Assessing the Use of Physiological Signals and Facial Behaviour to Gauge Drivers' Emotions as a UX Metric in Automotive User Studies

Christine Spencer
cspencer03@qub.ac.uk
Queen's University Belfast
Belfast, Antrim

İbrahim Alper Koç
iakoc@sensum.co
Sensum
Belfast, Antrim

Chihiro Suga
chihiro.suga@volvocars.com
Volvo Cars USA
Sunnyvale, California

Alexander Lee
alex.lee@volvocars.com
Volvo Cars USA
Sunnyvale, California

Anupama Mahesh Dhareshwar
anupama.dhareshwar@volvocars.com
Volvo Cars
Gothenburg, Västra Götaland

Elin Franzén
elin.franzen@volvocars.com
Volvo Cars
Gothenburg, Västra Götaland

Maria Iozzo
maria.iozzo@volvocars.com
Volvo Cars USA
Sunnyvale, California

Gawain Morrison
gawain@sensum.co
Sensum
Belfast, Antrim

Gary McKeown
g.mckeown@qub.ac.uk
Queen's University Belfast
Belfast, Antrim

ABSTRACT

Studies have shown that drivers' emotions can be assessed via changes in their physiology and facial behaviour. This study examined this approach as a means of gauging user experience (UX) in an automotive user study. 36 drivers' responses to typical UX-style questions were compared with computational estimates of their emotional state, based on changes in their cardiac, respiratory, electrodermal and facial signals. The drivers' arousal and valence levels were monitored in real-time as they drove a 23-mile route around Sunnyvale, CA. These estimates corresponded with two independent observers' judgments of the drivers' emotions. The results highlighted a disparity between the self-report and algorithmic scores—the drivers who answered the UX questions more positively experienced higher levels of stress—evidenced by higher arousal and lower valence algorithmic scores. The findings highlight the value of supplementing self-report measures with objective estimates of drivers' emotions in automotive UX research.

CCS CONCEPTS

• **Human-centered computing → User studies**; • **Applied computing → Psychology**.

KEYWORDS

Automotive user experience metrics, drivers' emotions, physiological signals, facial behaviour, multimodal data synchronisation

AutomotiveUI '20 Adjunct, September 21–22, 2020, Virtual Event, USA

ACM ISBN 978-1-4503-8066-9/20/09.
https://doi.org/10.1145/3409251.3411728

ACM Reference Format:
Christine Spencer, İbrahim Alper Koç, Chihiro Suga, Alexander Lee, Anupama Mahesh Dhareshwar, Elin Franzén, Maria Iozzo, Gawain Morrison, and Gary McKeown. 2020. Assessing the Use of Physiological Signals and Facial Behaviour to Gauge Drivers' Emotions as a UX Metric in Automotive User Studies. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '20 Adjunct), September 21–22, 2020, Virtual Event, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3409251.3411728

1 INTRODUCTION

How drivers respond emotionally at the wheel is of critical interest to researchers in the automotive user experience (UX) field [2]. Stress is a state of emotional or mental tension that people experience in response to taxing or adverse environmental demands or events which threaten their well-being [8]. Driver stress is known to impair driving performance [3][5][25][17][6], and can be triggered not only by external events in the road, but also by the design of a vehicle's assisted driving, navigation or HCI systems [29]. Although self-report measures are commonly used to gain insights into drivers' affective states [3][17][24][6], the objectivity of this approach is not clear. Self-report measures can suffer from social desirability bias [15]—people may feel obliged to report favourable feelings or opinions about their experience to avoid causing offence. They also do not allow for continuous measurement, and so may fail to capture important changes in a person's emotional state as they interact with a driving system. Research indicates that it is possible to estimate whether a person is feeling stressed by measuring changes in their physiology [11]—with stress being characterised by heightened physiological arousal—elevated heart rate [13][7], breathing rate [12][26] and skin conductance [19][14]. Facial cues can also offer information about a drivers' emotional state [9][10], although this is context-dependent [1].
According to [18], it is possible to conceptualise any emotion in terms of two dimensions—valence and arousal. Valence refers to

the subjective "positive" or "negative"quality of an emotional experience while arousal refers to a person's level of physiological activation. This on-road study assessed the value of gauging drivers' emotions through objective measures as an automotive UX metric by comparing how well drivers' responses to typical user-experience style questions correlated with computational estimates of their arousal and valence levels, derived from their physiological and facial data.

2 METHOD

2.1 Participants

36 employees within a leading international automotive brand were recruited for an internal study to test user reactions to an assisted driving system. Participants were all at least 18 years old, possessed a full driving licence, drove a minimum of three to five days a week on average and had previously used the brand's assisted "Pilot Assist" driving functionality. The recruited sample was 75% male, with a mean age of 38.73 years.

2.2 Procedure

Each participants' session consisted of a 23-mile route around Sunnyvale, CA, which involved industrial and pedestrianised downtown areas, as well as a highway section where participants used an assisted driving system. Each session took between 40 and 45 minutes on average. Data-recording difficulties resulted in incomplete data for 3 participants' sessions—these were excluded from the analyses.

2.2.1 ***Physiological and face data.*** Participants' heart and breathing rates were recorded by an upper body Equivital EQ02+ LifeMonitor sensor belt, while their skin conductance level was measured via electrodes placed on the underside of two fingers on their left hand, as shown in Figure 1. Participants' facial expressions were captured by a dash-mounted ELP Webcam USB 1080P camera and were analysed by the SHORE [21] facial classification library. Drivers' speech was recorded with a RØDE lapel microphone, and their view of the road was recorded with an outward-facing, dash-mounted Logitech RGB webcam camera. The placement of the data capture devices is illustrated in Figure 2.
All of the devices were connected to a Sensum Synsis™Empathic AI Developer Kit [20]—which predicted the drivers' arousal and valence levels in real-time. These algorithms were developed using a decision-tree machine learning approach, using data which was obtained in previous studies (e.g. [23]) where people were required to complete stressful and calming activities in a simulated driving context, and where the participants' behaviour was then coded for arousal, valence and stress on a visual analogue scale by multiple annotators. Decision-tree models were then trained to recognise changes in physiological activation and "positivity" or "negativity" based on these annotated labels. The validity of these estimates was corroborated by the judgments of independent observers who reviewed sample session footage after the study—see [22] for greater details of this validation.

2.2.2 ***Self-Reported Emotion.*** At regular intervals during the drive, the participants also verbally rated how they were feeling on a scale of 1 ("*not at all positive*") to 10 ("*highly positive*")—these scores were shown to have little correlation with their observed behaviour and so were not considered to offer an objective measure of emotion—an interested reader should refer to [22] for further details.

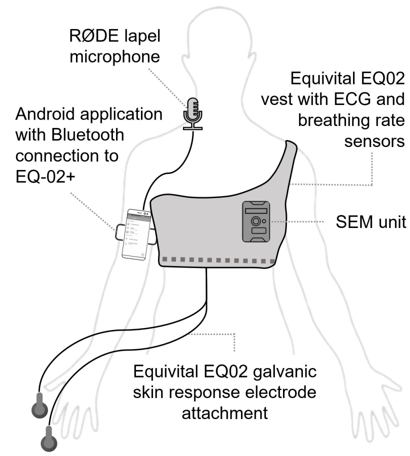

Figure 1: Placement of physiological sensors

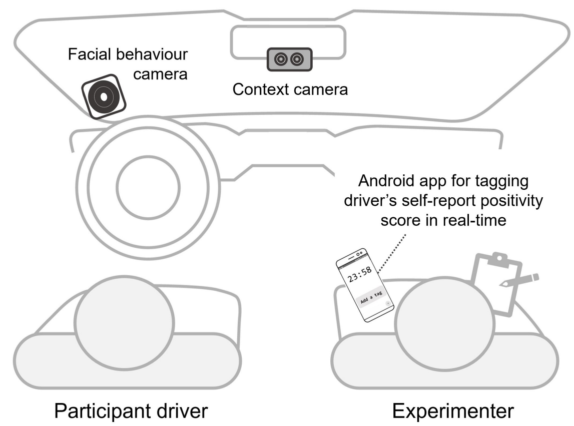

Figure 2: Placement of cameras

2.2.3 ***Self-Report UX Evaluation Scores.*** After the drive, participants completed a comprehensive 60-item questionnaire. 17 items were user experience style questions—participants were asked to rate the degree to which the driving system worked accurately, was reliable and easy to understand, as well as the degree to which they felt comfortable using it. They also indicated whether they understood how to interact with the user interface, and whether the interface provided them with appropriate and useful feedback. They also rated the degree to which the system seemed intelligent, perceptive, responsive and intuitive, as well as the degree to which it helped them to relax. All questions were answered on a 6-point Likert scale ranging from "Strongly disagree" to "Strongly agree". Participants' responses were totaled to derive a "Self-report UX Evaluation" score. Participants who scored above the median were presumed to have enjoyed a more positive user experience with the system. These individuals formed the "Highly Positive Self-report UX Evaluation" group, while those scoring less than the median formed the "Less Positive" group.

3 RESULTS

Analysis and data visualisation was performed using R's [16] mgcv [28], ggplot2 [27] and tidymv [4] packages. The scatter diagrams in Figure 3 show that both groups' arousal and valence predictions were most densely populated in the lower-right quadrant of the valence-arousal space, which is theoretically characterised by feelings of calmness [18]. However, the "Positive" group had a greater number of predictions in the upper-left quadrant, which is theoretically characterised by negative emotions such as stress or frustration [18].

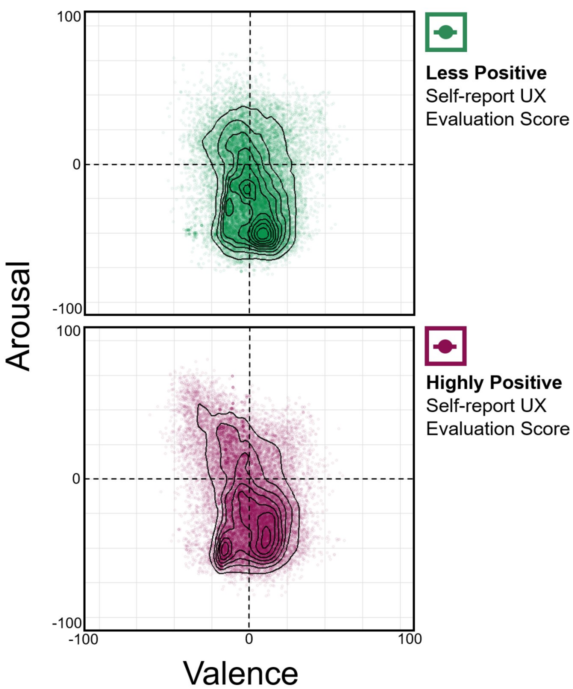

Figure 3: A Comparison of Each Group's Arousal and Valence Scores
Note. The contour lines demarcate zones in the VA space where the valence-arousal predictions are populated to the same degree. The innermost contour line represents the emotional zone most experienced by participants in each group.

Generalised additive models (GAMs) with restricted maximum likelihood estimation were also used to examine how each groups' arousal and valence levels varied over time. Each model took the generalised form:

$$y = \alpha + f(time, \beta) + \epsilon \sim N(0, \sigma^2)$$

where y represents either arousal or valence, α represents the intercept, $f(time, \beta)$ represents the smooth factor interaction, and ϵ represents normally distributed errors. The pointwise 95% confidence intervals are shown as grey bands around each smooth in Figure 4. Where the confidence intervals of the two group's GAM smooths do not overlap with each other, there is a significant difference between the groups.

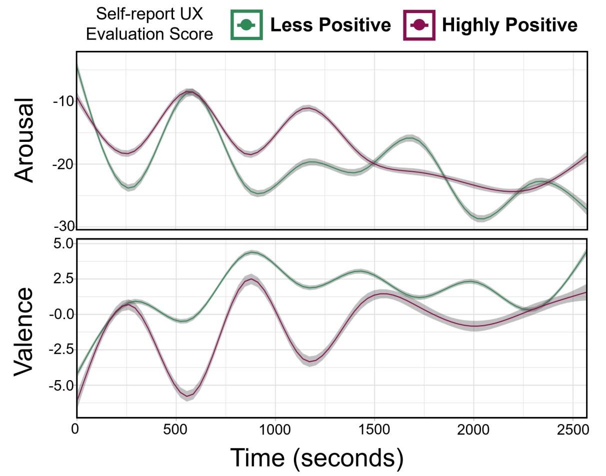

Figure 4: Changes in Each Group's Arousal and Valence Scores Over Time
Note. The grey variability bands around each smooth represent pointwise 95% confidence intervals.

As shown in Figure 4, both groups' arousal levels were similar at around 600-800 seconds into the session, at which time participants were driving on the highway and using the assisted driving system. However, there were interesting differences between the groups for the rest of the session. In the case of arousal, the smooth factor interaction term was significant (edf = 17.86, F = 88.75, p<.001)—the participants who answered more positively tended to have significantly higher arousal levels across the session compared to the "less positive" group. In the case of valence, the "highly positive" participants had significantly lower valence levels across the session compared to the "less positive" participants who had almost entirely positive valence values across the session. The smooth factor interaction term in this case was significant (edf = 18.73, F = 55.41, p<.001).

4 DISCUSSION

This study evaluated the practice of using algorithmic estimates of drivers' emotions—derived from their physiology and facial expressions—to gauge user experience in automotive studies. The findings highlighted a disparity between drivers' responses to typical UX-style questions and the algorithmic estimates. It would be expected that the drivers who provided more favourable responses about their experience would have enjoyed lower levels of stress during their drives. However, the algorithmic emotion predictions revealed the opposite to be the case. The participants who reported highly positive opinions consistently experienced higher levels of arousal and lower levels of valence across the session, entering the "stress" quadrant of [18]'s valence-arousal space more often than those who provided less positive responses.

It should be noted that the "Self-Report UX Evaluation Score" was rather broad, covering a wide range of user-experience elements, such as the system's ease of use and understanding, and its level of comfort, intelligence, intuitiveness and comfort, as examples. It was presumed that if participants rated these elements favourably, this meant that they enjoyed a positive emotional experience when using the system. Perhaps a future study could delineate these elements more clearly.

It should be noted that the drivers' self-report responses were skewed in this study—although the totalled "Self-Report UX Evaluation Score" responses had a potential range of 17-102, the median value was 93. Participants tended to provide favourable responses to most of the questions, which suggests that they may have experienced social desirability concerns. This supports the practice of supplementing self-report UX measures with objective measures of drivers' emotions. Even though the "less positive" group did not provide highly negative responses, notable differences were still observed between the groups' arousal and valence levels.

In conclusion, the present findings highlight the value of recording drivers' physiological and facial signals to derive a more objective metric of their experience in automotive user studies. Future work will apply this approach in expanded user study scenarios. We believe that examining drivers' emotions at a fine-grained, objective level can provide valuable insights into the individual differences that may exist in the psychological perception of UX elements—such as human-machine interface design, or vehicle dynamics with on-road events—and can contribute to the automotive UX research community.

5 ACKNOWLEDGMENTS

Thank you to all of the participants who gave their time to this study.

REFERENCES

[1] Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix M Martinez, and Seth D Pollak. 2019. Emotional expressions reconsidered: challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest* 20, 1 (2019), 1–68.

[2] Michael Braun, Florian Weber, and Florian Alt. 2020. Affective Automotive User Interfaces – Reviewing the State of Emotion Regulation in the Car. *arXiv preprint arXiv:2003.13731* (03 2020).

[3] Hua Cai and Yingzi Lin. 2011. Modeling of operators' emotion and task performance in a virtual driving environment. *International Journal of Human-Computer Studies* 69, 9 (2011), 571–586.

[4] Stefano Coretta. 2020. *tidymv: Tidy Model Visualisation for Generalised Additive Models*. https://CRAN.R-project.org/package=tidymv R package version 3.0.0.

[5] Joseph F Coughlin, Bryan Reimer, and Bruce Mehler. 2011. Monitoring, managing, and motivating driver safety and well-being. *IEEE Pervasive Computing* 10, 3 (2011), 14–21.

[6] Na Du, Feng Zhou, Elizabeth M. Pulver, Dawn M. Tilbury, Lionel P. Robert, Anuj K. Pradhan, and X. Jessie Yang. 2020. Examining the effects of emotional valence and arousal on takeover performance in conditionally automated driving. *Transportation Research Part C: Emerging Technologies* 112 (2020), 78 – 87. https://doi.org/10.1016/j.trc.2020.01.006

[7] Stephen H Fairclough, Marjolein van der Zwaag, Elena Spiridon, and Joyce Westerink. 2014. Effects of mood induction via music on cardiovascular measures of negative emotion during simulated driving. *Physiology & behavior* 129 (2014), 173–180.

[8] Susan Folkman. 2013. *Stress: Appraisal and Coping*. Springer New York, New York, NY, 1913–1915. https://doi.org/10.1007/978-1-4419-1005-9_215

[9] Hua Gao, Anil Yüce, and Jean-Philippe Thiran. 2014. Detecting emotional stress from facial expressions for driving safety. In *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 5961–5965.

[10] G Giannakakis, Matthew Pediaditis, Dimitris Manousos, Eleni Kazantzaki, Franco Chiarugi, Panagiotis G Simos, Kostas Marias, and Manolis Tsiknakis. 2017. Stress and anxiety detection using facial cues from videos. *Biomedical Signal Processing and Control* 31 (2017), 89–101.

[11] Sylvia D Kreibig. 2010. Autonomic nervous system activity in emotion: A review. *Biological psychology* 84, 3 (2010), 394–421.

[12] Yuri Masaoka and Ikuo Homma. 1997. Anxiety and respiratory patterns: their relationship during mental stress and physical load. *International Journal of Psychophysiology* 27, 2 (1997), 153–159.

[13] Bruce Mehler, Bryan Reimer, and Joseph F. Coughlin. 2010. Physiological Reactivity to Graded Levels of Cognitive Workload across Three Age Groups: An On-Road Evaluation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 54, 24 (2010), 2062–2066. https://doi.org/10.1177/154193121005402409

[14] Yoko Nagai, Christopher Iain Jones, and Arjune Sen. 2019. Galvanic skin response (gsr)/electrodermal/skin conductance biofeedback on epilepsy: a systematic review and meta-analysis. *Frontiers in neurology* 10 (2019), 377.

[15] Delroy L Paulhus. 1984. Two-component models of socially desirable responding. *Journal of personality and social psychology* 46, 3 (1984), 598.

[16] R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[17] Ernst Roidl, Berit Frehse, and Rainer Höger. 2014. Emotional states of drivers and the impact on speed, acceleration and traffic violations—A simulator study. *Accident Analysis & Prevention* 70 (2014), 282–292.

[18] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.

[19] H. Sawai, G. Inou, and E. Koyama. 2015. Evaluating Optimal Arousal Level during the Task Based on Performance and Positive Mood: Extracting Indices Reflecting the Relationship among Arousal, Performance, and Mood. In *2015 3rd International Conference on Applied Computing and Information Technology/2nd International Conference on Computational Science and Intelligence*. 213–218.

[20] Sensum. 2020. Sensum Empathic AI Developer Kit. https://sensum.co/.

[21] SHORE. 2020. Emotion Recognition Software SHORE®: Fast, Reliable and Real-time Capable. http://www.iis.fraunhofer.de/shore.

[22] Christine Spencer, Ibrahim A Koc, Chihiro Suga, Alex Lee, Anupama M Dhareshwar, Elin Franzén, Maria Iozzo, Gawain Morrison, and Gary J McKeown. 2020. A Comparison of Unimodal and Multimodal Measurements of Driver Stress in Real-World Driving Conditions. https://doi.org/10.31234/osf.io/en5r3

[23] Christine Spencer, Daniel Moore, Gary McKeown, Lucy Rutherford, and Gawain Morrison. 2019. Context matters: protocol ordering effects on physiological arousal and experienced stress during a simulated driving task. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, Association for Computing Machinery, Cambridge, UK, 1–7.

[24] Fabius Steinberger, Ronald Schroeter, and Diana Babiac. 2017. Engaged Drivers - Safe Drivers: Gathering Real-Time Data from Mobile and Wearable Devices for Safe-Driving Apps. In *Automotive user interfaces*. Springer.

[25] Lana Trick, Seneca Brandigampola, and James Enns. 2012. How fleeting emotions affect hazard perception and steering while driving: The impact of image arousal and valence. *Accident; analysis and prevention* 45 (03 2012), 222–9. https://doi.org/10.1016/j.aap.2011.07.006

[26] Elke Vlemincx, Joachim Taelman, Steven De Peuter, Ilse Van Diest, and Omer Van Den Bergh. 2011. Sigh rate and respiratory variability during mental load and sustained attention. *Psychophysiology* 48, 1 (2011), 117–120.

[27] Hadley Wickham. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

[28] Simon N Wood. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73, 1 (2011), 3–36.

[29] Sebastian Zepf, Monique Dittrich, Javier Hernandez, and Alexander Schmitt. 2019. Towards Empathetic Car Interfaces: Emotional Triggers while Driving. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6.