Systematic Evaluation of Driver's Behavior: A Multimodal Biometric Study

Michela Minen[1,2], Luisina Gregoret[1](✉), Divya Seernani[1], and Jessica Wilson[1]

[1] iMotions A/S, Copenhagen, Denmark
Luisina.Gregoret@imotions.com
[2] University of Padua, Padua, Italy

Abstract. Complex traffic areas and high cognitive workload while driving are leading contributors to traffic crashes. Even though cognitive workload and stress have been previously assessed through various neurophysiological responses, they are rarely characterized simultaneously, limiting the triangulation of behavioral metrics (like drivers' visual attention and facial coding) with physiological measures to investigate their interplay. The aim of the present study was to systematically characterize stress and cognitive workload through a multimodal assessment comprising eye-tracking, facial expressions, galvanic skin response (GSR), electromyography (EMG), electrocardiography (ECG) and respiration in three controlled driving simulations of varying complexity: 1. Baseline driving on an open road (Baseline); 2. Navigating between traffic cones (Cones); and 3. Driving in a neighborhood with multiple stressors (Traffic). The selected metrics were eye tracking dwell time, the presence of facial brow furrow, GSR peaks/minute, EMG activity of the upper trapezius muscle, heart rate and heart rate variability (HRV) and respiration cycles/minute. Physiological responses showed significant increases in GSR, heart rate and trapezius EMG activity with Cones and Traffic compared to Baseline. Eye tracking metrics were shown to be indicative of driving behavior in different conditions. There were no significant differences in facial expressions, HRV or respiration. These results are somewhat consistent with previous literature, suggesting that a multimodal approach to physiological signals can characterize affective and cognitive states in driving scenarios.

Keywords: Driving simulator · Cognitive workload · Eye tracking · Physiology

1 Introduction

The interaction between complex driving conditions and high cognitive workload while driving is one of the leading contributing factors to traffic crashes. Recent studies have employed several physiological and behavioral tools to help assess driver behavior in controlled studies. These can include eye tracking, which can reveal information about a driver's visual processing of the road. Other promising biosensors include electromyography (EMG), electrocardiography (ECG), and galvanic skin response (GSR) which measure muscle contraction, heart rate and heart rate variability and eccrine gland activity, respectively [1–3]. EMG can be used to study motor strategies and ergonomics, and

C. Stephanidis et al. (Eds.): HCII 2023, CCIS 1836, pp. 57–64, 2023.
https://doi.org/10.1007/978-3-031-36004-6_9

detect driver fatigue. Heart rate and heart rate variability (HRV), derived from the ECG signal, is one of the most frequently used techniques for assessing driver workload and has been shown to decrease significantly during periods of increased mental workload and stress while driving [4–6]. GSR, by comparison, is a pure reflection of sympathetic nervous system activity [7] and is thus increased in periods of high physiological arousal [8]. Lastly, the use of camera-based facial expression analysis has been to predict unsafe driving behaviors [9].

The combination of a comprehensive biometric assessment with realistic driving simulators can help pinpoint specific instances where drivers may be distracted, overwhelmed cognitively, or dealing with navigation-related issues that could compromise their safety while operating a motor vehicle. The overall aim of the present study, therefore, was to systematically characterize stress and cognitive workload through a multimodal assessment comprising eye-tracking, facial expressions, GSR, EMG, ECG and respiration in three randomized and controlled driving simulations to holistically measure driver performance.

2 Methods

Ten healthy respondents (31.4 ± 7.1 years old; 5 female) were recruited and participated in this evaluation. The inclusion criteria included individuals within an age range of 18–40 years who have a drivers license and no previous experience driving a car simulator, whereas the exclusion criteria included visual and motor disabilities, chronic or current acute pain at the time of the experiment, and cardiorespiratory disorders.

2.1 Experimental Procedure

Respondents were instructed to drive a professional driving simulator (VI-GRADE, Italy) in a 3-screen desktop set-up with dedicated steering wheel and pedals (Logitech G29, Logitech, Switzerland). The simulation software used was “MCity Traffic” (World-Sim, VI-GRADE, Italy). All signals were recorded and analyzed by a biometric software (iMotions 9.3.3, iMotions A/S, Denmark).

The study protocol consisted of:

1. Low-workload Baseline scenario (Baseline): Drive 3 min on an open highway, without traffic.
2. Medium-workload Cones scenario (Cones): Slalom the car between traffic cones and perform a double lane change maneuver as fast as reasonably possible, followed by 1.5 min of straight driving recovery.
3. High-workload city traffic scenario (Traffic): Drive 5 min through a city with random traffic conditions, traffic lights, other vehicles, and pedestrians.

The Baseline scenario was fixed at the start of the study protocol, while Traffic and Cones scenarios were pseudorandomized.

2.2 Data Collection

Eye-Tracking. Eye-tracking was performed with a multicam system of 3 infrared cameras (n = 10, Smart Eye Pro, Smart Eye, Sweden), positioning one infrared camera per simulator screen. Fixations were classified using a velocity-based I-VT filter (iMotions 9.3.3, iMotions A/S, Denmark) set at 30°/s. Dwell time (as a percentage of total task time) and number of revisits were selected as primary variables to evaluate the visual attention of the drivers.

Facial Expression Analysis. Drivers were recorded using a Logitech C920 webcam (n = 10, 640 × 480 resolution, 30 fps) positioned on top of the middle screen of the 3-screen setup. Affectiva's AFFDEX 2.0, embedded in the iMotions software (iMotions 9.3.3, iMotions A/S, Denmark) was used to detect and quantify facial movements. Brow Furrow was selected as a metric because of its affiliation with negative valence as well as workload.

Respiration. Respiration rate was recorded using a stretch belt connected to a respiration amplifier (n = 5, RSP 100 and MP160, Biopac systems Inc, US). Respiration rate (cycles/min) was determined based on the respiration count and the duration of the signal.

EMG. EMG activity was recorded from the right upper trapezius muscle using bipolar surface EMG electrodes connected to either a stationary EMG device (n = 5, Biopac Systems EMG100C and MP160, USA) or a bluetooth wireless EMG device (n = 2, Shimmer EXG, Shimmer Sensing, Ireland). The reference electrode was placed on the right acromion. EMG signals were sampled at 512 Hz and bandpass filtered from 20–500 Hz with a 4th order Butterworth filter, rectified, then smoothed using a zero-phase-shift 4th-order Butterworth filter with a 10-Hz cut-off frequency. The area under the curve (AUC) was calculated from the smoothed EMG signal.

GSR. GSR were registered by a portable device (n = 10, Shimmer3 GSR+, Shimmer Sensing, Ireland), positioning the two electrodes on the volar side of the ring and middle fingers of the non-dominant hand. A threshold of 0.01 uS and 0 uS was established to detect the onset and offset of every GSR peak, respectively. After the labeling of GSR peaks, a peaks/minute metric was computed.

ECG. ECG activity was obtained through a wireless strap (n = 8, Polar Belt H10, Polar, Switzerland) applied on the chest. Heart rate and HRV was calculated from R-R intervals; HRV was expressed using the standard deviation of the N-N intervals (SDNN).

2.3 Statistics

All signal processing was conducted using the iMotions R-Library (iMotions 9.3.3, iMotions A/S, Denmark). Statistical analyses were performed in Matlab using the statistics and machine learning toolbox v12.3 (Mathworks, Massachusetts, US). Threshold for statistical significance was set as $p < 0.05$. All data was assessed for normality by the Shapiro-Wilk test. Friedman tests were applied for all analyses across the three driving conditions, with p-values adjusted by Bonferroni correction for multiple comparisons. Results are reported as median (interquartile ranges) in both text and plots.

3 Results

Driver physiology and behavior across multiple biosensor modalities (eye tracking, GSR, FEA, ECG, EMG and respiration) were compared across three driving conditions of varying complexity (Baseline, Cones and Traffic). A summary of the descriptive statistics is presented in Table 1.

Table 1. Descriptive statistics of multimodal metrics during three driving conditions (Baseline, Cones and Traffic). Values are represented as median (interquartile range). † denotes a significant main effect, $p < 0.05$. †† denotes a significant main effect, $p < 0.01$. *a* denotes significant differences compared to Baseline, $p < 0.05$. *b* denotes significant differences compared to Baseline, $p < 0.01$. *c* denotes significant differences compared to Cones, $p < 0.05$.

Sensor	Metric	Condition		
		Baseline	Cones	Traffic
Eye Tracking	Dwell Time (speedometer)††	26.54 (15.29–35.56)	**a**22.41 (12.78–34.87)	**b**6.65 (4.00–8.61)
Eye Tracking	Dwell Time (left screen)†	0.21 (0.12–0.43)	0.12 (0.07–0.14)	**c**1.02 (0.57–2.40)
Eye Tracking	Dwell Time (center screen)	65.39 (55.4–73.5)	72.38 (69.07–75.61)	70.86 (68.65–75.75)
Eye Tracking	Dwell Time (right screen)†	0.25 (0.22–0.32)	0.15 (0.06–0.16)	**c**1.04 (0.57–1.99)
Facial Coding	% Time Spent Brow Furrow	0.90 (0.15–1.72)	0.0 (0.0–1.89)	1.12 (.12–5.18)
GSR	Peaks/Minute††	3.06 (1.77–4.25)	6.41 (4.64–7.79)	a6.05 (4.88–8.30)
EMG	AUC Upper Trapezius (mV*s)††	2.93 (2.35–3.94)	3.66 (2.34–4.61)	**b, c**5.92 (4.58–6.60)
ECG	Heart Rate (beats/minute)†	74.95 (66.67–81.67)	78.81 (71.07–81.99)	**c**74.88 (66.99–80.83)
ECG	Heart Rate Variability (SDNN)	42.33 (41.17–65.60)	55.29 (44.97–81.76)	47.28 (45.3–66.50)
Respiration	Cycles/Minute	13.76 (12.79–14.59)	12.77 (9.76–13.35)	12.23 (11.89–12.34)

There was a significant effect of dwell time on the speedometer across tasks ($F(2) = 14.888$, $p < 0.001$, $N = 10$). Post hoc analyses show higher dwell time % on the speedometer in the Baseline scenario as compared to both the Cones ($p < 0.05$) and Traffic ($p < 0.001$) scenarios. While dwell time on the center screen did not change across conditions, there was a significantly higher dwell time on the left ($F(2) = 8.857$, $p = < 0.02$, post hoc $p = 0.01$, $N = 10$) and right ($F(2) = 6.400$, $p < 0.05$, $N = 10$; post hoc $p < 0.05$) screens in the Traffic scenario compared to the Cones scenario.

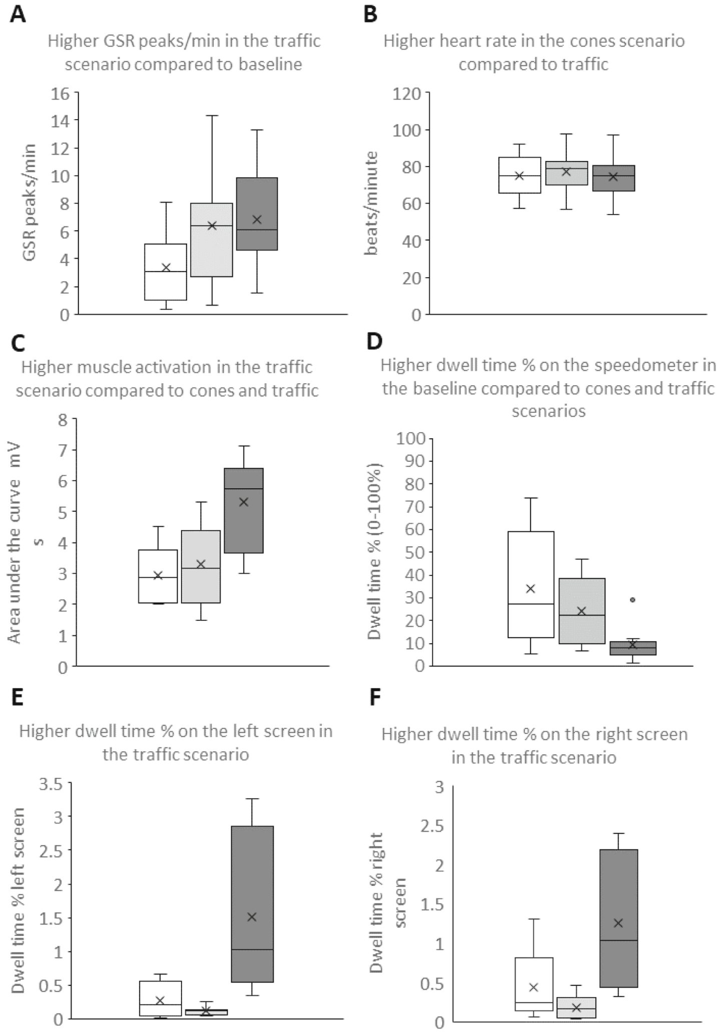

Fig. 1. Box plots of the biometric responses in the Baseline (white), Cones (light gray) and Traffic (dark gray) scenarios showing the mean (cross), the median (horizontal line), quartile ranges and minimum and maximum values. **A)** Higher GSR peaks/min were observed in the Traffic scenario compared to Baseline. **B)** Higher heart rate was observed in the Cones scenario compared to Traffic. **C)** Higher muscle activation was observed in the Traffic scenario compared to Cones and Baseline. **D)** Higher dwell time % over the speedometer occurred in the Baseline compared to Cones and Traffic scenarios. **E** and **F)** Higher dwell time % over the left and right screens were observed in the Traffic scenario, as compared to Cones and Baseline.

There was also a significant effect of task on GSR peaks/minute ($F(2) = 9.750$; $p = 0.008$, $N = 10$). Post hoc analysis showed that drivers exhibited higher GSR peaks/min while driving in the Traffic scenario as compared to the Baseline scenario (adjusted $p <$

0.01, Fig. 1A). A tendency of higher GSR peaks/min is observed while navigating Cones as compared to the Baseline scenario, but this was not statistically significant (adjusted $p = 0.073$).

Friedman analysis indicated a significant effect on heart rate ($F(2) = 6.867$, $p < 0.05$, $N = 8$). Post hoc analysis revealed that HR was higher while navigating the Cones than driving in the Traffic scenario (Fig. 1B - $F(2) = 1.250$, $p < 0.05$), but not compared to the Baseline. HRV, however, did not show any significant differences across the different scenarios ($F(2) = 0.250$, $p = 0.882$, $N = 8$).

Friedman tests on the AUC of smoothed EMG activity of the upper trapezius muscle resulted in a significant effect ($F(2) = 13.000$; $p = 0.002$, $N = 8$) showing higher activation while driving in the Traffic scenario compared to the Baseline ($p = 0.001$) as well in the Traffic scenario compared to the Cones ($p < 0.05$, Fig. 1C).

4 Discussion

The present evaluation expands the evidence of modulated physiological responses while navigating driving scenarios of varying complexity (Baseline, Cones and Traffic scenarios) by also including behavioral measures like eye tracking and facial expression analysis with which we can more clearly establish context.

4.1 Differences in Eye Tracking Behavior Confirm Different Attentional Strategies Between Tasks

Dwell time has been used before to represent attention allocation, especially in seeing how long attention is allocated to important or competing stimuli like cell phones or signposts [10]. In the Baseline task, there was an increase in dwell time on the speedometer, indicative of an increased attention on the speedometer compared to the Cones or Traffic scenarios. This could be because the Baseline task was the simplest, and the nature of the Cones and Traffic scenarios required attentional resources to be devoted elsewhere. This is supported by the increased dwell time on the left and right screens in the Traffic condition compared to Cones. The Traffic scenario is the most complex of the three, as it contains multiple competing elements like pedestrians, cars and intersections that require different attentional strategies [11]. In short, we see differences in eye tracking behavior across the scenarios due to the varying attentional requirements of the driver.

4.2 Modulation of GSR, EMG and HR in Complex Driving Scenarios

The overall hypothesis of the electrophysiological evaluation was that GSR peaks/min, heart rate, muscle activity and respiration rate would be facilitated and heart rate variability would be diminished under the most demanding scenarios (Cones and Traffic). We did find that GSR peaks/min and muscle activity were significantly facilitated in the Traffic scenario as compared to Baseline, possibly in response to increased task loads. This is only partially in agreement with Healey and Picard (2005), who did not see EMG changes, but found that GSR and ECG together correlated the highest with drivers' perceived stress in real-life drives [12].

Respiration and heart rate variability did not show significant differences across tasks. A lack of significant change in HRV was surprising given previous literature but HRV can also be influenced by a multitude of environmental, neurophysiological and lifestyle factors [13]. Studies involving HRV as a variable might benefit most from a multimodal approach.

We did not see a significant change in brow furrow across the tasks, although other head and facial movements have been shown to correlate with driving behaviors [9]. As the technology continues to advance in this arena, it will be interesting to see what combination of head and face movements can be most predictive of a driver's internal state.

5 Conclusion

The application of multiple biosensors in driver research has the potential to advance our comprehension of driving behavior, guide the creation of fresh driver-training initiatives and safety interventions, and eventually aid in lowering the rate of traffic accidents. The results of the present study indicate higher muscle tension (EMG activity), modulated arousal (heart rate and GSR peaks/min) and higher perceptual load (dwell time %) in complex driving scenarios. Most, but not all our findings are consistent with what we have seen in prior literature. The interaction between the driver's cognitive and physiological states and the environment inside and outside the vehicle produce a complex system that can be difficult to model and predict. It is only through repeated studies, with the triangulation of multiple physiological and behavioral measures, that we can gain deeper insights that can be leveraged to protect the driver.

References

1. Yang, X., et al.: The effects of traveling in different transport modes on galvanic skin response (GSR) as a measure of stress: an observational study. Environ. Int. **156**, 106764 (2021). https://doi.org/10.1016/j.envint.2021.106764
2. Mehler B., Reimer, B., Wang, Y.: A comparison of the heart rate and heart rate variability indices in distinguishing single-task driving and driving under secondary cognitive workload. In: Proceedings of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, pp. 590–597. Lake Tahoe, CA (2011)
3. Brookhuis, K.A., de Waard, D.: Monitoring drivers' mental workload in driving simulators using physiological measures. Accid. Anal. Prev. **42**, 898–903 (2010). https://doi.org/10.1016/j.aap.2009.06.001
4. Lenneman, J.K., Backs, R.W.: Cardiac autonomic control during simulated driving with a concurrent verbal working memory task. Hum. Factors **51**, 404–418 (2009). https://doi.org/10.1177/0018720809337716
5. Lenneman, J.K., Backs, R.W.: Enhancing assessment of in-vehicle technology attention demands with cardiac measures. In: Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular applications, pp. 20–21. ACM, New York (2010)
6. Collet, C., Clarion, A., Morel, M., Chapon, A., Petit, C.: Physiological and behavioral changes associated to the management of secondary tasks while driving. App. Ergon. **40**(6), 1041–1046 (2009). https://doi.org/10.1016/j.apergo.2009.01.007

7. Boucsein, W.: Electrodemal Activity. 2nd edn. Springer, New York (2012). https://doi.org/10.1007/978-1-4614-1126-0
8. Mehler, B., Reimer, B., Coughlin, J.F., Dusek, J.A.: Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. Transp. Res. Rec. **2138**(1), 6–12 (2009). https://doi.org/10.3141/2138-02
9. Jabon, M., Bailenson, J., Pontikakis, E., Takayama, L., Nass, C.: Facial expression analysis for predicting unsafe driving behavior. IEEE Pervasive Comput. **10**(4), 84–95 (2011). https://doi.org/10.1109/MPRV.2010.46
10. Strayer, D.L., Drews, F.A., Johnston, W.A.: Cell phone-induced failures of visual attention during simulated driving. J. Exp. Psychol. Appl. **9**(1), 23–32 (2003). https://doi.org/10.1037/1076-898x.9.1.23
11. Liang, N., et al.: Using eye-tracking to investigate the effects of pre-takeover visual engagement on situation awareness during automated driving. Accid. Anal. Prev. **157**, 106143 (2021). https://doi.org/10.1016/j.aap.2021.106143
12. Healey, J.A., Picard, R.W.: Detecting stress during real-world driving tasks using physiological sensors. IEEE Trans. Intell. Transp. Syst. **6**(2), 156–166 (2005). https://doi.org/10.1109/TITS.2005.848368
13. Fatisson, J., Oswald, V., Lalonde, F.: Influence diagram of physiological and environmental factors affecting heart rate variability: an extended literature overview. Heart Int. **11**(1), 32–40 (2016). https://doi.org/10.5301/heartint.5000232