario modes, plot (b) represents mean emotion responses before and after emergency situations in urban scenario mode, plot (c) represents facial expressions, namely eye closure, eye widen, and mouth open, categorised by scenario modes, plot (d) represents mean attention categorised by scenario modes.

The variance of Attention index (see Figure 7d), which accounts for glances that are in direct contact with smartphone's camera, was also studied using ANOVA; it was found to be significantly affected by various scenario modes, $F(5, 212524) = 512.067$, $p<0.001$.

Finally, correlations between participants' subjective workload ratings (see Figure 3) and physiological measures as well as some facial expressions were identified. A significant correlation across the whole data sample was found in GSR (Spearman's rho = 0.280, $p<0.01$) and eye closures (Spearman's rho = -0.314, $p<0.01$). However, neither RR intervals nor HRV derivatives were found to be significantly correlated to OWS ratings. Despite this, some individual participants' physiological indicators, including RR, RMSSD, and HF, have shown significant correlation with OWS.

V. Discussion and Conclusions

The aim of this study was to contribute to future development of hybrid Driver State Monitoring (DSM) systems as described in [13]. Such systems might rely on various sources of driver state related data for enhanced reliability and validity of state estimation, including data captured by consumer grade electronic devices. Therefore, this paper presented analysis of physiological indicators and emotions measured by means of several CEDs in simulated driving environment. In order to facilitate wireless collection, synchronisation, and storage of various measures, the mobile application was developed. The application has proven to successfully process substantial amounts of incoming data from multiple devices at frequencies up to 64Hz. Additionally, emotions were derived on-board in real-time by means of image processing at 6Hz using Affectiva's Affdex SDK [15]. This demonstrates how some of the DSM functionality can be performed by a mobile device, which could share processed driver state indicators with a vehicle.

As part of this user trial numerous physiological indicators were captured including GSR, PST, and RRs. Analysis of these indicators revealed some patterns. For instance, throughout the whole duration of driving task PST was monotonically decreasing (see Figure 4c); similar pattern was observed by Yamakoshi et al. in [19]. In contrast, GSR did not counteract pattern observed in PST (see Figure 4a), but instead was found to strongly correlate to participants' subjective overall workload ratings. However, GSR has predominantly been used as an indicator of distress, and not as a measure of workload [20].

The interbeat intervals (RRs) (see Figure 4e) have revealed following patterns: highly demanding scenario modes, including Urban and Racing sections, resulted in elevated heart rate, whereas low-demand tasks, such as Motorway and Autonomous driving, caused heart rate to slow down. This heart rate behaviour was previously observed in [21]. Despite the fact that heart metrics were found to be the most sensitive to mental workload [22], RR intervals in our sample did not correlate to overall workload ratings. Moreover, during the Resting period, which was the least demanding task, heart rate was insignificantly different to Rural driving and more elevated in comparison to Motorway driving. This might have been caused by distress due to anticipation during Resting period. Therefore, future simulated driving studies should allow longer resting period.

Furthermore, time and frequency domain analysis of Heart Rate Variability (HRV) was performed. Overall decrease of HRV due to increase of workload was observed, this phenomena was previously described in [12] and was previously replicated in our preliminary DSM user trials [21]. However, some inconsistences in HRV results were present. For instance, RMSSD in Urban section was significantly higher than HRV in Rural and Motorway sections. Despite this, low workload tasks, such as Rest and Autonomous driving, and high workload task, such as Racing, could be successfully distinguished. Moreover, the HRV was derived using multiple moving windows, which could be used for short and medium term workload estimation. However, it should be noted that suggested moving window for short term HRV estimation is 120 seconds [12]. Therefore, HRV is not a preferred measure for instantaneous workload estimation. In the other hand, RR intervals (see Figure 4f) were proven to be a good indicator of short term change in drivers' state due to some distressful stimuli i.e., emergency situations on roads.

Frequency domain metrics (i.e., VLF, LF, and HF) were derived using 2-minute moving window to address all frequency spectrum components as suggested by [12]. The HF band has revealed some pattern that was previously described in [22]. It should be noted that HF reflects

Parasympathetic Nervous System (PNS) activity and is mostly influenced by breathing or Respiratory Sinus Arrhythmia [23]. Our results suggest that PNS was mostly active during Resting and least active during task of high complexity that is, Racing and Urban driving. Moreover, some participants' HF data showed negative correlation with subjective workload evaluations. Neither LF, nor the ratio between LF and HF, which meant to reveal Sympathetic Nervous System (SNS) activity, did not follow expected patterns i.e., no significant correlation to OWS was found. The VLF range was not interpreted because the physiological explanation of the VLF component is much less defined in literature [12].

Seven basic emotions were detected in accordance to FACS [11]. Emotions are important in human intelligence, rational decision making, social interaction, perception, memory, learning, creativity, and more. Furthermore, emotions were found to be a motivating and a guiding force in perception and attention [24]. Thus, an attempt was made to capture emotional state in driving context in order to enhance DSM. Previous attempts have successfully captured mostly negative valence in various driving conditions [25]. Considering that the high relevance of positive valence to driving was documented by a substantial body of literature [7], future DSM systems could benefit from detection of positive emotions. Therefore, specific scenario mode was designed to elicit positive emotion onto participating drivers. Racing mode resulted in significantly elevated emotional response including joy, and surprise (see Figure 7a). The emotional response was similar in evaluation of extreme situations (see Figure 7b), where prevalence of joy and surprise could be observed. In addition to emotions some individual facial expressions were analysed, namely eye closure, eye widen, and mouth open (see Figure 7c). Eye closure was more prevalent during the least demanding tasks that is, Rest, Autonomous, and Motorway driving, which can be useful in detecting if a driver is becoming tired as eye closures were found to negatively correlate with subjective workload ratings. Finally, attention index provided by Affectiva (see Figure 7d) could to be used to indicate drivers' glance direction. It was found that participants were less focused on the road in-front of them during highly demanding scenario modes such as Urban and Racing. On the other hand, the highest attention measure was achieved during Motorway driving, where participants were required to drive on a straight road without any traffic.

To conclude, this study has provided some supporting evidence for adoption of consumer grade electronic device in driver state monitoring. It was shown how a mobile device can act as a host for DSM system, support connectivity, synchronisation, and storage of DSM related measures from multiple devices. Also, the appropriate suggestions were made on relevance of various physiology and emotions indicators, sourced from CEDs. Moreover, it was shown how task complexity can be reflected in drivers' physiological and emotional responses in simulated environment.

ACKNOWLEDGMENT

The support for this work was provided by Engineering & Physical Sciences Research Council (EPSRC) and Jaguar Land Rover (JLR).

REFERENCES

[1] Department for Transport, "Reported Road Casualties in Great Britain : Main Results 2014," 2014.

[2] NHTSA, "Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey," Washington, US, 2015.

[3] S. Lal and A. Craig, "A critical review of the psychophysiology of driver fatigue," *Biol. Psychol.*, vol. 55, pp. 173–194, 2007.

[4] B. Reimer and B. Mehler, "The impact of cognitive workload on physiological arousal in young adult drivers: a field study and simulation validation," *Ergonomics*, vol. 54, pp. 932–942, 2011.

[5] T. a. Ranney, W. R. Garrott, and M. J. Goodman, "NHTSA driver distraction research: Past, present, and future," *Driv. Distraction Internet Forum*, no. 233, p. 9, 2000.

[6] E. Roidl, F. W. Siebert, M. Oehl, and R. Höger, "Introducing a multivariate model for predicting driving performance: The role of driving anger and personal characteristics," *J. Safety Res.*, vol. 47, pp. 47–56, 2013.

[7] F. Eyben, M. Wöllmer, T. Poitschke, B. Schuller, C. Blaschke, B. Färber, and N. Nguyen-Thien, "Emotion on the Road—Necessity, Acceptance, and Feasibility of Affective Computing in the Car," *Adv. Human-Computer Interact.*, vol. 2010, pp. 1–17, 2010.

[8] Frost & Sullivan, "Executive Outlook of Health , Wellness , and Wellbeing Technologies in the Global Automotive Industry Volume-driving OEMs Leading the First Wave of Proliferation," 2015.

[9] N. Stanton, A. Hedge, K. Brookhuis, E. Salas, and H. Hendrick, *Handbook of Human Factors and Ergonomics Methods*. CRC Press, 2004.

[10] A. Kramer and T. Weber, "Application of psychophysiology to human factors," in *Handbook of Psychophysiology*, 2000, pp. 794–814.

[11] P. Ekman and E. Rosenberg, *What the face reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*, 2nd ed. Oxford University Press, 2005.

[12] M. Malik, A. Camm, B. J. J, G. Breithardt, S. Cerutti, R. . Cohen, P. Coumel, E. Fallen, H. Kennedy, R. Kleiger, F. Lombardi, A. Malliani, A. Moss, J. Rottman, G. Schmidt, P. Schwartz, and D. Singer, "Heart rate variability. Standards of measurement, physiological interpretation, and clinical use," *Eur. Heart J.*, pp. 354–381, 1996.

[13] V. Melnicuk, S. Birrell, E. Crundall, and P. Jennings, "Towards Hybrid Driver State Monitoring: Review, Future Perspectives and the Role of Consumer Electronics," in *IEEE Intelligent Vehicle Symposium*, 2016, pp. 19–22.

[14] Empatica Inc., "Empatica E4 User Manual." pp. 1–32, 2015.

[15] Affectiva Inc., "Emotion Recognition Software and Analysis - Affectiva." [Online]. Available: http://www.affectiva.com/.

[16] WMG, "Intelligent Vehicles Research," 2017. [Online]. Available: http://www2.warwick.ac.uk/fac/sci/wmg/research/automotive/smarter/

[17] S. G. Hill, H. P. Iavecchia, J. C. Byers, a. C. Bittner, a. L. Zaklad, and R. E. Christ, "Comparison of four subjective workload rating scales," *Hum. Factors*, vol. 34, no. 4, pp. 429–439, 1992.

[18] J. a. Healey and R. W. Picard, "Detecting Stress During Real-World Driving Tasks Using Physiological Sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, Jun. 2005.

[19] T. Yamakoshi, K. Yamakoshi, S. Tanaka, M. Nogawa, M. Shibata, Y. Sawada, P. Rolfe, and Y. Hirose, "A Preliminary Study on Driver's Stress Index Using a New Method Based on Differential Skin Temperature Measurement," in *29th Annual International Conference of the IEEE EMBS*, 2007, pp. 722–725.

[20] D. De Waard, "The Measurement of Drivers ' Mental Workload," University of Groningen, 1996.

[21] V. Melnicuk, S. Birrell, P. Konstantopoulos, E. Crundall, and P. Jennings, "JLR heart: Employing wearable technology in non-intrusive driver state monitoring. Preliminary study," *IEEE Intell. Veh. Symp. Proc.*, vol. 2016-Augus, no. Iv, pp. 55–60, 2016.

[22] J. Paxion, E. Galy, and C. Berthelon, "Mental workload and driving," *Front. Psychol.*, vol. 5, no. December, pp. 1–11, 2014.

[23] L. K. McCorry, "Physiology of the autonomic nervous system," *Am. J. Pharm. Educ.*, vol. 71, no. 4, 2007.

[24] R. W. Picard, *Affective Computing*, 1st ed. Cambridge, United States: MIT Press, 1997.

[25] E. Schmidt, R. Decke, and R. Rasshofer, "Correlation between subjective driver state measures and psychophysiological and vehicular data in simulated driving," *IEEE Intell. Veh. Symp. Proc.*, vol. 2016-Augus, no. Iv, pp. 1380–1385, 2016.