



# PSYCHOPHYSIOLOGICAL CONTRIBUTIONS TO TRAFFIC SAFETY

EDITED BY: Guido P. H. Band, Karel Brookhuis, Bruce Mehler and  
Gianluca Borghini

PUBLISHED IN: Frontiers in Human Neuroscience





# frontiers

## Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88963-352-4

DOI 10.3389/978-2-88963-352-4

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)

# PSYCHOPHYSIOLOGICAL CONTRIBUTIONS TO TRAFFIC SAFETY

Topic Editors:

**Guido P. H. Band**, Leiden University, Netherlands

**Karel Brookhuis**, University of Groningen, Netherlands

**Bruce Mehler**, Massachusetts Institute of Technology, United States

**Gianluca Borghini**, Sapienza University of Rome, Italy

**Citation:** Band, G. P. H., Brookhuis, K., Mehler, B., Borghini, G., eds. (2020).

Psychophysiological Contributions to Traffic Safety. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-88963-352-4

# Table of Contents

|     |   |
|-----|---|
| 04  | <b><i>Editorial: Psychophysiological Contributions to Traffic Safety</i></b><br>Guido P. H. Band, Gianluca Borghini, Karel Brookhuis and Bruce Mehler   |
| 07  | <b><i>Age-Related Differences in Pro-active Driving Behavior Revealed by EEG Measures</i></b><br>Stephan Getzmann, Stefan Arnau, Melanie Karthaus, Julian Elias Reiser and Edmund Wascher   |
| 17  | <b><i>Recognizing Frustration of Drivers From Face Video Recordings and Brain Activation Measurements With Functional Near-Infrared Spectroscopy</i></b><br>Klas Ihme, Anirudh Unni, Meng Zhang, Jochem W. Rieger and Meike Jipp  |
| 32  | <b><i>Using Smartbands, Pupillometry and Body Motion to Detect Discomfort in Automated Driving</i></b><br>Matthias Beggiato, Franziska Hartwich and Josef Krems   |
| 44  | <b><i>A Novel Method for Classifying Driver Mental Workload Under Naturalistic Conditions With Information From Near-Infrared Spectroscopy</i></b><br>Anh Son Le, Hirofumi Aoki, Fumihiko Murase and Kenji Ishida   |
| 56  | <b><i>Brain Network Changes in Fatigued Drivers: A Longitudinal Study in a Real-World Environment Based on the Effective Connectivity Analysis and Actigraphy Data</i></b><br>André Fonseca, Scott Kerick, Jung-Tai King, Chin-Teng Lin and Tzyy-Ping Jung  |
| 69  | <b><i>EEG-Based Mental Workload Neurometric to Evaluate the Impact of Different Traffic and Road Conditions in Real Driving Settings</i></b><br>Gianluca Di Flumeri, Gianluca Borghini, Pietro Aricò, Nicolina Sciaraffa, Paola Lanzi, Simone Pozzi, Valeria Vignali, Claudio Lantieri, Arianna Bichicchi, Andrea Simone and Fabio Babiloni |
| 87  | <b><i>EEG-Based Neurocognitive Metrics may Predict Simulated and On-Road Driving Performance in Older Drivers</i></b><br>Greg Rupp, Chris Berka, Amir H. Meghdadi, Marija Stevanović Karić, Marc Casillas, Stephanie Smith, Theodore Rosenthal, Kevin McShea, Emily Sones and Thomas D. Marcotte  |
| 101 | <b><i>Demonstrating Brain-Level Interactions Between Visuospatial Attentional Demands and Working Memory Load While Driving Using Functional Near-Infrared Spectroscopy</i></b><br>Jakob Scheunemann, Anirudh Unni, Klas Ihme, Meike Jipp and Jochem W. Rieger  |
| 118 | <b><i>A Review of Psychophysiological Measures to Assess Cognitive States in Real-World Driving</i></b><br>Monika Lohani, Brennan R. Payne and David L. Strayer   |
| 145 | <b><i>Comparing the Relative Strengths of EEG and Low-Cost Physiological Devices in Modeling Attention Allocation in Semiautonomous Vehicles</i></b><br>Dean Cisler, Pamela M. Greenwood, Daniel M. Roberts, Ryan McKendrick and Carryl L. Baldwin  |
| 156 | <b><i>Associating Vehicles Automation With Drivers Functional State Assessment Systems: A Challenge for Road Safety in the Future</i></b><br>Christian Collet and Oren Musicant   |





# Editorial: Psychophysiological Contributions to Traffic Safety

Guido P. H. Band<sup>1\*</sup>, Gianluca Borghini<sup>2,3</sup>, Karel Brookhuis<sup>4</sup> and Bruce Mehler<sup>5</sup>

<sup>1</sup> Leiden Institute for Brain and Cognition, Leiden University Institute of Psychology, Leiden, Netherlands, <sup>2</sup> Department of Molecular Medicine, Sapienza University of Rome, Rome, Italy, <sup>3</sup> Neuroelectrical Imaging and BCI Lab, Fondazione Santa Lucia (IRCCS), Rome, Italy, <sup>4</sup> Faculty of Behavioural and Social Sciences, Groningen University, Groningen, Netherlands, <sup>5</sup> Massachusetts Institute of Technology, Center for Transportation and Logistics, Cambridge, MA, United States

**Keywords:** attention, traffic performance, human factors, automation, psychophysiology, fatigue, workload

## Editorial on the Research Topic

### Psychophysiological Contributions to Traffic Safety

Research shows the dominant contribution of human factors to incidents and accidents in air (Wiegmann and Shappell, 2001), road (Petridou and Moustaki, 2000), rail (Baysari et al., 2009), and maritime traffic participation (Hetherington et al., 2006), as well as in (air) traffic control (Isaac and Ruitenberg, 2016). Operator errors are in majority associated with a non-optimal mental state, such as fatigue, drowsiness, stress, elevated mental workload, distraction from the main task, limited vigilance, and failing situation awareness (Borghini et al., 2014). In turn, most of these functional limitations can be expressed as an aberration of arousal (Collet and Musicant; Lohani et al.) and difficulties maintaining relevant information in working memory (Wu et al., 2017). In an attempt to further reduce traffic casualties, there is an increasing interest in the potential of monitoring the mental state of both professionals and non-professional users. The current Research Topic deals with the question about how mental states can be optimally tracked in simulated as well as naturalistic contexts; now and when technology progresses further toward autonomous driving.

Assessment of fitness to drive by psychometric tools (e.g., self-report such as NASA-TLX; Hart and Staveland, 1988) has serious limitations. Construct validity, sensitivity, and reliability are limited because questionnaires rely on introspection and require a subjective judgment. More importantly, these techniques are not capable of capturing real-time changes, as they are typically not administered during action. Limited gain in traffic safety can be expected from identifying risk only after the fact.

In contrast, dynamic measures have great added value in monitoring the operator's tendencies in real-time during simulated or naturalistic traffic participation. Parameters like steering variability and route compliance (e.g., Getzmann et al.) are directly relevant for operation safety. Similarly, subtle bodily motions can provide clues about the operator's behavioral and muscular tendencies as related to safety. Beggiato et al. showed an increase in backward pressure on the driver's seat when autonomous navigation led to the proximity of a truck. Ihme et al. classified video recordings of facial expressions and were able to identify muscular indicators of frustration, a predictor of less responsible driving. Previous studies have shown the value of tracking head tilt and yawning as indices of drowsiness or fatigue (Reyes-Muñoz et al., 2016).

In contrast with yawning or tightening muscles, which can be perceived as byproducts of mental state, ocular behavior is a functional characteristic that may predict performance. Eye movements reflect overt attention and as such are an index of task-relevant behavior, as defined by areas with and without relevant information. Van de Merwe et al. (2012) demonstrated the value of eye movement parameters such as fixation time, focus and entropy to index situation awareness in simulated flying. Although eye trackers record more than only gaze, other parameters are

## OPEN ACCESS

### Edited and reviewed by:

Lutz Jäncke,  
University of Zurich, Switzerland

### \*Correspondence:

Guido P. H. Band  
band@fsw.leidenuniv.nl

### Specialty section:

This article was submitted to  
Cognitive Neuroscience,  
a section of the journal  
Frontiers in Human Neuroscience

**Received:** 31 October 2019

**Accepted:** 05 November 2019

**Published:** 19 November 2019

### Citation:

Band GPH, Borghini G, Brookhuis K  
and Mehler B (2019) Editorial:  
Psychophysiological Contributions to  
Traffic Safety.  
Front. Hum. Neurosci. 13:410.  
doi: 10.3389/fnhum.2019.00410

currently of limited value in naturalistic settings. In well-controlled laboratory or simulator settings, pupillometry has merit in detecting fatigue and workload (Wiegand et al., 2008). However, the pupil's strong response to variable lighting makes it virtually impossible to recognize subtle pupil dilation associated with arousal levels under naturalistic conditions. Eye blink duration, however, has successfully been linked to workload and fatigue (Benedetto et al., 2011).

In comparison with behavioral tendencies, psychophysiological and neuroimaging indicators tap even more directly into mental states during traffic participation (Van Erp et al., 2015; Borghini et al., 2017b). They have the potential to detect adverse changes before a change in the user functional state is visible in behavior. Near infrared spectroscopy (NIRS) is successful at recognizing frustration (Ihme et al.) and elevated workload (Le et al.; Scheunemann et al.) with accuracy of classification ranging from 78 to 90%. Electroencephalographic activity (EEG) has traditionally been used to distinguish spectral contributions associated with higher cognitive activity (Borghini et al., 2017a; Di Flumeri et al., 2018) vs. sleep-like activity (Simon et al., 2011; Fonseca et al., 2018). EEG is superior to blood-oxygenation based recordings in its temporal resolution, which allows for the identification of transient stimulus-induced changes using event-related potentials (Brookhuis and de Waard, 2010; Rupp et al., 2019) or time-frequency analyses (Gurudath and Riley, 2014).

Mental states are not only reflected in brain activity, but also in the activity and balance of the autonomic nervous system. In particular, arousal, vigilance (Schmidt et al., 2009), and fatigue (Wang et al., 2018) can be tracked with cardiovascular recording techniques (see Lohani et al. for a review). In addition, electrodermal activity can index elevated mental workload (Mehler et al., 2012) and stress (Boucsein, 2012). As more psychophysiological signals are monitored, the reliability of

estimating the user's mental state can only improve (Sahayadhas et al., 2012).

Neuroscientific methods have long suffered from practical limitations, such as non-portability, intolerance to motion, invasive or intruding sensors, and computational demands that prohibited real-time use. With the advance of technology, however, more and more of these neuroscientific approaches become accessible for real-time applications, and occasionally also for improving real-world traffic safety, as in driver assistance systems. Unfortunately, not all affordable sensors are suited to monitor performance potential. Cisler et al. showed that high-tech EEG recordings of midline alpha band could model speed of responding to faulty behavior of an autonomous car, but that low-tech indicators of eye gaze and heart-rate variability lacked predictive power.

The collection of papers in this Research Topic illustrates current topics in transportation research. Technology is moving forward. This introduces the challenge to maintain safety in the context of the increasingly popular, but yet imperfect operator assistance and automated driving systems. At the same time, new technology is rapidly providing new hardware, data processing algorithms and artificial intelligence that may make it more feasible and acceptable to track the operator's mental state and actively support, as needed, situation awareness. We know from the relative successes in aviation that it is possible to keep pilots aware and capable of taking over control despite extensive use of the autopilot mode. An important challenge is now to reach a similar level of capability in non-professionals, even in adverse conditions. Psychophysiological techniques can play a critical role in achieving this goal.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## REFERENCES

- Baysari, M. T., Caponecchia, C., McIntosh, A. S., and Wilson, J. R. (2009). Classification of errors contributing to rail incidents and accidents: a comparison of two human error identification techniques. *Saf. Sci.* 47, 948–957. doi: 10.1016/j.ssci.2008.09.012
- Benedetto, S., Pedrotti, M., Minin, L., Baccino, T., Re, A., and Montanari, R. (2011). Driver workload and eye blink duration. *Transport. Res. Part F* 14, 199–208. doi: 10.1016/j.trf.2010.12.001
- Borghini, G., Aricò, P., Di Flumeri, G., and Babiloni, F. (2017a). *Industrial Neuroscience in Aviation: Evaluation of Mental States in Aviation Personnel*, Vol. 18. Cham: Springer. doi: 10.1007/978-3-319-58598-7
- Borghini, G., Aricò, P., Di Flumeri, G., Cartocci, G., Colosimo, A., Bonelli, S., et al. (2017b). EEG-based cognitive control behaviour assessment: an ecological study with professional air traffic controllers. *Sci. Rep.* 7:547. doi: 10.1038/s41598-017-00633-7
- Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., and Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev.* 44, 58–75. doi: 10.1016/j.neubiorev.2012.10.003
- Boucsein, W. (2012). *Electrodermal Activity*. Springer Science & Business Media. Boston, MA: Springer. doi: 10.1007/978-1-4614-1126-0\_1
- Brookhuis, K. A., and de Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. *Accid. Anal. Prev.* 42, 898–903. doi: 10.1016/j.aap.2009.06.001
- Di Flumeri, G., Borghini, G., Aricò, P., Sciaraffa, N., Lanzi, P., Pozzi, S., et al. (2018). EEG-based mental workload neurometric to evaluate the impact of different traffic and road conditions in real driving settings. *Front. Hum. Neurosci.* 12:509. doi: 10.3389/fnhum.2018.00509
- Fonseca, A., Kerick, S., King, J.-T., Lin, C.-T., and Jung, T.-P. (2018). Brain network changes in fatigued drivers: a longitudinal study in a real-world environment based on the effective connectivity analysis and actigraphy data. *Front. Hum. Neurosci.* 12:418. doi: 10.3389/fnhum.2018.00418
- Gurudath, N., and Riley, H. B. (2014). Drowsy driving detection by EEG analysis using wavelet transform and k-means clustering. *Proc. Comput. Sci.* 34, 400–409. doi: 10.1016/j.procs.2014.07.045
- Hart, S. G., and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. *Adv. Psychol.* 52, 139–183. doi: 10.1016/S0166-4115(08)62386-9
- Hetherington, C., Flin, R., and Mearns, K. (2006). Safety in shipping: the human element. *J. Saf. Res.* 37, 401–411. doi: 10.1016/j.jsr.2006.04.007
- Isaac, A. R., and Ruitenberg, B. (2016). *Air Traffic Control: Human Performance Factors*. New York, NY: Routledge.
- 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. *Hum. Factors* 54, 396–412. doi: 10.1177/0018720812442086
- Petridou, E., and Moustaki, M. (2000). Human factors in the causation of road traffic crashes. *Eur. J. Epidemiol.* 16, 819–826. doi: 10.1023/A:1007649804201
- Reyes-Muñoz, A., Domingo, M. C., López-Trinidad, M. A., and Delgado, J. L. (2016). Integration of body sensor networks and vehicular *ad-hoc* networks for traffic safety. *Sensors* 16:E107. doi: 10.3390/s16010107
- Rupp, G., Berka, C., Meghdadi, A. H., Karic, M., Casillas, M., Smith, S., et al. (2019). EEG-based neurocognitive metrics may predict simulated and on-road driving performance in older drivers. *Front. Hum. Neurosci.* 12:532. doi: 10.3389/fnhum.2018.00532
- Sahayadhas, A., Sundaraj, K., and Murugappan, M. (2012). Detecting driver drowsiness based on sensors: a review. *Sensors* 12, 16937–16953. doi: 10.3390/s121216937
- Schmidt, E. A., Schrauf, M., Simon, M., Fritzsche, M., Buchner, A., and Kincses, W. E. (2009). Drivers' misjudgement of vigilance state during prolonged monotonous daytime driving. *Accid. Anal. Prev.* 41, 1087–1093. doi: 10.1016/j.aap.2009.06.007
- Simon, M., Schmidt, E. A., Kincses, W. E., Fritzsche, M., Bruns, A., Aufmuth, C., et al. (2011). EEG alpha spindle measures as indicators of driver fatigue under real traffic conditions. *Clin. Neurophysiol.* 122, 1168–1178. doi: 10.1016/j.clinph.2010.10.044
- Van de Merwe, K., Van Dijk, H., and Zon, R. (2012). Eye movements as an indicator of situation awareness in a flight simulator experiment. *Int. J. Aviat. Psychol.* 22, 78–95. doi: 10.1080/10508414.2012.635129
- Van Erp, J. B. F., Brouwer, A.-M., and Zander, T. O. (2015). Editorial: using neurophysiological signals that reflect cognitive or affective state. *Front. Neurosci.* 9:193. doi: 10.3389/fnins.2015.00193
- Wang, F., Wang, H., and Fu, R. (2018). Real-Time ECG-based detection of fatigue driving using sample entropy. *Entropy* 20:196. doi: 10.3390/e20030196
- Wiegand, D. M., Hanowski, R. J., Olson, R. L., and Melvin, W. (2008). *Fatigue Analyses: From 16 Months of Naturalistic Commercial Motor Vehicle Driving Data*. The National Surface Transportation Center for Excellence.
- Wiegmann, D. A., and Shappell, S. A. (2001). *A Human Error Analysis of Commercial Aviation Accidents in Using the Human Factors Analysis and Classification System (HFACS)*. Retrieved from: <https://commons.erau.edu/publication/1216>
- Wu, Y., Miwa, T., and Uchida, M. (2017). Using physiological signals to measure operator's mental workload in shipping – an engine room simulator study. *J. Mar. Eng. Technol.* 16, 61–69. doi: 10.1080/20464177.2016.1275496

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Band, Borghini, Brookhuis and Mehler. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Age-Related Differences in Pro-active Driving Behavior Revealed by EEG Measures

Stephan Getzmann\*, Stefan Arnau, Melanie Karthaus, Julian Elias Reiser and Edmund Wascher

Leibniz Research Centre for Working Environment and Human Factors, Technical University of Dortmund, Dortmund, Germany

## OPEN ACCESS

### Edited by:

Guido P. H. Band,  
Leiden University, Netherlands

### Reviewed by:

Vasil Kolev,  
Institute of Neurobiology (BAS),  
Bulgaria  
Mario Tombini,  
Università Campus Bio-Medico, Italy

### \*Correspondence:

Stephan Getzmann  
getzmann@ifado.de

**Received:** 08 May 2018

**Accepted:** 23 July 2018

**Published:** 07 August 2018

### Citation:

Getzmann S, Arnau S, Karthaus M, Reiser JE and Wascher E (2018) Age-Related Differences in Pro-active Driving Behavior Revealed by EEG Measures. *Front. Hum. Neurosci.* 12:321. doi: 10.3389/fnhum.2018.00321

Healthy aging is associated with a decline in cognitive functions. This may become an issue when complex tasks have to be performed like driving a car in a demanding traffic situation. On the other hand, older people are able to compensate for age-related deficits, e.g., by deploying extra mental effort and other compensatory strategies. The present study investigated the interplay of age, task workload, and mental effort using EEG measures and a proactive driving task, in which 16 younger and 16 older participants had to keep a virtual car on track on a curvy road. Total oscillatory power and relative power in Theta and Alpha bands were analyzed, as well as event-related potentials (ERPs) to task-irrelevant regular and irregular sound stimuli. Steering variability and Theta power increased with increasing task load (i.e., with sharper bends of the road), while Alpha power decreased. This pattern of workload and mental effort was found in both age groups. However, only in the older group a relationship between steering variability and Theta power occurred: better steering performance was associated with higher Theta power, reflecting higher mental effort. Higher Theta power while driving was also associated with a stronger increase in reported subjective fatigue in the older group. In the younger group, lower steering variability came along with lower ERP responses to deviant sound stimuli, reflecting reduced processing of task-irrelevant environmental stimuli. In sum, better performance in proactive driving (i.e., more alert steering behavior) was associated with increased mental effort in the older group, and higher attentional focus on the task in the younger group, indicating age-specific strategies in the way younger and older drivers manage demanding (driving) tasks.

**Keywords:** EEG, aging, proactive driving, mental effort, workload, alpha oscillations, theta oscillations, event-related potentials

## INTRODUCTION

Healthy aging is usually associated with a decline in sensory, cognitive and motor functions (Park, 2000; Lindenberger and Ghisletta, 2009). All these abilities are required when complex tasks have to be performed, like driving a car through dense city traffic or on a monotonous road where attention towards driving-related events has to be kept over a longer period of time. In fact, car driving can be regarded as a prototypical example of a complex task in which an adequate interplay of information intake, cognitive processing, and motor responses is necessary. Each of these instances may be prone to age-related deficits. For example, the sensory intake

may be reduced due to vision problems or responses to unexpected critical traffic events may be slowed down due to motor impairments (Anstey and Wood, 2011; Park et al., 2011). In addition, driving under challenging conditions may lead to a greater mental workload for older drivers (Cantin et al., 2009). Besides these negative consequences of aging, increasing driving experience and concurrent emergence of enhanced driving strategies are advantages that can help older drivers to manage complex traffic situations. Furthermore, it is known that older drivers are able to compensate for age-related decline, at least in part. On the one hand, compensation comprises strategies to adapt driving behavior to individual abilities, for example, by reducing driving speed and avoiding driving in the rain or for long periods of time (e.g., Molnar and Eby, 2008; Meng and Siren, 2012). On the other hand, older adults tend to increase their mental effort in order to counteract age-related neurocognitive decline (Cabeza et al., 2002). In the driving context, older drivers can allocate extra mental resources to keep performance high for adequate responses to a critical traffic situation. While adaptation of driving behavior is immediately visible, adjustments of mental effort (as well as a driver's mental state in general) are not. Also, increasing mental fatigue or attentional disengagement when driving on a monotonous road are not directly measurable, but they may have negative consequences for traffic safety. Therefore, monitoring a driver's mental state during driving can help us to learn more about age-related differences in traffic-related performance and hidden processes of compensatory activity (for review Da Silva, 2014).

A powerful approach to objectively determine mental states while driving are physiological measures (De Waard, 2002; Brookhuis and de Waard, 2010; Borghini et al., 2014). Especially neurophysiological measures like the brain oscillatory activity derived from the EEG have a long tradition. Here, the overall oscillatory power as well as the relative power in specific frequency bands like the Theta band (4–7 Hz) and Alpha band (8–12 Hz) are of special interest, as these are assumed to reflect different mental states. *Decreased* Alpha power is usually regarded as a marker of allocation of attention (Herrmann and Knight, 2001) and higher working memory demands (Klimesch, 2012). In contrast, *increased* Alpha power is related to mental fatigue, as well as to attentional withdrawal and disengagement (Hanslmayr et al., 2012; Wascher et al., 2014, 2016) as observed during monotonous and boring tasks (Borghini et al., 2014). Accordingly, increases in Alpha power during monotonous driving situations and low perceptual demands may reflect periods of inattention and mind-wandering (Lin et al., 2016). Activation in the Theta band, on the other hand, is assumed to reflect aspects of executive functioning and—more generally—cognitive control (Cavanagh and Frank, 2014; Cavanagh and Shackman, 2015). Accordingly, Theta power is usually increased with higher task demands (Jensen and Tesche, 2002; Onton et al., 2005) and—when time on task increases—with the effort to keep performance high (Wascher et al., 2014; Arnau et al., 2017).

In a recent driving simulator study, we employed oscillatory EEG measures to investigate underlying cognitive processes that

may explain inter-individual variability in driving performance (Karthaus et al., 2018). In a lane-keeping scenario, in which younger and older drivers had to respond to variable levels of crosswind by compensatory steering movements, both age groups showed comparable overall performance. However, the analysis of Alpha and Theta power suggested subtle differences in driving styles, on the one hand within the older group and, on the other hand, between the younger and older drivers. In accordance with previous results (Garcia et al., 2017), these driving styles could be described as either re-active or pro-active: while re-active driving was characterized by high driving lane variability and higher Alpha power, pro-active driving was indicated by low driving lane variability and lower activity in Alpha (and Beta) band. The latter has been associated with a more alert mental state, a better anticipation and active use of ongoing sensory driving information, and a more proactive planning of future responses. The re-active driving style, in contrast, led to situations in which the driver rather re-acts to environmental information, resulting in delayed compensatory steering activity (see also Braver, 2012).

Whether a driver uses a more re-active or pro-active driving style does not only depend on the driver him/herself, but also on the external conditions, i.e., the degree to which a situation in principle can be controlled. A highly controllable situation enables the anticipation and planning of future actions (e.g., steering movements when driving on a curvy road under good visibility conditions), whereas a poorly controllable situation forces the driver to respond exclusively to unpredictable outer stimulation (Garcia et al., 2017). Depending on the driving situation, the drivers' mental states may vary profoundly: a recent EEG study, in which pro-active and re-active driving scenarios were contrasted by employing either a curve-taking or crosswind-compensation task, it could be demonstrated that the latter task results much faster in a mental state of attentional disengagement and withdrawal of attentional resources (as indicated by an increasing Alpha). Taking bends, in contrast, was associated with a more focused driving activity, as indicated by a higher relative Theta power in general, and an additional increase in Theta power in narrow curves (Wascher et al., 2018). This higher Theta power can be interpreted as the consequences of the need for higher cognitive control in more demanding driving situations (see also Cavanagh and Frank, 2014). In line with this, higher steering demands in a pro-active driving scenario clearly resulted in increases in mental effort (Dijksterhuis et al., 2011).

Mental workload, task demands and driving performance are closely interrelated (Da Silva, 2014). For driving safety, a core question is therefore to what degree a driver actively uses the information provided by the environment and whether he or she adequately and flexibly adapts the processing of this information to a current driving situation. This critically depends on the interaction of: (a) the driving situation that might be more or less controllable; (b) the current workload of a given situation; and (c) the driver's individual mental capacities that might be reduced due to, e.g., temporal states of fatigue or boredom, or long-term age-related declines in cognitive functioning. The present study investigated this interplay in a pro-active lane-keeping driving task, in which younger and older drivers had to keep track



on a curvy road. The interaction of age, varying degrees of task workload (operationalized by bends of different radii), and variations in mental effort over time was tested by analysis of behavioral, neurophysiological and subjective measures. In particular, steering variability was analyzed as a measure of driving effort, with higher variability being associated with higher effort. By the analysis of oscillatory power in the Alpha and Theta bands, mental states of attentional withdrawal and effort were determined.

In addition to brain oscillatory power, event-related brain responses (ERPs) were analyzed that offer a further approach to objectively measure mental effort and task load of a primary task. Therefore, an auditory oddball paradigm has been applied in which ERPs were measured to a stream of task-irrelevant tone stimuli consisting of frequently presented standard and rare deviant stimuli. Two different fronto-central ERP components were analyzed, the mismatch negativity (MMN) reflecting the automatic context-dependent pre-attentive information processing of deviant tone stimuli (irrespective of the subject's focus of attention; for review Näätänen et al., 2007), and the P3a indicating an involuntary shifting of attention towards a deviant stimulus (e.g., Näätänen, 1992). Results of previous studies showed that MMN and P3a provide suitable approaches to mental states, like mental fatigue (Massar et al., 2010; Yang et al., 2013) or attentional load (Yucel et al., 2005; Zhang et al., 2006) as well as to task workload and time on task (Kramer et al., 1995; Wascher et al., 2016). In particular, in a steering-task paradigm, the P3a elicited by task-irrelevant auditory probes was reduced with higher steering difficulty, suggesting the P3a to be an indirect measure for evaluating mental workload (Brouwer et al., 2012; Scheer et al., 2016). Thus, in the present study, higher amplitudes of MMN and P3a would indicate a deeper processing of the task-irrelevant probes, either at a pre-attentive or attentional level, potentially impairing performance in the driving task.

## MATERIALS AND METHODS

### Participants

Sixteen younger (eight female, mean age 24.1 years, age range 20–30 years) and 16 older (eight female, mean age 63.3 years, age range 55–69 years) active car drivers (at least two drives per week during the last 3 years) participated in the study. The data of one (older) participants were excluded from analysis due to profound EEG artifacts. As could be expected, the older drivers hold their driving licenses longer than the younger ones (young: 6.7 years, SE 0.7 years; older: 43.7 years, SE 1.7 years;  $t_{(29)} = 20.66$ ;  $p < 0.001$ ), but the two groups did not differ in their mean annual mileage (young: 12207 km, SE 4435 km; older: 12455 km, SE 1609 km;  $t_{(24)} = 0.05$ ;  $p > 0.05$ ; reduced number of participants). All participants had normal or corrected-to-normal vision, and none of them reported any known neurological or psychiatric disorder. They received 30 € for participation in the experiment and provided written informed consent prior to entering the experiment. This study was carried out in accordance with the recommendations of Code of Ethics

of the World Medical Association (Declaration of Helsinki). The protocol was approved by the local Ethical Committee of the Leibniz Research Centre for Working Environment and Human Factors, Dortmund, Germany. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

### Task and Procedure

The experiment took place in a static driving simulator, consisting of three 32 inch displays and 200 degrees horizontal field of view (ST Sim; ST Software B.V. Groningen, Netherlands). The participants' task was to keep a car on track on a one-lane road with curves of varying radii. The driving environment consisted of monotonous grassland without any additional visual distraction. The driving speed was held constant at 31 mph to prevent the participants from compensatory slowing down in narrow curves. The radii of the curves varied randomly every 2 min between three levels (task load: low, middle, high): At the low level, there was a straight road without curves. At the high level, the curves were adjusted to radii that allowed the participants to keep the car on track for about 95% of the time (as determined in a pilot study). At the middle level, the curve radii were distributed in between the low and high levels. Left and right turns varied randomly within each curve segment. To smooth the transfer between adjacent segments and to avoid abrupt changes, 1-s transfer-intervals were introduced. Three different curve segments were combined to triplets in randomized order, each lasting for 6 min. The first triplet was used to familiarize the participants with the task. This practice triplet was followed by nine experimental triplets that were separated into three blocks. All in all, the experimental blocks lasted for 54 min without any break or interruption.

During the entire experiment a random sequence of short auditory stimuli was presented via two broad-band loudspeakers in front of the participants at a sound level of 70 dB (A). The duration of each stimulus was 100 ms (including 5 ms rise and fall times) and the interstimulus-onset interval was 1,000 ms. The majority of the stimuli (80%) were standards intermixed with 10% higher and 10% lower deviants. The standard stimulus was a harmonic tone composed of three sinusoidal partials of 500, 1,000 and 1,500 Hz, with the intensity of the second and third partials being lower than that of the first partial by 3 and 6 dB, respectively. The two deviant stimuli differed from the standard stimulus in frequency, being either 10% higher (partials: 550, 1,100, 1,650 Hz) or lower (450, 900, 1,350 Hz) than the standard. The tones represented irrelevant distractor stimuli that should be ignored by the participants.

In order to measure possible changes in subjective fatigue over the driving session, the participants filled out the German version of the Stanford Sleepiness Scale (Hoddes et al., 1973) immediately before and after the driving session.

### Data Recording

EEG was recorded by 64 scalp electrodes placed according to the International 10-10 system (electrode impedance below 10 k $\Omega$ ) and a "BioSemi active 2" amplifier (BioSemi, Netherlands; sampling rate 2,048 Hz, bandwidth DC–140 Hz).



Six additional electrodes positioned around both eyes were used for electrooculography to measure horizontal and vertical eye positions. Two additional electrodes were placed on the left and right mastoids.

## Data Analysis

### Behavioral Data

Two different behavioral measures were analyzed, the time off track and the steering variability. Time off track was defined as the percentage of time that the car left the track. It was used as main index for individual driving accuracy. Steering variability was defined as the steering activity in degrees per second and was operationalized as an index for workload (Verwey and Veltman, 1996). Both parameters were entered into mixed-design ANOVAs with the between-subject factor Age (2; young, old) and the within-subject factors Task Load (3; low, medium, high) and Time on Task (3; Blocks 1, 2, 3).

### EEG Data

After re-referencing the EEG data to common average reference, a bandpass filter between 0.1 Hz and 35 Hz was applied. Broken channels were detected and excluded based on kurtosis and probability criteria. Afterwards, the filtered data were resampled to 128 Hz (dataset 1) and epoched into segments ranging from –600 ms to 1,200 ms with respect to the onset of the sound stimuli. For the ERP analysis, 1,024 Hz sampled data (dataset 2) were segmented, also ranging from –600 ms to 1,200 ms, and a baseline ranging from –200 ms to 0 ms was subtracted. Segments containing artifacts were identified in dataset 1 and removed from both datasets 1 and 2. An independent component analysis (ICA) was performed on dataset 1 and ICs reflecting artifacts were identified using ADJUST (Mognon et al., 2011). The IC weights were then copied to dataset 2 to again remove artifactual ICs from both datasets.

The spectral properties of the EEG were obtained by calculating Fast Fourier Transformations on dataset 1. Due to substantial differences in the raw spectra between the two age groups, a two-step analysis was chosen. First, to address the different levels in general power, total power between 3 Hz and 30 Hz was calculated. Thereafter, the mean power was extracted for the Theta band (4–7 Hz) and the Alpha band (8–12 Hz) to subsequently compute relative power values. The proportional contribution of Theta and Alpha power to total power was entered into analyses.

Due to relatively liberal criteria of the statistical rejection of segments, an additional amplitude criterion was applied to the data (maximum voltage difference of  $\pm 50 \mu\text{V}$  per segment). To compute the deviance-related MMN and P3a components, difference waveforms were calculated by subtracting the standard-tone ERPs from the deviant-tone ERPs. For analysis of MMN and P3a amplitudes, the fronto-central FCz electrode was chosen where the most prominent responses are usually obtained (for reviews see Escera et al., 2000; Näätänen et al., 2007). The MMN and P3a amplitudes were calculated as a mean voltage within the 40-ms period centered at the peak latencies in the grand-average waveforms (MMN: 125 ms; P3a: 230 ms; relative to tone onsets).

It should be noted that the sequence of the tone stimuli temporarily overlapped with the driving task, and that the oscillatory measures were computed in epochs, in which the cortical processing of the auditory standards and deviants took place. Alpha and Theta frequency bands represent important portions of auditory event-related oscillations (e.g., Kolev and Yordanova, 1997; Yordanova et al., 2000) and possible effects of these event-related oscillations might be assumed. However, given that the tone sequence was kept constant throughout the driving session, and because the analysis was mainly focused on within-subjects effect of task workload and time on task as well as the interaction of these factors with age, such effects of event-related oscillations should not play a significant role for the analysis of Alpha and Theta power measures.

Total power, percentages of total power in Theta and Alpha bands (relative power), and MMN and P3a amplitudes were entered into mixed-design ANOVAs with the between-subject factor Age (2; young, old) and within-subject factors Task Load (3; low, medium, high) and Time on Task (3; Blocks 1, 2, 3). The statistical analysis of the spectral power of the EEG based upon the averaged spectrograms of four anterior channels (F1, Fz, F2, FCz) and four posterior channels (PO3, POz, Pz, PO4), respectively. Average measures of electrode patches were used in order to gain a better stability of these measures by accounting for minor topographical deviations of spectral power across subjects. Anterior and posterior electrode patches were chosen as spectral power modulations at these locations have been linked to changes in mental states affecting performance. Theta power, especially at frontal recording sites, was shown to reflect cognitive effort and the exertion of cognitive control, thus also reflecting task demands (Jensen and Tesche, 2002; Maurer et al., 2014). Posterior alpha power, on the other hand, has been linked to cognitive disengagement and sensory withdrawal (Hanslmayr et al., 2011). In the context of time on task effects, recent studies report reliable increases of alpha power at anterior leads (Barwick et al., 2012; Wascher et al., 2014; Fan et al., 2015). False discovery rate (FDR) correction was applied to account for this multiple testing, and only FDR-corrected  $p$ -values are provided. Effect sizes were provided for interpretation of the practical significance of the results, using the partial eta-squared coefficient ( $\eta_p^2$ ).

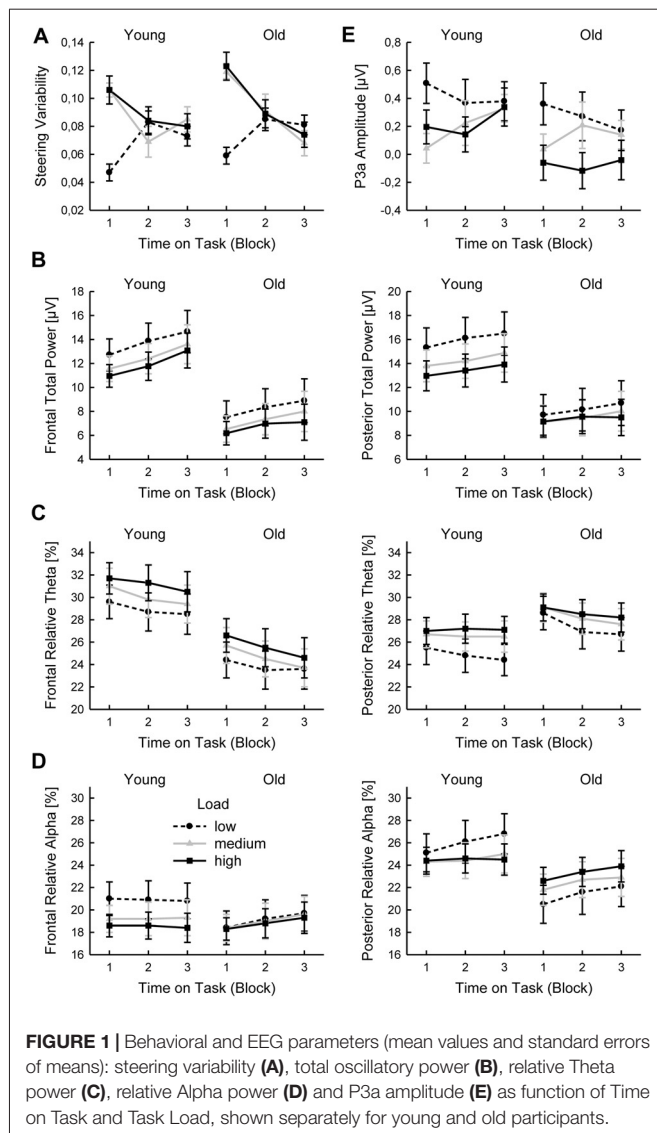
### Subjective Data

Sleepiness ratings on the Stanford Sleepiness Scale were entered into mixed-design ANOVAs with the between-subject factor Age (2; young, old) and the within-subject factors Time (2; pre, post). In addition, changes in subjective mental states over the course of the driving session were computed as relative differences in sleepiness before and after driving, using the formula  $(\text{Post} - \text{Pre}) / \text{Pre} * 100\%$ .

## RESULTS

### Behavioral Data

Time off track was slightly increased with higher Task Load ( $F_{(2,58)} = 2.75$ ;  $p = 0.09$ ;  $\eta_p^2 = 0.09$ ) but did not depend on Time



on Task or Age. None of the interactions reached significance (all  $p > 0.22$ ; all  $\eta_p^2 < 0.05$ ).

Steering variability was overall higher with higher Task Load ( $F_{(2,58)} = 9.08$ ;  $p < 0.005$ ;  $\eta_p^2 = 0.24$ ) and decreased with Time on Task ( $F_{(2,58)} = 7.77$ ;  $p < 0.005$ ;  $\eta_p^2 = 0.21$ ). Also, there was an interaction of Task Load and Time on Task ( $F_{(4,116)} = 12.72$ ;  $p < 0.001$ ;  $\eta_p^2 = 0.31$ ) that was due a decrease of steering variability with Time on Task for medium and high task load, but rather an increase for low task load (Figure 1A). There was no main effect of Age ( $F_{(1,29)} = 1.47$ ;  $p = 0.23$ ;  $\eta_p^2 = 0.05$ ) and no further interaction (all  $p > 0.10$ ; all  $\eta_p^2 < 0.09$ ).

## EEG Data

### Total Power

The total power increased significantly with Time on Task (frontal:  $F_{(2,58)} = 11.66$ ;  $p < 0.005$ ;  $\eta_p^2 = 0.29$ ; posterior:  $F_{(2,58)} = 8.00$ ;  $p < 0.005$ ;  $\eta_p^2 = 0.22$ ) and decreased with increasing Task Load (frontal:  $F_{(2,58)} = 24.64$ ;  $p < 0.001$ ;  $\eta_p^2 = 0.46$ ; posterior:

$F_{(2,58)} = 9.85$ ;  $p < 0.005$ ;  $\eta_p^2 = 0.25$ ; Figure 1B). Total power was stronger in the younger than in the older group (frontal:  $F_{(1,29)} = 7.75$ ;  $p < 0.05$ ;  $\eta_p^2 = 0.21$ ; posterior:  $F_{(1,29)} = 5.24$ ;  $p < 0.05$ ;  $\eta_p^2 = 0.15$ ).

### Relative Theta Power

The relative Theta power decreased with Time on Task (frontal:  $F_{(2,58)} = 12.58$ ;  $p < 0.001$ ;  $\eta_p^2 = 0.30$ ; posterior:  $F_{(2,58)} = 8.28$ ;  $p < 0.005$ ;  $\eta_p^2 = 0.22$ ) and was increased with increasing Task Load (frontal:  $F_{(2,58)} = 22.99$ ;  $p < 0.001$ ;  $\eta_p^2 = 0.44$ ; posterior:  $F_{(2,58)} = 18, 10$ ;  $p < 0.001$ ;  $\eta_p^2 = 0.38$ ). The decrease of posterior relative Theta power with Time on Task was more pronounced with lower than with higher Task Load (Time on Task by Task Load interaction:  $F_{(4,116)} = 4.02$ ;  $p < 0.01$ ;  $\eta_p^2 = 0.12$ ; Figure 1C). Frontal relative Theta power was slightly stronger in the younger than in the older group ( $F_{(1,29)} = 5.58$ ;  $p = 0.05$ ;  $\eta_p^2 = 0.16$ ).

### Relative Alpha Power

The posterior relative Alpha power increased with Time on Task ( $F_{(2,58)} = 5.11$ ;  $p < 0.05$ ;  $\eta_p^2 = 0.15$ ), and the frontal relative Alpha power was decreased with higher Task Load ( $F_{(2,58)} = 9.10$ ;  $p < 0.01$ ;  $\eta_p^2 = 0.24$ ). The latter effect on frontal relative Alpha power was more pronounced in the younger than in the older group (Age by Task Load interaction:  $F_{(2,58)} = 5.79$ ;  $p < 0.05$ ;  $\eta_p^2 = 0.17$ ; Figure 1D). Accordingly, separate ANOVAs for the two groups indicated a significant effect of Task Load for the younger ( $F_{(2,60)} = 18.41$ ;  $p < 0.001$ ;  $\eta_p^2 = 0.55$ ), but not the older group ( $F_{(2,56)} = 0.18$ ;  $p = 0.84$ ;  $\eta_p^2 = 0.01$ ), indicating that the increase of relative Alpha with lower Task Load was confined to the younger drivers. On posterior side, the effect of Task Load on relative Alpha power was also modulated by Age (Age by Task Load interaction:  $F_{(2,58)} = 5.85$ ;  $p < 0.05$ ;  $\eta_p^2 = 0.17$ ). Here, separate ANOVAs indicated a significant interaction of Task Load and Time on Task for the younger group ( $F_{(4,60)} = 2.74$ ;  $p < 0.05$ ;  $\eta_p^2 = 0.16$ ), demonstrating a more pronounced increase in relative Alpha power over time with lower than with higher Task Load. In contrast, the older group showed an overall decrease in relative Alpha power with lower Task Load ( $F_{(2,56)} = 4.46$ ;  $p < 0.05$ ;  $\eta_p^2 = 0.24$ ).

### MMN and P3a

The MMN amplitude (mean  $-0.33 \mu V$ , SE  $0.05 \mu V$ ) did not depend on Time on Task or Task Load and did not differ between age group (all  $p > 0.21$ ; all  $\eta_p^2 < 0.06$ ). The P3a amplitude decreased with higher Task Load ( $F_{(2,58)} = 6.33$ ;  $p < 0.005$ ;  $\eta_p^2 = 0.18$ ; Figure 1E) and was slightly higher in the younger group ( $F_{(1,29)} = 3.48$ ;  $p = 0.07$ ;  $\eta_p^2 = 0.11$ ).

## Relationship Between Steering Variability and EEG

In order to evaluate possible relationships between driving performance and mental states, steering variability and EEG measures were averaged across Task Load and Time on Task, and a correlation was calculated for total power, relative Theta and Alpha power, and ERPs separately. The Kendall's Tau correlation coefficient was used as a non-parametric test for the statistical dependance of behavior and EEG measures.

## Total Power

There were no significant correlations of steering variability and total power, neither for the younger (frontal:  $\tau = 0.217$ ; posterior:  $\tau = 0.265$ ), nor for the older group (frontal:  $\tau = -0.067$ ; posterior:  $\tau = -0.105$ ; all  $p > 0.14$ ).

## Relative Theta Power

A significant negative correlation of steering variability and relative Theta power occurred in the older group (frontal:  $\tau = -0.524$ ;  $p = 0.024$ ; posterior:  $\tau = -0.467$ ;  $p = 0.030$ ; FDR-corrected  $p$ -values), indicating that higher Theta power was associated with lower steering variability (**Figure 2A**). No such relationship was found in the younger group (frontal:  $\tau = -0.150$ ; posterior:  $\tau = 0.100$ ; both  $p > 0.55$ ).

## Relative Alpha Power

There was no correlation of steering variability and relative Alpha power, neither for the younger group (frontal:  $\tau = -0.050$ ; posterior:  $\tau = 0.133$ ), nor for the older group (frontal:  $\tau = -0.143$ ; posterior:  $\tau = 0.048$ ; all  $p > 0.45$ ).

## MMN and P3a

There was a significant positive correlation of steering variability and P3a amplitude in the younger group ( $\tau = 0.550$ ;  $p = 0.012$ ), indicating that a more pronounced P3a was associated with higher steering variability (**Figure 2B**). In the older group, the correlation of steering variability and P3a slightly failed to reach significance ( $\tau = 0.333$ ;  $p = 0.083$ ). In addition, there was a trend to a correlation of steering variability and MMN (younger:  $\tau = 0.383$ ;  $p = 0.051$ ; older:  $\tau = 0.410$ ;  $p = 0.051$ ; FDR-corrected  $p$ -values).

## Subjective Data

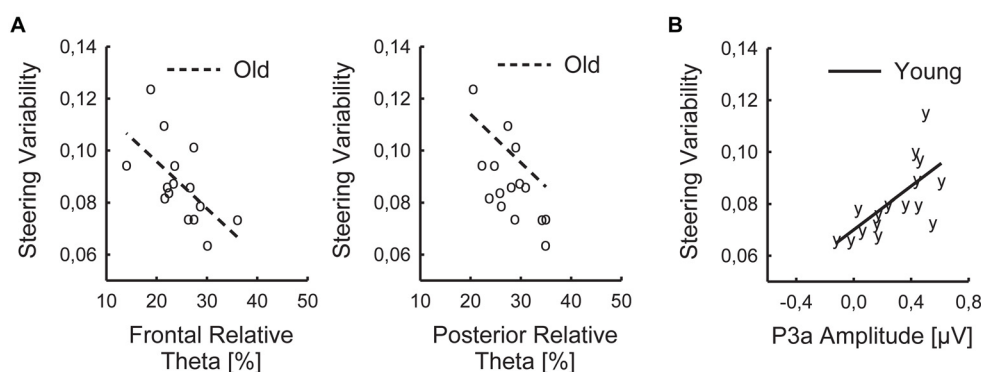
The analysis of the pre and post ratings of fatigue indicated an increase in subjective sleepiness over the course of the driving session ( $F_{(1,29)} = 42.12$ ;  $p < 0.001$ ;  $\eta_p^2 = 0.59$ ; **Figure 3A**). In addition, the younger group scored higher in overall sleepiness ( $F_{(1,29)} = 16.46$ ;  $p < 0.001$ ;  $\eta_p^2 = 0.36$ ) than the older group. A significant interaction of Age and Time on Task on sleepiness

( $F_{(1,29)} = 16.46$ ;  $p < 0.001$ ;  $\eta_p^2 = 0.36$ ) was due to a higher increase in sleepiness in the younger, than older, group.

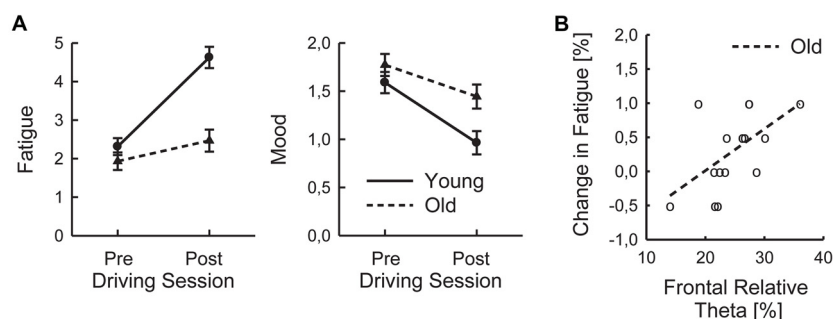
In order to test whether relative frontal Theta power and P3a amplitude, which turned out to be related to steering variability in the older (respective younger) group, were also associated with the subjective fatigue, the relative changes in sleepiness over the driving session were computed and correlations were determined for frontal relative Theta power and P3a amplitude for both groups separately. In the older group, a significant correlation was found for frontal relative Theta ( $\tau = 0.490$ ;  $p = 0.036$ ; FDR-corrected  $p$ -value; **Figure 3B**), indicating that higher relative Theta power during the driving session was associated with a stronger increase in sleepiness. The relationship of frontal relative Theta and change in sleepiness in the younger group did not reach significance ( $\tau = 0.188$ ;  $p = 0.315$ ). No significant correlations of P3a and subjective measures were found, neither for the younger ( $\tau = 0.103$ ), nor for the older group ( $\tau = 0.000$ ; both  $p > 0.58$ ).

## DISCUSSION

Goal of the present study was to explore the interaction of task workload, variations in mental effort, and age-related difference in performance in a simulated (lane-keeping) driving task, using electrophysiological EEG and ERP measures. Lane keeping in general was efficient in younger and older drivers. There was a slight increase in time off track in tighter bends that could be expected. However, there were no significant changes over time, neither in the younger, nor in the older group. Steering variability on curvy road sections was increased relative to straight sections in both groups but decreased with time on task. This could be due to learning effects resulting in a more and more experienced anticipation of steering behavior in dependence of the curve radius. In this regard, it should be noted that the driving scenario allowed for a pro-active driving strategy, in which the future steering behavior could be dynamically adapted to the course of the road in an anticipatory manner. This is in sharp contrast to the cross-wind scenario of our previous



**FIGURE 2 |** Significant relationships of behavioral and EEG parameters: Individual values of steering variability, frontal and posterior relative Theta (**A**), and P3a amplitude (**B**), averaged across Task Load and Time on Task, shown for young (y) and old (o) participants with regression lines.



**FIGURE 3 |** Subjective parameters: reported pre and post driving fatigue ratings (A) and individual changes in fatigue and frontal relative Theta (B) of old (o) participants (as averaged across Task Load and Time on Task) with regression lines.

study (Karthaus et al., 2018), in which the driver could only react to the different degrees of crosswind but had no chance to prepare for an upcoming steering event over a longer time frame.

The analyses of the EEG oscillatory data indicated an overall stronger total power in the younger group. Such age-related differences in oscillatory activity have been reported in a number of previous studies (e.g., Polich, 1997; for review Klimesch, 1999), but the relationship between EEG measures, cognitive performance, and age are still poorly understood and obviously depend on the cognitive functions studied (e.g., Vlahou et al., 2014; Trammell et al., 2017). Beside this age effect, there was an increase in total power over time and a decrease with increasing workload. These effects of time on task and workload were found for frontal and posterior positions and appear plausible, given that strong oscillatory activity in lower bands (in particular the Alpha band) reflects mental states of boredom and attentional disengagement (e.g., Klimesch, 1999, 2012; Hanslmayr et al., 2012; Wascher et al., 2016) as usually observed during monotonous and boring tasks (Borghini et al., 2014). These mental states might be more pronounced on straight road sections and after the participants had become more familiar with the driving task because of automatization of a well-learned task.

A quite similar overall pattern was found for the relative Alpha power. However, there were differences between the two age groups: stronger frontal relative Alpha power and increases in posterior relative Alpha power on straight road sections, relative to curvy sections, were only observed in the younger group. In contrast, in the older group frontal relative Alpha power did not differ between task loads, and posterior relative Alpha power was even smaller on straight road sections. Thus, the afore-mentioned attentional disengagement was only found in the younger group, whereas the older drivers showed less signs of boredom and mental fatigue even when the task was less demanding. This observation can be interpreted with regard to the decline-compensation hypothesis (Cabeza et al., 2002), assuming that older adults maintain or even increase their mental effort to counteract possible age-related neurocognitive declines. Accordingly, neurophysiological studies have frequently demonstrated

greater activation especially in frontal brain areas in older adults than in younger adults, even when the performance did not differ (for review Reuter-Lorenz and Cappell, 2008).

In contrast to relative Alpha power, relative Theta power decreased over time and was stronger with increasing workload. This observation is well in line with the view that Theta power reflects mental effort (e.g., Jensen and Tesche, 2002; Onton et al., 2005; Cavanagh and Frank, 2014; Cavanagh and Shackman, 2015; Arnau et al., 2017), which should decrease over time due to learning and automatisms, but which should increase on demanding road sections. While this interplay of time on task, workload, and Theta power did not differ between the two age groups, a significant negative relationship of relative Theta power and steering variability was found only in the older group: Older drivers with higher Theta power showed a lower steering variability. Assuming that steering variability represents a measure for workload (Verwey and Veltman, 1996), and frontal Theta power constitutes an indicator for mental effort, this negative correlation does not appear plausible at first sight. However, for the interpretation of the present relationship, it should be considered that in a pro-active driving task lower steering variability reflects a more adequate steering behavior that requires (and results from) a better anticipation of the ongoing and future road course (Garcia et al., 2017). In other words: to perform the task successfully, the driver had to perceive the direction and strength of the curves continuously, respond by steering movements, monitor the movements of the car relative to the driving lane, and correct the steering movement if necessary. Higher steering variability is therefore the consequence of deficiencies in this ongoing loop of perceiving, responding, and correction. Lower steering variability, on the other hand, should be associated with higher mental effort in the proactive driving task, as indicated by higher frontal Theta power.

Interestingly, the correlation of steering variability and relative Theta power was significant only in the older group. The fact that the relationship of steering performance and mental effort exclusively occurred in the older group is—on the one hand—in line with the decline-compensation hypothesis



(Cabeza et al., 2002). On the other hand, the differences between the age groups might reflect the higher inter-individual variability in elderly that is usually found (e.g., Hultsch et al., 2002). Higher differences in steering variability in the older group might promote correlations with brain measures on a between-subject level. In addition to differences *between* the older drivers (rather suggesting individual longer-term driving strategies), the question of whether a similar relationship exists on a within-subject level would also be of interest, as this would reflect short-term variability in the interplay of task workload, mental effort and driving outcome. Here, it would be expected that current states of higher mental effort should be accompanied by lower actual steering variability. In order to address this question, recent research approaches using ongoing EEG recordings have been proposed for the assessment of mental workload, fatigue and drowsiness in car drivers (e.g., Borghini et al., 2014; Charbonnier et al., 2016; Kong et al., 2017).

The assumption that better steering behavior of older drivers was associated with higher mental effort is further supported by the significant correlation of relative Theta power and the increase in the subjective measure of fatigue: Older drivers showing higher relative Theta power (indicating higher mental effort while driving) reported a higher pre-/post-increase on the sleepiness scale. Such an association of mental effort and relative Theta power has also been reported in previous studies (Klimesch, 1999; Smith et al., 2001; Smit et al., 2004). Although a similar relationship could be observed in the younger group (see **Figure 3B**), the correlation failed to reach significance. The younger group also reported an overall higher level of sleepiness that even increased more than that of the older group. This suggests a higher tendency of the younger group to respond to monotonous task with boredom and attentional withdrawal than the older group, as it was also found in a previous study (Arnau et al., 2017). This also corresponds nicely to the observation that an increase in posterior relative Alpha power on straight, less demanding road sections was only observed in the younger group (see above).

The analysis of the ERPs to deviance in the acoustic stimulation indicated an MMN and a P3a to the rare deviant tones, relative to the regular standard tone. The MMN is regarded as a correlate of deviance detection at an early, rather pre-attentive level, while the P3a reflects the allocation of attentional resources toward a change in the individual's environment (Näätänen, 1992; for review Näätänen et al., 2007). In classical distraction paradigms, in which participants are instructed to attend to a task-relevant stimulus feature while ignoring task-irrelevant features, the P3a is usually assumed to reflect an involuntary shift of attention toward the distractor, away from the task at hand (Escera et al., 2000). In the present driving task, MMN and P3a can be regarded as correlates of detection of and attention to the deviant tone stimulus. The first process did not depend on workload and time on task (which could be expected, given that the MMN in passive auditory oddball paradigms can be found even in coma patients, e.g., Morlet and Fischer, 2014). The second one, however, decreased in amplitude with higher workload. This can be interpreted in

the framework of a resource allocation approach (for review see Wickens, 2008), in which less allocation of cognitive resources to a secondary task shows that more resources are required by a primary task. In line with this hypothesis, diminished P3 amplitudes to task-irrelevant sounds have been observed in a primary steering task when steering demands increased (Scheer et al., 2016). In the present task setting, keeping track in narrow curves might tie up attentional resources, which are no longer available for attending to the changes in the tone stimuli. In this regard, it should be noted that MMN and P3a are determined as the difference waveforms of ERPs to deviant and standard tones. Thus, the workload effect on P3a amplitude does not simply mirror the pattern of total oscillatory power but reflects genuine workload-related differences in stimulus deviance processing. The slightly higher P3a amplitude in younger participants is in line with the literature (Polich, 1997). More importantly, the younger group showed a significant correlation of P3a amplitude and steering variability, suggesting that worse steering performance of younger drivers came along with a higher involuntary shift of attention toward the deviant tone stimuli. Conversely, younger drivers who attended less to changes in the environment showed a better driving performance. This relationship clearly stresses the implication of the P3a to task-irrelevant sound as an indirect measure for evaluating mental workload, which has also been shown previously (Scheer et al., 2016). Furthermore, assuming that shifts of attention toward task-irrelevant stimulation are closely related to distraction, the present results suggest that the susceptibility to distraction of younger and older drivers is based on different cognitive mechanisms.

As a final methodological remark, it should be noted that the most prominent results of the present study were found in relative values of Theta and Alpha power, whereas the analysis of total oscillatory power indicated only basic effects of workload and time on task. Thus, relative power in EEG bands appears to be the more fine-grained measure that should be considered in future studies.

## CONCLUSION

In the present pro-active driving scenario, younger and older drivers did not differ in behavioral measures of performance, i.e., in time off track and steering variability. However, electrophysiological measures demonstrated age-related differences in the way the two groups reached this goal: high performance was associated with increased mental effort (and a higher increase in fatigue) in the older group. Yet, high performance in the younger group was related to higher attentional focusing on the driving task (respective less attention to distractors). What do these results mean for traffic safety? On the one hand, it appears that older drivers invest extra mental resources to manage demanding (driving) tasks. Thus, they should always start a longer trip sufficiently rested and otherwise fit for driving. Also, it is advisable for them to take breaks regularly. Driver assistance systems detecting drowsiness can additionally help to avoid driving situations in which

reduced mental resources might lead to critical driving behavior. On the other hand, younger drivers showing a tendency to attend to traffic-irrelevant stimulation should avoid being distracted by this information, especially in monotonous driving situations. Here, more initiatives towards younger drivers addressing the dangers of using mobile phones and communication devices while driving would be recommended.

## AUTHOR CONTRIBUTIONS

SG, MK and EW: study conception and design. MK: acquisition of data. SG, MK, SA, JER and EW: analysis and interpretation of data; critical revision. SG and EW: drafting of manuscript.

## REFERENCES

- Anstey, K. J., and Wood, J. (2011). Chronological age and age-related cognitive deficits are associated with an increase in multiple types of driving errors in late life. *Neuropsychology* 25, 613–621. doi: 10.1037/a0023835
- Arnau, S., Möckel, T., Rinkenauer, G., and Wascher, E. (2017). The interconnection of mental fatigue and aging: an EEG study. *Int. J. Psychophysiol.* 117, 17–25. doi: 10.1016/j.ijpsycho.2017.04.003
- Barwick, F., Arnett, P., and Slobounov, S. (2012). EEG correlates of fatigue during administration of a neuropsychological test battery. *Clin. Neurophysiol.* 123, 278–284. doi: 10.1016/j.clinph.2011.06.027
- Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., and Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev.* 44, 58–75. doi: 10.1016/j.neubiorev.2012.10.003
- Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends Cogn. Sci.* 16, 106–113. doi: 10.1016/j.tics.2011.12.010
- Brookhuis, K. A., and de Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. *Accid. Anal. Prev.* 42, 898–903. doi: 10.1016/j.aap.2009.06.001
- Brouwer, A.-M., Hogervorst, M. A., van Erp, J. B. F., Heffelaar, T., Zimmerman, P. H., and Oostenveld, R. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. *J. Neural Eng.* 9:045008. doi: 10.1088/1741-2560/9/4/045008
- Cabeza, R., Anderson, N. D., Locantore, J. K., and McIntosh, A. R. (2002). Aging gracefully: compensatory brain activity in high-performing older adults. *Neuroimage* 17, 1394–1402. doi: 10.1006/nimg.2002.1280
- Cantin, V., Lavallière, M., Simoneau, M., and Teasdale, N. (2009). Mental workload when driving in a simulator: effects of age and driving complexity. *Accid. Anal. Prev.* 41, 763–771. doi: 10.1016/j.aap.2009.03.019
- Cavanagh, J. F., and Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends Cogn. Sci.* 18, 414–421. doi: 10.1016/j.tics.2014.04.012
- Cavanagh, J. F., and Shackman, A. J. (2015). Frontal midline theta reflects anxiety and cognitive control: meta-analytic evidence. *J. Physiol. Paris* 109, 3–15. doi: 10.1016/j.jphysparis.2014.04.003
- Charbonnier, S., Roy, R. N., Bonnet, S., and Campagne, A. (2016). EEG index for control operators' mental fatigue monitoring using interactions between brain regions. *Expert Syst. Appl.* 52, 91–98. doi: 10.1016/j.eswa.2016.01.013
- Da Silva, F. P. (2014). Mental workload, task demand driving performance: what relation? *Procedia Soc. Behav. Sci.* 162, 310–319. doi: 10.1016/j.sbspro.2014.12.212
- De Waard, D. (2002). "Mental workload," in *Human Factors for Highway Engineers*, eds R. Fuller and J. A. Santos (Netherlands: Pergamon Press), 161–175.
- Dijksterhuis, C., Brookhuis, K. A., and De Waard, D. (2011). Effects of steering demand on lane keeping behaviour, self-reports, and physiology. *Accid. Anal. Prev.* 43, 1074–1081. doi: 10.1016/j.aap.2010.12.014
- Escera, C., Alho, K., Schröger, E., and Winkler, I. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiol. Neurotol.* 5, 151–166. doi: 10.1159/000013877

## FUNDING

The study was supported by a grant from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG KA 4120/2-1) to MK. The publication of this article was funded by the Open Access Fund of the Leibniz Association.

## ACKNOWLEDGMENTS

We are grateful to Ludger Blanke for technical assistance and to Christiane Westedt for her help in running the experiments, and to the two reviewers for their very constructive comments on an earlier version of this manuscript.

- Fan, X., Zhou, Q., Liu, Z., and Xie, F. (2015). Electroencephalogram assessment of mental fatigue in visual search. *Biomed. Mater. Eng.* 26, S1455–S1463. doi: 10.3233/bme-151444
- Garcia, J. O., Brooks, J., Kerick, S., Johnson, T., Mullen, T. R., and Vettel, J. M. (2017). Estimating direction in brain-behavior interactions: proactive and reactive brain states in driving. *Neuroimage* 150, 239–249. doi: 10.1016/j.neuroimage.2017.02.057
- Hanslmayr, S., Gross, J., Klimesch, W., and Shapiro, K. L. (2011). The role of  $\alpha$  oscillations in temporal attention. *Brain Res. Rev.* 67, 331–343. doi: 10.1016/j.brainresrev.2011.04.002
- Hanslmayr, S., Staudigl, T., and Fellner, M.-C. (2012). Oscillatory power decreases and long-term memory: the information via desynchronization hypothesis. *Front. Hum. Neurosci.* 6:74. doi: 10.3389/fnhum.2012.00074
- Herrmann, C. S., and Knight, R. T. (2001). Mechanisms of human attention: event-related potentials and oscillations. *Neurosci. Biobehav. Rev.* 25, 465–476. doi: 10.1016/s0149-7634(01)00027-6
- Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., and Dement, W. C. (1973). Quantification of sleepiness: a new approach. *Psychophysiology* 10, 431–436. doi: 10.1111/j.1469-8986.1973.tb00801.x
- Hultsch, D. F., MacDonald, S. W., and Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 57, P101–P115. doi: 10.1093/geronb/57.2.p101
- Jensen, O., and Tesche, C. D. (2002). Frontal theta activity in humans increases with memory load in a working memory task. *Eur. J. Neurosci.* 15, 1395–1399. doi: 10.1046/j.1460-9568.2002.01975.x
- Karthus, M., Wascher, E., and Getzmann, S. (2018). Proactive vs. reactive car driving: EEG evidence for different driving strategies of older drivers. *PLoS One* 13:e0191500. doi: 10.1371/journal.pone.0191500
- Klimesch, W. (1999). EEG  $\alpha$  and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res. Rev.* 29, 169–195. doi: 10.1016/s0165-0173(98)00056-3
- Klimesch, W. (2012).  $\alpha$ -band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci.* 16, 606–617. doi: 10.1016/j.tics.2012.10.007
- Kolev, V., and Yordanova, J. (1997). Analysis of phase-locking is informative for studying event-related EEG activity. *Biol. Cybern.* 76, 229–235. doi: 10.1007/s004220050335
- Kong, W., Zhou, Z., Jiang, B., Babiloni, F., and Borghini, G. (2017). Assessment of driving fatigue based on intra/inter-region phase synchronization. *Neurocomputing* 219, 474–482. doi: 10.1016/j.neucom.2016.09.057
- Kramer, A. F., Trejo, L. J., and Humphrey, D. (1995). Assessment of mental workload with task-irrelevant auditory probes. *Biol. Psychol.* 40, 83–100. doi: 10.1016/0301-0511(95)05108-2
- Lin, C.-T., Chuang, C.-H., Kerick, S., Mullen, T., Jung, T.-P., Ko, L.-W., et al. (2016). Mind-wandering tends to occur under low perceptual demands during driving. *Sci. Rep.* 6:21353. doi: 10.1038/srep21353
- Lindenberger, U., and Ghisletta, P. (2009). Cognitive and sensory declines in old age: gauging the evidence for a common cause. *Psychol. Aging* 24, 1–16. doi: 10.1037/a0014986
- Massar, S. A., Wester, A. E., Volkerts, E. R., and Kenemans, J. L. (2010). Manipulation specific effects of mental fatigue: evidence from novelty



- processing and simulated driving. *Psychophysiology* 47, 1119–1126. doi: 10.1111/j.1469-8986.2010.01028.x
- Maurer, U., Brem, S., Liechti, M., Maurizio, S., Michels, L., and Brandeis, D. (2014). Frontal midline theta reflects individual task performance in a working memory task. *Brain Topogr.* 28, 127–134. doi: 10.1007/s10548-014-0361-y
- Meng, A., and Siren, A. (2012). Cognitive problems, self-rated changes in driving skills, driving-related discomfort and self-regulation of driving in old drivers. *Accid. Anal. Prev.* 49, 322–329. doi: 10.1016/j.aap.2012.01.023
- Mognon, A., Jovicich, J., Bruzzone, L., and Buiatti, M. (2011). ADJUST: an automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology* 48, 229–240. doi: 10.1111/j.1469-8986.2010.01061.x
- Molnar, L. J., and Eby, D. W. (2008). The relationship between self-regulation and driving-related abilities in older drivers: an exploratory study. *Traffic Inj. Prev.* 9, 314–319. doi: 10.1080/15389580801895319
- Morlet, D., and Fischer, C. (2014). MMN and Novelty P3 in coma and other altered states of consciousness: a review. *Brain Topogr.* 27, 467–479. doi: 10.1007/s10548-013-0335-5
- Näätänen, R. (1992). *Attention and Brain Function*. Hillsdale, NJ: Erlbaum.
- Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590. doi: 10.1016/j.clinph.2007.04.026
- Onton, J., Delorme, A., and Makeig, S. (2005). Frontal midline EEG dynamics during working memory. *Neuroimage* 27, 341–356. doi: 10.1016/j.neuroimage.2005.04.014
- Park, D. C. (2000). “The basic mechanisms accounting for age-related decline in cognitive function,” in *Cognitive Aging: A Primer*, eds D. C. Park and N. Schwarz (New York, NY: Psychology Press), 3–21.
- Park, S. W., Choi, E. S., Lim, M. H., Kim, E. J., Hwang, S. I., Choi, K. I., et al. (2011). Association between unsafe driving performance and cognitive-perceptual dysfunction in older drivers. *PM R* 3, 198–203. doi: 10.1016/j.pmrj.2010.12.008
- Polich, J. (1997). EEG and ERP assessment of normal aging. *Electroencephalogr. Clin. Neurophysiol.* 104, 244–256. doi: 10.1016/s0168-5597(97)96139-6
- Reuter-Lorenz, P. A., and Cappell, K. A. (2008). Neurocognitive aging and the compensation hypothesis. *Curr. Dir. Psychol. Sci.* 17, 177–182. doi: 10.1111/j.1467-8721.2008.00570.x
- Scheer, M., Bülthoff, H., and Chuang, L. (2016). Steering demands diminish the early-P3, late-P3 and RON components of the event-related potential of task-irrelevant environmental sounds. *Front. Hum. Neurosci.* 10:73. doi: 10.3389/fnhum.2016.00073
- Smit, A. S., Eling, P. A., and Coenen, A. M. (2004). Mental effort affects vigilance enduringly: after-effects in EEG and behavior. *Int. J. Psychophysiol.* 53, 239–243. doi: 10.1016/j.ijpsycho.2004.04.005
- Smith, M. E., Gevins, A., Brown, H., Karnik, A., and Du, R. (2001). Monitoring task loading with multivariate EEG measures during complex forms of human-computer interaction. *Hum. Factors* 43, 366–380. doi: 10.1518/001872001775898287
- Trammell, J. P., MacRae, P. G., Davis, G., Bergstedt, D., and Anderson, A. E. (2017). The relationship of cognitive performance and the theta- $\alpha$  power ratio is age-dependent: an EEG study of short term memory and reasoning during task and resting-state in healthy young and old adults. *Front. Aging Neurosci.* 9:364. doi: 10.3389/fnagi.2017.00364
- Verwey, W. B., and Veltman, H. A. (1996). Detecting short periods of elevated workload: a comparison of nine workload assessment techniques. *J. Exp. Psychol. Appl.* 2, 270–285. doi: 10.1037/1076-898x.2.3.270
- Vlahou, E. L., Thurm, F., Kolassa, I. T., and Schlee, W. (2014). Resting-state slow wave power, healthy aging and cognitive performance. *Sci. Rep.* 4:5101. doi: 10.1038/srep05101
- Wascher, E., Arnau, S., Gutberlet, I., Karthaus, M., and Getzmann, S. (2018). Evaluating mental states in pro-active and re-active driving scenarios. *Front. Hum. Neurosci.* 12:205. doi: 10.3389/fnhum.2018.00205
- Wascher, E., Getzmann, S., and Karthaus, M. (2016). Driver state examination—Treading new paths. *Accid. Anal. Prev.* 91, 157–165. doi: 10.1016/j.aap.2016.02.029
- Wascher, E., Rasch, B., Sängner, J., Hoffmann, S., Schneider, D., Rinkenauer, G., et al. (2014). Frontal theta activity reflects distinct aspects of mental fatigue. *Biol. Psychol.* 96, 57–65. doi: 10.1016/j.biopsycho.2013.11.010
- Wickens, C. D. (2008). Multiple resources and mental workload. *Hum. Factors* 50, 449–455. doi: 10.1518/001872008x288394
- Yang, B., Xiao, W., Liu, X., Wu, S., and Miao, D. (2013). Mental fatigue impairs pre-attentive processing: a MMN study. *Neurosci. Lett.* 532, 12–16. doi: 10.1016/j.neulet.2012.08.080
- Yordanova, J., Devrim, M., Kolev, V., Ademoglu, A., and Demiralp, T. (2000). Multiple time-frequency components account for the complex functional reactivity of P300. *Neuroreport* 11, 1097–1103. doi: 10.1097/00001756-200004070-00038
- Yucel, G., Petty, C., McCarthy, G., and Belger, A. (2005). Graded visual attention modulates brain responses evoked by task-irrelevant auditory pitch changes. *J. Cogn. Neurosci.* 17, 1819–1828. doi: 10.1162/089892905775008698
- Zhang, P., Chen, X., Yuan, P., Zhang, D., and He, S. (2006). The effect of visuospatial attentional load on the processing of irrelevant acoustic distractors. *Neuroimage* 33, 715–724. doi: 10.1016/j.neuroimage.2006.07.015

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Getzmann, Arnau, Karthaus, Reiser and Wascher. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Recognizing Frustration of Drivers From Face Video Recordings and Brain Activation Measurements With Functional Near-Infrared Spectroscopy

Klas Ihme<sup>1\*</sup>, Anirudh Unni<sup>2</sup>, Meng Zhang<sup>1</sup>, Jochem W. Rieger<sup>2</sup> and Meike Jipp<sup>1</sup>

<sup>1</sup>Department of Human Factors, Institute of Transportation Systems, German Aerospace Center (DLR), Braunschweig, Germany, <sup>2</sup>Department of Psychology, University of Oldenburg, Oldenburg, Germany

## OPEN ACCESS

### Edited by:

Guido P. H. Band,  
Leiden University, Netherlands

### Reviewed by:

Paola Pinti,  
University College London,  
United Kingdom  
Edmund Wascher,  
Leibniz-Institut für Arbeitsforschung  
an der TU Dortmund (IfADo),  
Germany

### \*Correspondence:

Klas Ihme  
klas.ihme@dlr.de

**Received:** 17 April 2018

**Accepted:** 25 July 2018

**Published:** 17 August 2018

### Citation:

Ihme K, Unni A, Zhang M, Rieger JW  
and Jipp M (2018) Recognizing  
Frustration of Drivers From Face  
Video Recordings and Brain  
Activation Measurements With  
Functional Near-Infrared  
Spectroscopy.  
*Front. Hum. Neurosci.* 12:327.  
doi: 10.3389/fnhum.2018.00327

Experiencing frustration while driving can harm cognitive processing, result in aggressive behavior and hence negatively influence driving performance and traffic safety. Being able to automatically detect frustration would allow adaptive driver assistance and automation systems to adequately react to a driver's frustration and mitigate potential negative consequences. To identify reliable and valid indicators of driver's frustration, we conducted two driving simulator experiments. In the first experiment, we aimed to reveal facial expressions that indicate frustration in continuous video recordings of the driver's face taken while driving highly realistic simulator scenarios in which frustrated or non-frustrated emotional states were experienced. An automated analysis of facial expressions combined with multivariate logistic regression classification revealed that frustrated time intervals can be discriminated from non-frustrated ones with accuracy of 62.0% (mean over 30 participants). A further analysis of the facial expressions revealed that frustrated drivers tend to activate muscles in the mouth region (chin raiser, lip pucker, lip pressor). In the second experiment, we measured cortical activation with almost whole-head functional near-infrared spectroscopy (fNIRS) while participants experienced frustrating and non-frustrating driving simulator scenarios. Multivariate logistic regression applied to the fNIRS measurements allowed us to discriminate between frustrated and non-frustrated driving intervals with higher accuracy of 78.1% (mean over 12 participants). Frustrated driving intervals were indicated by increased activation in the inferior frontal, putative premotor and occipito-temporal cortices. Our results show that facial and cortical markers of frustration can be informative for time resolved driver state identification in complex realistic driving situations. The markers derived here can potentially be used as an input for future adaptive driver assistance and automation systems that detect driver frustration and adaptively react to mitigate it.

**Keywords:** frustration, driver state recognition, facial expressions, functional near-infrared spectroscopy, adaptive automation

## INTRODUCTION

Imagine driving through a city during rush hour on the way to an important meeting. You started a little late and realize that with the dense traffic conditions, it will be hard to arrive at the meeting in time. You are becoming increasingly annoyed by the driver in front of you who is driving provocatively slowly and causes unnecessary extra stops at traffic lights. In addition, the myriads of construction sites along your way further worsen the situation. After yet another red light, you are really frustrated and it appears unbearable to you to wait behind the bus right after the light turned green. You accelerate to overtake the bus but fail to see the pedestrian crossing the street and heading to the bus.

The above story is one example of how frustration can affect driving in a negative way and most readers have likely experienced one or the other situation. Frustration can be seen as an aversive emotional state resulting when goal-directed behavior is blocked (Lazarus, 1991) that is associated with negative valence and slightly elevated arousal (Russell, 1980; Scherer, 2005). As driving is a goal-directed behavior (e.g., reaching the destination in time in the above example), blocking events, as described above, can induce frustration and eventually lead to aggressive (driving) behavior (Ekman and Friesen, 2003; Lee, 2010; Jeon, 2015). In addition, frustration can have negative effects on cognitive processes important for driving such as attention, judgment and decision making (Jeon, 2015). Jeon (2012) suggested that negative emotions may have a worse influence on driving performance than distraction or secondary tasks, as drivers normally are aware of the secondary task and can pro-actively compensate for it. Conversely, negative emotions may degrade driving performance without attempts for compensation (Jeon, 2012). Taken together, reducing frustration during driving is an important step towards improving road safety (e.g., zero-vision of the European Commission, as expressed in the White Paper on Transportation, European Commission, 2011).

In order to reduce frustration and its potential negative consequences in future intelligent driver assistance systems by means of emotion-aware systems (e.g., Picard and Klein, 2002), it is necessary to automatically assess the drivers' current level of frustration. One potential indicator of emotions is the momentary facial expression. Humans use facial expressions to communicate their emotions and these facial expressions appear to be relatively idiosyncratic to specific emotions (Ekman and Friesen, 2003; Erickson and Schulkin, 2003) and hence may discriminate between different emotions. Moreover, brain activations are the physiological basis of emotions, appraisal processes as well as subjective experiences (Scherer, 2005) and may allow to objectively discriminate frustrated from non-frustrated subjective states. Therefore, we will investigate whether facial expressions and brain activation patterns are indicative for frustration while driving.

Humans communicate emotions by changing the configuration of the activation of their facial muscles which is fundamental to understand each other in social interaction (Erickson and Schulkin, 2003). Following from this, the idea is

to equip machines like vehicles with the same capability to read facial expressions in order to gain the ability to interpret the driver's current emotional state and eventually become empathic (e.g., Bruce, 1992; Picard and Klein, 2002). This vision becomes realistic with the recent progress in the fields of image processing and machine learning making it possible to automatically track changes in facial activity from video recordings of the face (Bartlett et al., 2008; Hamm et al., 2011; Gao et al., 2014). Still, to the best of our knowledge, so far, only few studies investigated the facial features accompanying frustration and whether these can be used to discriminate frustration from other emotional states. For example, a study by Malta et al. (2011) used facial features to detect frustration in real world situations but did not report the discriminative features. Studies from human-computer interaction (HCI) linked frustration to increased facial muscle movement in the eye brow and mouth area (D'Mello et al., 2005; Grafsgaard et al., 2013). In addition, a recent study investigating facial activity of frustrated drivers found that muscles in the mouth region (e.g., tightening and pressing of lips) were more activated when participants were frustrated compared to a neutral affective state (Ihme et al., in press). However, the authors employed a manual technique for coding the facial muscle behavior and did not evaluate the potential of automatic frustration recognition. Based on these earlier results, we reasoned that *automated* recognition of frustration is possible by combining lip, mouth and eye brow movements.

Still, the goal of this work is not only to evaluate whether it is possible to discriminate frustrated from non-frustrated drivers based on video recordings from the face, but also to describe the patterns of facial muscle configuration related to frustration. For this, we used a multistep approach. First of all, we used a tool to extract the activity of facial action units (AUs) frame by frame from video recordings of the face (we used a commercial tool based on Bartlett et al., 2008). AUs are concepts from the Facial Action Coding System (FACS, Ekman et al., 2002) that can be regarded as the atomic units of facial behavior related to activation of facial muscles. The frame-wise activations of the facial AUs were then used as an input for time-resolved prediction of participants' frustration using a machine learning approach which served to evaluate whether an automated discrimination of frustration is possible. In a second step, we aimed to identify the AU activations patterns that are indicative for frustration. For this, we clustered the frame-wise AU data in order to derive frequently occurring facial muscle configurations. Because facial expressions are described as momentary configurations of AU activations (Ekman, 1993), the resulting cluster centroids can be interpreted as representations of the frequently occurring facial expressions. The AU activations in the cluster centroids are then used to describe which AUs are activated and compared with previous results on facial expressions of frustration (D'Mello et al., 2005; Grafsgaard et al., 2013; Ihme et al., in press). In this way, we can determine which facial expressions are shown by frustrated drivers and whether these are in line with our expectation that facial expressions of frustration are related to lip, mouth and eye brow movements.

Only few studies investigated neural correlates of frustration despite it being a common emotional state. A functional

magnetic resonance imaging (fMRI) study by Siegrist et al. (2005) investigated chronic social reward frustration experienced in a mental calculation task and revealed neural correlates of reward frustration primarily in the medial prefrontal, anterior cingulate and the dorsolateral prefrontal cortex (DLPFC). Bierzyńska et al. (2016) induced frustration in a somatosensory discrimination task. The fMRI results revealed increased activation in the striatum, cingulate cortex, insula, middle frontal gyrus and precuneus with increasing frustration. Another fMRI study from Yu et al. (2014) using speeded reaction times found that experienced frustration correlated with brain activation in PFC and in deep brain structures like the amygdala, and the midbrain periaqueductal gray (PAG). These authors suggested that experienced frustration can serve as an energizing function translating unfulfilled motivation into aggressive-like surges via a cortical, amygdala and PAG network (Yu et al., 2014). Interestingly, other fMRI studies suggested a role of the anterior insula in the subjective experience of feeling frustrated (Abler et al., 2005; Brass and Haggard, 2007). Together, these fMRI studies provided detailed anatomical information about potential neural correlates of frustration, but mostly employed relatively simple experimental paradigms to induce frustration. Thus, in our study, we considered it desirable to employ a brain imaging technology that is better compatible with real-world applications in vehicles.

Functional near-infrared spectroscopy (fNIRS) is a non-invasive optical imaging technique that uses near-infrared light (600–900 nm) to measure cerebral blood flow changes when neural activity is elicited (Jöbsis, 1977; Villringer et al., 1993) based on neurovascular coupling similar to fMRI. fNIRS can be used to measure brain activation in realistic driving simulations and is relatively robust to movement artifacts (Unni et al., 2017). However, fNIRS measurements focus on cortical activation and their spatial resolution (around 3 cm) is lower than fMRI. Perlman et al. (2014) recorded fNIRS data from prefrontal cortical areas in 3–5-year-old children while they played a computer game where the expected prize was sometimes stolen by an animated dog to induce frustration. The results suggest a role for the lateral PFC in emotion regulation during frustration. Hirshfield et al. (2014) induced frustration by slowing down the internet speed while participants performed the task of shopping online for the least expensive model of a specified product given limited time constraints. The fNIRS results indicate increased activation in the DLPFC and the middle temporal gyrus when frustrated. A more recent study by Grabell et al. (2018) investigated the association between the prefrontal activation from fNIRS measurements and irritability scores in children. These authors reported an inverted U-shaped function between the children's self-ratings of emotion during frustration and lateral prefrontal activation such that children who reported mild distress showed greater activation than peers who reported no or high distress highlighting the role of the lateral prefrontal areas and their involvement in emotion regulation. In sum, fNIRS and fMRI neuroimaging studies revealed that activation in prefrontal cortices plus several other brain areas, potentially specific to the exact task demands, are modulated by frustration. Based on this, in our study, we

hypothesize that the lateral prefrontal areas might be indicative of frustration while driving.

To the best of our knowledge, no study exists that investigated the brain activation of drivers experiencing frustration and that uses brain activity measured with fNIRS for automated recognition of frustration. In addition, we are not aware of any study that aimed at continuous, time resolved prediction of driver frustration from facial expression or brain activation measurements. Therefore, the goals of this study were to evaluate whether it is possible to detect spontaneously experienced frustration of drivers based on: (1) video recordings of the face; and (2) brain activation as measured with fNIRS. In addition, we aimed to reveal facial muscle features and cortical brain activation patterns linked to frustration. To this end, we conducted two driving simulator experiments, in which frustration was induced through a combination of time pressure and blocking a goal, while videotaping the faces of the participants (Experiment 1) and recording brain activation using fNIRS (Experiment 2). We employed a multivariate data-driven approach to evaluate whether a discrimination of frustration from a non-frustrated state is possible using the data at hand. This analysis provided us with an estimate for the discriminability of the two induced affective states but could not tell about the underlying patterns of facial and brain activity that are related to frustration. Therefore, additionally, we investigated the underlying facial expressions and brain activity patterns in a second step and report the results thereof.

## MATERIALS AND METHODS

### Experiment 1

#### Participants

Thirty-one volunteers participated in Experiment 1. The video recording of one participant failed due to a technical problem. Consequently, the data of 30 participants (twelve females, age mean [M] = 26.2 years, standard deviation [SD] = 3.5 years) were included in the analysis. All participants held a valid driver's license, gave written informed consent prior to the experiment and received a financial compensation of 21 € for their participation. The experiments of this study were carried out in accordance with the recommendations of the guidelines of the German Aerospace Center and approved by the ethics committee of the Carl von Ossietzky University, Oldenburg, Germany. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

#### Experimental Set-Up

The study was accomplished in a driving simulator consisting of a 46-inch screen (NEC MultiSync LCD4610) with a resolution of 1366 × 768 pixels, a G27 Racing gaming steering wheel (Logitech, Newark, CA, USA) including throttle and brake pedal and a gaming racing seat. Via the steering wheel and the pedals, the participants could control a virtual car in a driving simulation (Virtual Test Drive, Vires Simulationstechnologie, Bad Aibling, Germany). Sounds of the driving simulation were presented via loudspeakers (Logitech Surround Sound Speakers Z506).



During the experiment, the participant's face was filmed using a standard IP-Camera (ABUS, Wetter, Germany) with a resolution of  $1280 \times 720$  pixels at a sampling rate of 10 frames per second.

### Experimental Design and Cover Story

Frustration is experienced when goal-directed behavior is blocked (Lazarus, 1991) and can be intensified by time pressure (e.g., Rendon-Velez et al., 2016). Therefore, a cover story was created that told the participants to imagine being a driver at a parcel delivery service and having to deliver a parcel to a client (goal-directed behavior) within 6 min (time pressure). Participants were told that they received 15 € reimbursement for the experiment plus a bonus of 2 € for every parcel delivered within the given time. The drives took place in a simulated urban environment with two lanes (one per direction). Participants were told to stick to the traffic rules and to not exceed the speed limit of 50 km/h (~31 mph, which is the standard speed limit in urban areas in Germany). The experiment started with a short training with moderate traffic which lasted about 10 min. All participants drove the drives of all conditions (within-subjects block design), which are specified in the following sections.

#### Frust Condition

Three drives were used to induce frustration. In these, the participants had to deliver the parcel in 6 min, but their driving flow was blocked by events on the street. It was tried to distribute the occurrence of the frustrating events roughly equal over time in the scenario. However, the exact timing differed and also depended on participants' speed. In addition, the nature of the events was varied (e.g., red lights, construction sites, slow lead cars that could not be overtaken, or a pedestrian crossing the street slowly) with the goal to create a scenario feeling as natural as possible to the participant. In two Frust drives, the participants were told after 6 min that the parcel could not be delivered in time, they will win no extra money and they should stop the car (of these two drives, the first drive had six frustrating events and the second one had eight). In the third drive participants were told after 5:40 min that they successfully delivered the parcel and won 2 € extra (this drive had eight frustrating events).

#### NoFrust Condition

Three further drives served as control condition. The participants had to deliver the parcel in 6 min with only little or moderate traffic on the ego lane (i.e., the lane they drove on), so that driving at the maximally allowed speed was almost always possible. In two of the drives, the participants were told after a fixed amount of time below 6 min (5:41 and 5:36 min) that they successfully delivered the parcel and won 2 € extra. The third non-frustrated drive ended after 6 min with a message that the time is over and no extra money was won.

The design of the frustrating drives was similar to the experimental manipulations of earlier studies on driver frustration (Lee, 2010; Lee and LaVoie, 2014). The participants drove the experimental conditions in random order and were not informed whether the current drive was a Frust or a NoFrust drive, i.e., they only experienced more or less frequently blocking events during a given drive. In order to reduce carry-over effects between the experimental

drives, the participants had to drive for about 2 min through the same urban setting without any concurrent traffic between two experimental drives. Between the drives, there were breaks, in which participants had to fill in the questionnaires mentioned below and could take some time to relax.

### Subjective Rating

As a manipulation check, the participants rated their subjectively experienced emotion using the SAM (Bradley and Lang, 1994) after each drive. In addition, they filled in the NASA-Task Load Index (NASA-TLX) after each Frust and NoFrust drive (Hart and Staveland, 1988). Here, we specifically focused on the frustration scale. One participant did not fill in the NASA-TLX, so only 29 questionnaires could be analyzed for that scale.

### Data Analyses

#### Subjective Rating

The subjective ratings for the three used questionnaire items were compared to each by means of analysis of variance (ANOVA). Partial eta-squared ( $\eta_p^2$ ) was calculated for each test as an indicator for effect size.

#### Pre-processing of Video Data

The software FACET (Imotions, Copenhagen, Denmark), which is based on the CERT toolbox (Bartlett et al., 2008), was used to extract information regarding the frame-wise AU activity. FACET makes use of the FACS (Ekman et al., 2002) and can determine the activation of 18 AUs as well as head motion (pitch, roll and yaw in °). An overview of the AUs recorded by FACET can be found in **Table 1**. The activation of AUs is coded as evidence, which indicates the likelihood of activation of the respective AU. For instance, an evidence value of 0 means that the software is uncertain about whether or not the respective AU is activated, a positive evidence value refers to an increase in certainty and a negative value to decreasing certainty. In order to reduce inter-individual difference in the evidence value, we subtracted the mean evidence value of the first minute of each drive per AU from the remaining values. In addition, a motion correction was accomplished as FACET operates optimally if the participants' face is located frontally to the camera. Therefore, we analyzed only the frames with a pitch value between  $-10^\circ$  to  $20^\circ$  as well as roll and yaw values between  $-10^\circ$  and  $+10^\circ$ . About 10.6% of the data were removed in this step.

#### Multivariate Cross-Validated Prediction of Frust and NoFrust Drives Based on AU Data

We used a multivariate logistic ridge regression (Hastie et al., 2009) decoding model implemented in the Glmnet toolbox (Qian et al., 2013) for the prediction of Frust and NoFrust drives from the z-scored AU activation (i.e., time resolved evidence). A 10-fold cross-validation approach was used to validate the model. For this, the time series data were split into 10 intervals. This approach avoids overfitting of the data to the model and provides an estimate of how well a decoding approach would predict new data in an online analysis (Reichert et al., 2014). In the logistic ridge regression, the  $\lambda$  parameter (also as hyper-

**TABLE 1** | Overview of recorded action units (AUs).

| AU | Description          |
|----|----------------------|
| 1  | Inner brow raiser    |
| 2  | Outer brow raiser    |
| 4  | Brow lowerer         |
| 5  | Upper lid raiser     |
| 6  | Cheek raiser         |
| 7  | Lid tightener        |
| 9  | Nose wrinkler        |
| 10 | Upper lip raiser     |
| 12 | Lip corner puller    |
| 14 | Dimpler              |
| 15 | Lip corner depressor |
| 17 | Chin raiser          |
| 18 | Lip pucker           |
| 20 | Lip stretcher        |
| 23 | Lip tightener        |
| 24 | Lip pressor          |
| 25 | Lips part            |
| 28 | Lip suck             |

The list and the descriptions of the AUs have been taken from the website of *Imotions* (Farnsworth, 2016).

parameter) determines the overall intensity of regularization. We used a standard nested cross-validation procedure to train the model and test generalization performance. The  $\lambda$  parameter was optimized in an inner 10-fold cross-validation loop implemented in the training. The outer cross-validation loop tested the generalization of the regression model with the optimized  $\lambda$  on the held-out test dataset. The input features that went into the decoding model were the pre-processed AU activations averaged across 10 data frames (= 1 s, no overlapping windows) to reduce the amount of data and increase the signal to noise ratio without increasing the model complexity. The model weights these input features and provides an output which is between 0 and 1. This output value indicates the likelihood for the test data classified as either the Frust class or the NoFrust class. An output  $\geq 0.5$  is considered to be classified as Frust drive whereas an output  $< 0.5$  is considered to be classified as a NoFrust drive.

The accuracy of the classification model of the individual participants was calculated as follows:

Model Accuracy (%) =

$$\frac{TPR_{\text{Frust}} + TPR_{\text{NoFrust}}}{TPR_{\text{Frust}} + TPR_{\text{NoFrust}} + FPR_{\text{Frust}} + FPR_{\text{NoFrust}}} * 100 \quad (1)$$

In Equation 1, the TPR is the true positive rate and FPR is the false positive rate of the two conditions as denoted by Frust or noFrust. The model accuracy by itself is not a sufficient measure for evaluating the robustness of the model. Other performance measures like recall and precision are important indicators to evaluate whether the model exploits group information contained in the data and are insensitive to group size differences (Rieger et al., 2008). Recall is the proportion of trials which belong to a particular empirical class (Frust or NoFrust) and were assigned to the same class by the model. Precision provides information about how precise the model is in assigning the respective class (Frust or NoFrust). In this study, we report the F1-score, which is the harmonic average of precision and

recall. An F1-score of 1 indicates perfect precision and recall (Shalev-Shwartz and Ben-David, 2016). The F1-score for the Frust condition was calculated as follows:

$$F1\text{-score} = \frac{2 * TPR_{\text{Frust}}}{2 * TPR_{\text{Frust}} + FPR_{\text{Frust}} + FPR_{\text{NoFrust}}} \quad (2)$$

### Characterization of Facial Activation Patterns of Frustration: Clustering

A clustering approach was employed to identify patterns of co-activated AUs frequently occurring in the Frust and in the noFrust drives and to compare these between the two conditions. These patterns of co-activated AUs can be seen as the most frequently shown facial expressions during the driving scenarios. For the clustering, we separated the AU data in two sets: a training data set that contained two randomly selected drives from the Frust condition per participant and a test set including the remaining Frust drives and one randomly selected noFrust drive. For the cluster analysis, we used the data of all participants to ensure sufficient sample size. A recent work that simulated the effect of sample size on the quality of the cluster solution recommends using at least 70 times the number of variables considered (Dolcinar et al., 2014). The number of sample points in our training data set was roughly 14 times higher than this recommendation (30 participants  $\times$  2 drives  $\times$  5 min  $\times$  60 s = 18,000 data points  $> 70 \times 18$  AUs = 1,260). K-means clustering with  $k = 5$  was conducted on the training set. A value of  $k = 5$  was chosen after visually inspecting a random selection of video frames of the face recordings. It seemed as if five different expressions were shown predominantly. We applied the resulting cluster centroids to cluster the data from the test set (i.e., each data point was assigned to the cluster with the smallest distance to the centroid). From this, we could determine the percentage of data points per condition assigned to each of the five clusters (per participant) and compare the conditions by means of paired Wilcoxon tests. In addition, we characterized the resulting clusters by their patterns of activated AUs in the centroids. An AU was assumed to be activated if the evidence in the centroid was  $\geq 0.25$ . This criterion was adopted from Grafsgaard et al. (2013), who used the same threshold to select activated AUs in their work. We report the five resulting clusters with the AUs that characterize these as well as the results of the Wilcoxon test. Moreover, we investigated the relationship between the subjectively reported frustration levels (by means of the NASA-TLX frustration item) and the probability of the clusters in the test set. For this, we correlated both values with each other using Kendall's Tau (as the data were not normally distributed). In order to account for the variability between subjects, we additionally performed a linear mixed effects analysis of the relationship between the probability of Cluster 4 and the subjective frustration rating using the combined data of training and test set. With this, we wanted to estimate whether we can predict the probability of showing Cluster 4 using the subjective frustration rating as fixed effect. As random effects we had intercepts for participants and by-participant random slopes for the effect of the subjective rating.  $P$ -values were obtained by likelihood ratio tests ( $\chi^2$ )



of the full models with the effect in question against the models without the effect in question (see Winter, 2013). The models were calculated using the R package lme4 (Bates et al., 2015).

## Experiment 2

### Participants

Sixteen male volunteers aged between 19 and 32 ( $M = 25.3$ ,  $SD = 3.5$ ) years participated in Experiment 2. All participants possessed a valid German driving license and provided written informed consent to participate prior to the experiment. They received a financial compensation of 30 € for the participation in the experiment. The data from one participant was excluded because the participant suffered from simulation sickness during the course of the experiment. Data from three other participants were excluded due to a large number (>50%) of noisy channels in the fNIRS recordings. The mean age of the remaining participants was 25.2 ( $SD = 3.8$ ) years. This study was carried out in accordance with the recommendations of the guidelines of the German Aerospace Center and approved by the ethics committee of the Carl von Ossietzky University, Oldenburg, Germany. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

### Experimental Set-Up

The experiment was accomplished in the virtual reality (VR) lab with 360° full view at the German Aerospace Center (Fischer et al., 2014). Participants sat in a realistic vehicle mock-up and controlled the mock-up car in the driving simulation (Virtual Test Drive, Vires Simulationstechnologie, Bad Aibling, Germany) via a standard interface with throttle, brake pedal, steering wheel and indicators.

### Experimental Design and Cover Story

The same parcel delivery service cover story as in Experiment 1 was used. The only difference was that the participants received a slightly higher basic reimbursement of 18 € (instead of 15 € in Experiment 1) due to the longer overall duration of the experiment. The bonus of 2 € for every parcel delivered within the given time was the same as in Experiment 1. In the end, all participants were paid 30 € for their participation, irrespective of their success. The experiment was structured as a block design and began with a short training of roughly 10 min with moderate traffic. Thereafter, we recorded the baseline data for 2 min following which the participants drove the Frust and noFrust drives (six per condition) in alternation on the same urban track as in Experiment 1. The order of each type of drives was randomized. The experimental conditions are specified in the following sections.

### Frust Condition

In the Frust drives, the participants had to deliver the parcels within a maximum time of 6 min, but their driving was blocked by events on the street (similar to Experiment 1, but with a bit less complexity, e.g., no pedestrians involved).

The blocking events had an average time distance of 20 s (i.e., after 20 s of driving, an obstacle occurred). There were seven blocking events per drive. However, if the participant drove very slowly, it could be that less blocking events were passed. In case they reached the goal within 6 min, a message was presented telling them that they received 2 €. If they did not reach the goal after 6 min, they were informed that they did not succeed this time. Both messages ended the drives accordingly.

### NoFrust Condition

The noFrust drives served as control condition. Participants were told that they had to pick up the parcels from headquarters. There was moderate traffic on the ego lane, so that driving at the maximally allowed speed was almost always possible. The drives took 5 min. Between the drives, there were breaks, in which participants had to fill in the questionnaires mentioned below and could take some time to relax.

### Subjective Rating

As a manipulation check, the participants rated their subjectively experienced emotion using the SAM (Bradley and Lang, 1994) after each drive.

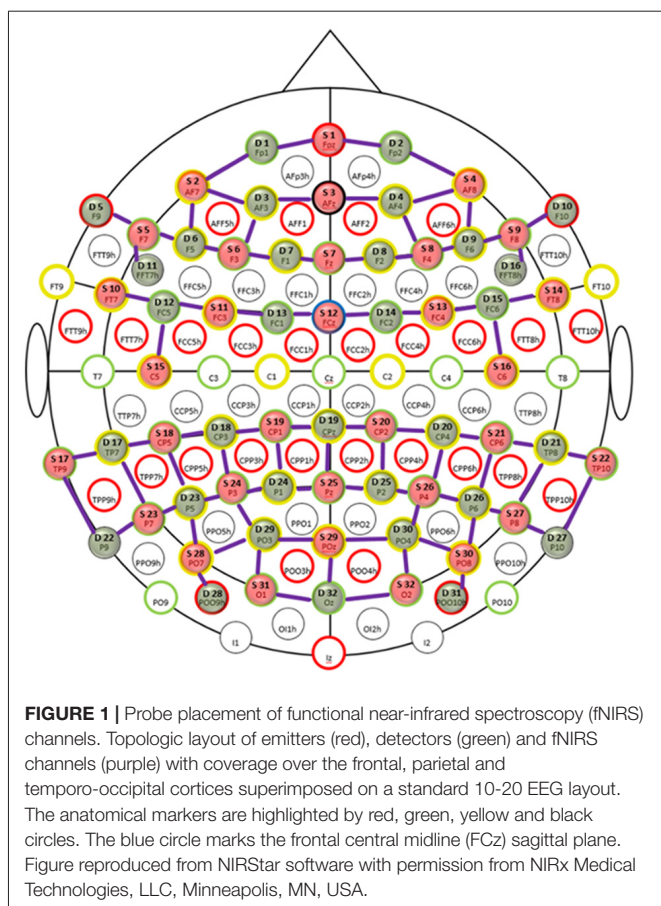
### fNIRS Set-Up

Functional near infrared spectroscopy is a non-invasive optical imaging technique that uses near-infrared light (600–900 nm) to measure hemodynamic responses in the brain (Jöbsis, 1977; Villringer et al., 1993). This is done by measuring the absorption changes in the near-infrared light that reflects the local concentration changes of oxy-hemoglobin (HbO) and deoxy-hemoglobin (HbR) in the sub-cortical brain areas as correlates of functional brain activity. We recorded fNIRS data from the frontal, parietal and temporo-occipital cortices using two NIRScout systems (NIRx Medical Technologies GmbH, Berlin, Germany) in tandem mode resulting in 32 detectors and emitters at two wavelengths (850 and 760 nm). In total, we had 80 channels (combinations of sources and detectors) each for HbO and HbR as shown in **Figure 1**. The distances for the channels ranged between 2 cm and 4 cm ( $M = 3.25$ ,  $SD = 0.45$ ). The shortest channels were the source-detector combinations S5-D11 and S9-D16 in the bilateral prefrontal areas whereas the longest channels were S25-D19 in the parietal midline and S28-D29 and S30-D30 in the bilateral occipital areas (see **Figure 1**). To ensure that the fNIRS cap was placed in a reliable way across all participants, we checked if the position of the optode holder on the fNIRS cap for the anatomical location Cz on the midline sagittal plane of the skull is equidistant to the nasion and the inion and equidistant to the ears. The sampling frequency of the NIRS tandem system was almost 2 Hz.

### Data Analyses

#### Subjective Rating

The subjective ratings for the two questionnaire items were compared to each by means of ANOVA. Partial eta-squared effect sizes ( $\eta_p^2$ ) were calculated for each test.



### fNIRS Data Pre-processing

The raw data from fNIRS measurements record the influence of not only cortical brain activity but also other systemic physiological artifacts (cardiac artifacts, respiration rate, Mayer waves) and movement artifacts causing the signal to be noisy. To reduce the influence of these artifacts, the raw data was pre-processed using the nirsLAB analysis package (Xu et al., 2014). We first computed the coefficient of variation (CV) which is a measure for the signal-to-noise ratio (SNR) from the unfiltered raw data using the mean and the standard deviation of each NIRS channel over the entire duration of the experiment (Schmitz et al., 2005; Schneider et al., 2011). All channels with a CV greater than 20% were excluded from further analysis. Additionally, we performed a visual inspection and deleted channels which were excessively noisy with various spikes. On average, 64 channels each for HbO and HbR per participant and were included in the analysis ( $SD = 7.46$ ). We then applied the modified Beer-Lambert's law to convert the data from voltage ( $\mu V$ ) to relative concentration change (mmol/l; Sassaroli and Fantini, 2004).

To reduce effects of movement artifacts and systemic physiology, we used an autoregressive model of order  $n$  ( $AR(n)$ ;  $n_{\max} = 30$ ) based on autoregressive iteratively reweighted least squares developed by Barker et al. (2013) implemented as a

function in the nirsLAB 2017.6 Toolbox<sup>1</sup>. The algorithm fit the residuals of each individual channel to an  $AR(n)$  model, where  $n$  is the order that minimized the Bayesian information criterion. With the resulting autoregressive coefficients, a pre-whitening filter was generated that was applied to the fNIRS data. The reason for that is the fact that fNIRS time series data are typically characterized by large outliers caused by movement artifacts and serially correlated noise from the physiological artifacts and the temporal correlation of the time samples. This generally leads to incorrect estimation of regressor weights when performing univariate regression analyses and results in overestimation while computing corresponding statistical values, causing an increase in false positives and false negatives (Tachtsidis and Scholkmann, 2016). Pre-whitening can handle such noise correlated time series data where an autoregressive model takes into account the correlation between the current time sample and its neighboring samples and models the temporal correlations.

### Multivariate Cross-Validated Prediction of Frust and NoFrust Drives From fNIRS Data

In line with Experiment 1, we used the multivariate logistic ridge regression (Hastie et al., 2009) decoding model implemented in the Glmnet toolbox (Qian et al., 2013) for the prediction of Frust and NoFrust drives from sample-by-sample fNIRS brain activation data. The input features that went into the decoding model were the pre-processed HbO and HbR values which were z-scored for the particular segments of Frust and noFrust drives. Both HbO and HbR features were used simultaneously. The model weighted these input features and provided an output between 0 and 1. This output value indicates the likelihood for the test data classified as either the Frust class or the NoFrust class. Like in Experiment 1, accuracy and F1 score are reported to estimate the model prediction.

### Characterization of Brain Areas Predictive to Frustration: Univariate Regression Analysis

In order to characterize the pattern of brain areas involved during frustrating drives, we performed univariate regression analyses on a single-subject level separately for each fNIRS channel using the generalized linear model (GLM) analysis module implemented in the nirsLAB Toolbox. Our design matrix consisted of two regressors that corresponded to the entire blocks of Frust and noFrust drives. The autoregressive model  $AR(n)$  from Barker et al. (2013) that generated the pre-whitening filter and was applied to the fNIRS time-series data was also applied to the design matrix. Regression co-efficients were estimated by convolving a boxcar function weighted corresponding to the entire blocks of Frust and noFrust drives with a canonical hemodynamic response function (HRF) implemented in the nirsLAB toolbox (Xu et al., 2014) which composed of one gamma function for HbO. The time parameters at which the response reached the peak and undershoot were 6 s and 16 s, respectively. The canonical HRF was reversed for HbR in order to match the effect-sizes for HbO and HbR brain maps for a particular contrast

<sup>1</sup>[https://www.nitrc.org/projects/fnirs\\_downstate](https://www.nitrc.org/projects/fnirs_downstate)

in the same direction since the HbO and HbR signals are correlated negatively. This setting is applied by default in nirsLAB while estimating the GLM co-efficients for HbR. Channel-wise beta values were used to compute t-statistic for each channel separately for the contrast (difference: Frust-noFrust). Finally, we performed a group level analysis for generalization of the brain areas predictive to frustration while driving. The beta values computed from GLM for each channel and each participant in the individual analysis was used for the group-level analyses. The group-level analyses represented the standard deviation for the beta values for each channel across participants.

## RESULTS

### Experiment 1

#### Subjective Rating

The participants rated the drives from the Frust condition as significantly more arousing, more negative and more frustrating than the drives from the NoFrust condition. The results of the subjective rating are presented in **Table 2**. The frustration rating showed a strong negative correlation with the valence rating ( $r = -0.61, p < 0.001$ ) and a marginal significant positive correlation with the arousal rating ( $r = 0.24, p = 0.06$ ). Valence and arousal were negatively correlated ( $r = -0.46, p < 0.001$ ).

#### Multivariate Prediction of Frust and NoFrust Drives Based on AU Data

The average classification accuracy for the Frust vs. the noFrust condition using the multivariate approach based on the AU activations was 62.0% ( $SD = 9.6\%$ ) and the mean F1 score was 0.617 ( $SD = 0.097$ ). The individual classification results for each participant from the 10-fold cross-validation are presented in **Table 3**. **Figure 2** depicts the results from the multivariate logistic ridge regression model from the participant with the highest classification accuracy of 77.8% (participant 11). The results presented in this figure show the model output for all test data. In **Figure 2**, each orange sample point is the output of the decoding model for the AU test data seen from the Frust drives. Similarly, each green sample point is the output of the model for the AU test data as seen from the noFrust drives. We present our results in the form of TPR (i.e., data from a particular drive is classified as the correct drive) and FPR (i.e., data from a particular drive is classified as the opposite drive). Here, one can see that a TPR of 78.5% and 77.2% and a

FPR of 21.5% and 22.8% were achieved for the Frust and NoFrust drives respectively. For the example participant, an F1-score of 0.78 was achieved.

#### Characterization of Facial Activation Patterns of Frustration: Clustering Approach

The resulting cluster centroids from the k-means clustering, which can be seen as the most frequently shown facial expressions during the drives, are presented in **Figure 3**. When applying the cluster centroids to the (unseen) test set, Cluster 4 and Cluster 5 displayed a significantly different relative frequency of occurrence (i.e., percentage of video frames attributed to this cluster centroid) between the two conditions. Specifically, Cluster 4 was found more often in the Frust condition ( $M = 27.3\%$ ,  $SD = 14.18\%$ ) compared to the noFrust condition ( $M = 10.9\%$ ,  $SD = 12.1\%$ ,  $Z = 3.95, p < 0.001$ ) and Cluster 5 more often in the noFrust ( $M = 33.7\%$ ,  $SD = 19.9\%$ ) than the Frust condition ( $M = 22.7\%$ ,  $SD = 10.9\%$ ,  $Z = -1.87, p < 0.05$ ). No significant differences were observed for the three other clusters (see **Table 4**). Cluster 4 is characterized by above threshold activity (i.e., evidence  $>0.25$ ) in AU9 (nose wrinkler), AU17 (chin raiser), AU18 (lip pucker) and AU24 (lip pressor). In comparison, Cluster 5 accounts for no AU with above threshold evidence. The other clusters are described by different patterns of AU activity: Cluster 1 shows little activity in all AUs (only AU12 [lip corner puller] has evidence  $>0.25$ ), Cluster 2 the highest activation in AU6 (cheek raiser), AU9 (nose wrinkler), AU10 (upper lip raiser) as well as AU12 (lip corner puller) and Cluster 3 mostly in AU4 (brow lowerer), AU9 (nose wrinkler) and AU28 (lip suck; see **Table 5** for an overview and a possible interpretation). Interestingly, the correlation analysis revealed that the frequency of occurrence of the cluster that was shown more often in the Frust condition (Cluster 4) also positively correlated with the subjective frustration rating ( $\tau = 0.27, p < 0.05$ , see **Figure 4**). No other cluster showed a significant relationship with the subjectively experienced frustration (Cluster 1:  $\tau = 0.13, p = 0.15, \chi^2_{(1)} = 3.57, p = 0.06$ ; Cluster 2:  $\tau = 0.02, p = 0.85, \chi^2_{(1)} = 0.75, p = 0.38$ ; Cluster 3:  $\tau = -0.03, p = 0.75, \chi^2_{(1)} = 0.03, p = 0.86$ ; Cluster 5:  $\tau = 0, p = 0.98, \chi^2_{(1)} = 0.03, p = 0.87$ ). The positive relationship between the frustration rating and probability of Cluster 4 was confirmed by results of the linear mixed effects analysis including intercepts for participants and by-participant random slopes as random effects, which revealed a significant relationship between the subjective frustration rating and Cluster 4 probability ( $\chi^2_{(1)} = 6.74, p < 0.01$ ). The analysis of the fixed effect rating

**TABLE 2 |** Means (M), standard deviations (SD) and results of the analysis of variance (ANOVA) for the subjective ratings (self-assessment manikin [SAM] valence, SAM arousal and NASA Task Load Index [NASA-TLX] frustration score).

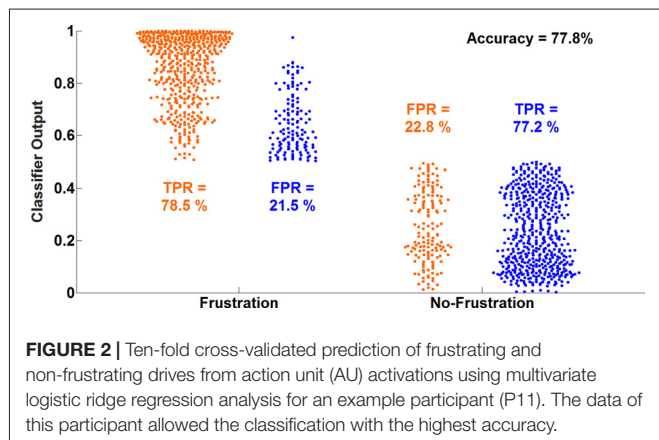
|                      | Frust |     | noFrust |     | ANOVA |         |         |            |
|----------------------|-------|-----|---------|-----|-------|---------|---------|------------|
|                      | M     | SD  | M       | SD  | F     | df      | p       | $\eta^2_p$ |
| SAM arousal          | 4.9   | 1.6 | 3.9     | 1.6 | 14.67 | (1, 29) | <0.01*  | 0.34       |
| SAM valence          | 0.0   | 1.3 | 1.8     | 1.2 | 73.06 | (1, 29) | <0.001* | 0.72       |
| NASA-TLX frustration | 6.3   | 2.2 | 4.4     | 2.0 | 33.70 | (1, 28) | <0.001* | 0.55       |

Note that one participant failed to fill in the NASA-TLX which explains the reduced number of degrees of freedom (df) for that item. Higher values indicate higher arousal (1–9), higher valence (–4 to +4) and higher frustration (1–9). Significant results are marked with a “\*”.



**TABLE 3** | Ten-fold cross-validated predictions of Frust and NoFrust drives from AU data using multivariate logistic ridge regression analysis for all participants.

| Participant       | P1   | P2   | P3   | P4   | P5   | P6   | P7   | P8   | P9   | P10  | P11  | P12  | P13  | P14  | P15  |
|-------------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| Mean accuracy (%) | 63   | 59   | 58   | 55   | 49   | 54   | 78   | 78   | 56   | 60   | 78   | 70   | 70   | 56   | 58   |
| F1-score          | 0.61 | 0.58 | 0.58 | 0.54 | 0.49 | 0.54 | 0.78 | 0.78 | 0.56 | 0.60 | 0.78 | 0.70 | 0.70 | 0.56 | 0.58 |
| Participant       | P16  | P17  | P18  | P19  | P20  | P21  | P22  | P23  | P24  | P25  | P26  | P27  | P28  | P29  | P30  |
| Mean accuracy (%) | 49   | 70   | 73   | 60   | 64   | 53   | 64   | 73   | 60   | 55   | 64   | 51   | 71   | 41   | 71   |
| F1-score          | 0.49 | 0.70 | 0.72 | 0.60 | 0.65 | 0.53 | 0.63 | 0.73 | 0.59 | 0.53 | 0.64 | 0.51 | 0.72 | 0.40 | 0.70 |



revealed that each increase of the subjective frustration rating by 1 on the scale increased the probability of showing Cluster 4 by 1.5% (standard error: 0.5%). Comparing the model with a simpler model without inclusion of random effects revealed that the Akaike information criterion (AIC, Akaike, 1998) was lower for the model with random effects (AIC = -110.2) compared to the one without the random effects (-95.5). This suggests that the model better explains the data if the random effects are included.

## Experiment 2

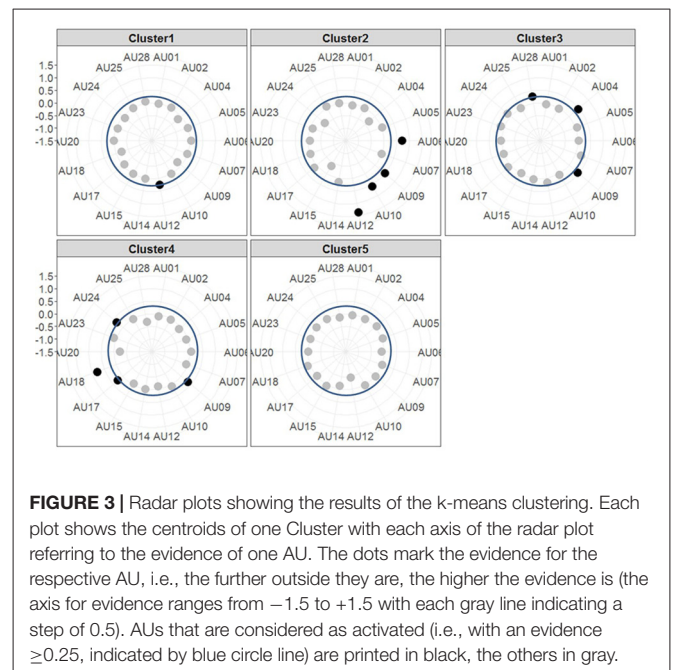
### Subjective Ratings

In concordance with the results of Experiment 1, the participants rated the Frust drives as more arousing (SAM arousal, Frust:  $M = 4.7$ ,  $SD = 1.4$ ; noFrust:  $M = 3.5$ ,  $SD = 1.1$ ,  $F_{(1,11)} = 26.87$ ,  $p < 0.01$ ,  $\eta_p^2 = 0.71$ ) and more negative (SAM valence, Frust:  $M = 0.75$ ,  $SD = 1.2$ ; noFrust:  $M = 1.5$ ,  $SD = 1.0$ ,  $F_{(1,11)} = 15.67$ ,  $p < 0.01$ ,  $\eta_p^2 = 0.59$ ) than the noFrust drives. Valence and arousal were negatively correlated ( $r = 0.55$ ,  $p < 0.01$ ).

### Multivariate Prediction of Frust and NoFrust Drives From fNIRS Data

The mean frustration prediction accuracy and F1-score obtained with fNIRS brain activation recordings across all participants were 78.1% ( $SD = 11.2\%$ ) and 0.776 ( $SD = 0.115$ ), respectively. **Table 6** lists the individual results for all participants.

**Figure 5** shows the distributions of single time interval predictions of the multivariate logistic ridge regression model for the participant with the highest classification accuracy of almost 95% for HbR and HbO data. In this participant, a TPR of 96.5%



and 93.3% and a FPR of 3.5% and 6.7% were achieved for the Frust and NoFrust drives, respectively.

### Characterization of Brain Areas Predictive to Frustration

We performed univariate GLM analyses separately for each channel in order to determine the localization of brain areas most predictive to frustration while driving. The univariate approach was chosen because the model weights of the multivariate fNIRS regression model are hard to interpret for various reasons (Reichert et al., 2014; Weichwald et al., 2015; Holdgraf et al., 2017).

**Figures 6A,B** show the results presented as unthresholded  $t$ -value maps (difference: Frust-noFrust) from the channel-wise linear regression of HbR and HbO data for the group level analysis. The  $t$ -value maps indicate the local effect sizes, in essence they are Cohen's  $d$  scaled by the square root of the number of samples included in their calculation. The  $t$ -values provide a univariate measure to estimate the importance of a feature for multivariate classification. The Bonferroni-corrected  $t$ -maps for the group-level analysis are included in the **Supplementary Figures S1A,B**. In **Figures 6A,B**, both HbR and HbO  $t$ -value maps show significant convergence in brain activation patterns bilaterally in the inferior frontal areas

**TABLE 4 |** Relative frequency of attribution of data points from the test set to the cluster centroids extracted from the test set for the two conditions Frust and noFrust.

|           | Frust    |           | noFrust  |           | Wilcoxon test |          |
|-----------|----------|-----------|----------|-----------|---------------|----------|
|           | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>Z</i>      | <i>p</i> |
| Cluster 1 | 20.6     | 13.4      | 22.6     | 16.9      | −0.32         | 0.63     |
| Cluster 2 | 10.3     | 16.9      | 10.1     | 11.0      | 1.40          | 0.08     |
| Cluster 3 | 19.2     | 14.8      | 22.8     | 12.1      | 0.03          | 0.49     |
| Cluster 4 | 27.3     | 14.8      | 10.9     | 12.1      | 3.95          | < 0.001* |
| Cluster 5 | 22.7     | 10.9      | 33.7     | 19.9      | −1.87         | < 0.05*  |

Means (*M*), standard deviations (*SD*) and results of Wilcoxon test are presented. Significant differences are marked with an asterisk.

**TABLE 5 |** Description of the five clusters including the involved AUs and a potential interpretation of the meaning.

| Cluster | Involved action units | Potential interpretation |
|---------|-----------------------|--------------------------|
| 1       | AU12                  | neutral to slight smile  |
| 2       | AU6, AU9, AU10, AU12  | smiling                  |
| 3       | AU4, AU9, AU28        | frowning                 |
| 4       | AU9, AU17, AU18, AU24 | frustrated               |
| 5       | No above threshold AU | neutral                  |

(putative BA45) and the ventral motor cortex (putative BA6). Additional informative channels can be seen in the right inferior parietal areas (putative BA22) only for the HbR maps but not for the HbO maps. This could be due to the averaging effects of the brain activation on a channel-level across participants who showed inter-individual variabilities. In both HbR and HbO maps, some channels in the left temporo-occipital areas (putative BA21) were found to be predictive to frustrated driving although the trend is not as strong there as it is in the frontal areas. **Figures 6C,D** show *t*-value maps for the same contrast from the channel-wise linear regression of HbR and HbO data for the participant with the highest prediction accuracies. These single participant brain activation patterns closely resemble the pattern of the group level map. However, the *t*-values are much higher in the single participant than in the group averaged map. Both, HbR and HbO signals indicate enhanced activation bilaterally in the inferior frontal and ventral motor areas ( $t > 10$ ) during Frust drives in the single participant *t*-maps whereas the group averaged *t*-values rarely exceed  $t = 4$ . The reduced *t*-values in the averaged maps are due to variability of the predictive brain activation patterns with respect to both, spatial distribution and local effects sizes. This can be seen, for example, in the HbO *t*-statistic value maps which show predictive activation in the left inferior parietal and the left temporo-occipital areas in the single participant maps but less so in the group level analysis.

We visualized the averaged brain map on the MNI 152 brain in Neurosynth<sup>2</sup> and used MRICron<sup>3</sup> to determine MNI co-ordinates and the corresponding Brodmann areas for the brain areas with increased activation differences between Frust and NoFrust drives. **Table 7** lists the brain areas, the MNI-coordinates of the difference maxima and *t*-values as indicators of the effect sizes.

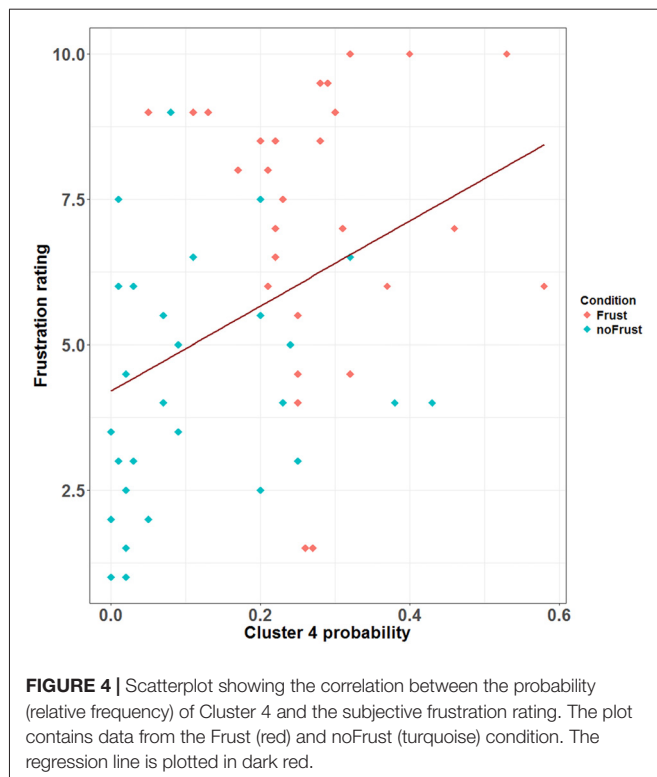
<sup>2</sup><http://neurosynth.org>

<sup>3</sup><https://www.nitrc.org/projects/mricron>

## DISCUSSION

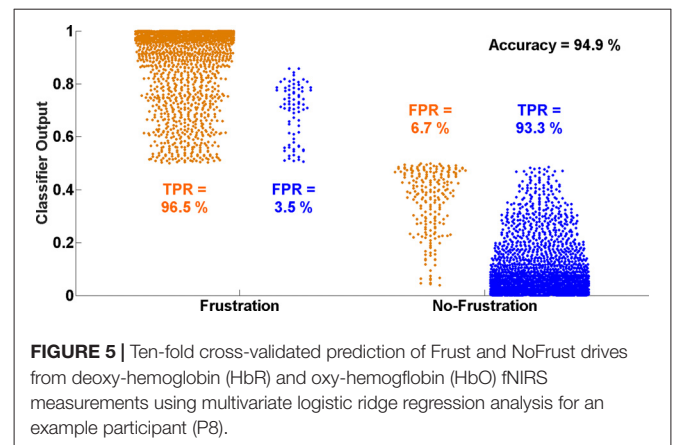
The goals of this study were to investigate discriminative properties of facial muscle activity extracted from video recordings and brain activation patterns using fNIRS for the automated detection of driver frustration. Therefore, two driving simulator experiments were conducted in which frustration was induced through a combination of time pressure and goal blocking. In Experiment 1, we videotaped the faces of the participants during the drives and extracted the activity of the facial muscles using automated video processing. We could show that the facial expression data can be used to classify frustration from a neutral state with an average classification accuracy of almost 62%. Frustration could be discriminated from a neutral state with above chance accuracy in most participants, with maximum accuracy up to 78% for the best participants. In addition, a detailed analysis comparing the muscle activation in both conditions revealed that the muscles nose wrinkler, chin raiser, lip pucker and lip pressor are activated in synchrony more often in the frustrating condition than in the neutral condition. The approach was then extended to fNIRS brain activation measurements in Experiment 2, where the discrimination of frustration from the neutral state improved to average classification accuracy of almost 78% and up to 95% for the best two participants. An additional univariate GLM analysis indicated that frustration during driving was reflected in reliable brain activation modulation bilaterally in ventrolateral inferior frontal areas in the group-level analysis. Our results demonstrate that frustration during driving could be detected time resolved from video recordings of the face and fNIRS recordings of the brain.

In both experiments, frustration was induced using a combination of events blocking goal-directed behavior and time pressure during simulated drives. According to research on frustration as well as previous studies on frustrated driving, this combination generally leads to a state of experienced frustration for the participants (Lazarus, 1991; Lee, 2010; Rendon-Velez et al., 2016). The accomplished manipulation checks showed that the participants rated the frustrating drives as more negative and more arousing than the non-frustrating drives in both experiments, which is in line with the classification of frustration in the valence-arousal space of emotions (Russell, 1980). Additionally, the participants assigned a higher score in the NASA-TLX frustration scale to the frustrating drives in Experiment 1. Therefore, we could conclude that the experimental manipulation indeed induced



frustration and was suitable to study the proposed research questions.

Our approach of using multivariate logistic ridge regression in combination with cross-validation enabled us to explore the feasibility of using facial muscle activity extracted from video recordings of the face and almost whole-head fNIRS as an estimate for cortical activity for time-resolved characterization of driver frustration. The multivariate modeling allowed us to predict frustrated drives from the non-frustrated drives in a continuous manner with relatively high accuracy. For the facial muscle activation, the decoding model made predictions from the evidence values of 18 AU input features which were pre-selected with the used software. For the fNIRS brain activation, the input features to the decoding model were the sample-by-sample pre-processed fNIRS data from all the selected channels for each participant. On average, we had about 128 input features ( $SD = 14.9$ ) across all participants. Our decoding models were able to discriminate driver frustration from non-frustration with a mean accuracy of 62% for facial muscle activation and almost 80% for cortical activation. The cross-validation approach allowed us to estimate the generalization of our decoding model to new data which our model had never seen before (i.e., the test dataset) indicating



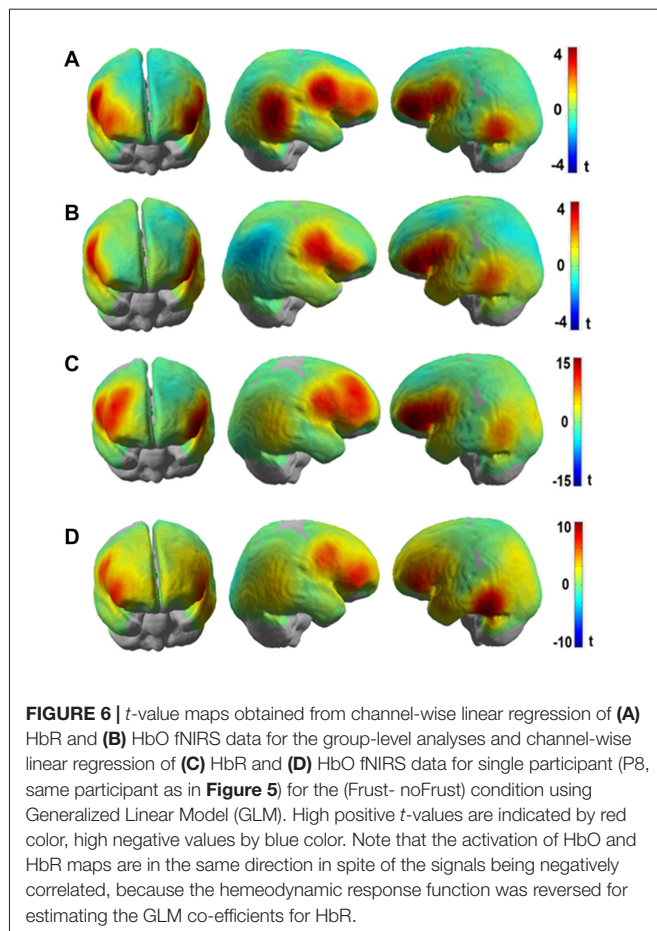
the true predictive power of our model necessary for online tracking of user states (Reichert et al., 2014; Holdgraf et al., 2017). The classification accuracy derived from the facial expression data is higher than chance level, but likely not high enough for robust usage in human-machine systems with adaptive automation. One reason for that may be the fact that humans do not show the same facial expression constantly over a period of several minutes, even though they report to be frustrated in that drive. Moreover, it is conceivable that the level of frustration also varied during the drives leading to fact that facial expression indicating other emotions or a neutral state may have been shown by participants. Together, this may have biased the training and test set as these not only included facial expressions of frustration, but also other facial expressions. This in turn could lead to the lower classification accuracy for facial expression data in comparison to the brain activation. Still, we can confirm our initial hypothesis that it is possible to discriminate driver frustration from a neutral affective state using facial muscle activity and cortical activation with above chance accuracy with cortical activation providing better classification results. It remains to be shown that the classification accuracy is high enough to ensure user acceptance in adaptive automation.

Since the supervised classification gives us only an estimate of how well we would be able to recognize frustration using the respective data frames, we also conducted a detailed analysis of the facial muscle and brain activation data to understand which features are indicative for frustration. For this, a clustering of the facial muscle activity was conducted in order to identify patterns of co-activated facial muscles that occur with increased likelihood if a driver is frustrated. The clustering approach revealed five different clusters of AU activity, which can be seen as the facial expressions that were shown (most frequently) during the drives. Cluster 4 was shown significantly more often

**TABLE 6 |** Ten-fold cross-validated predictions of Frustrated and No-Frustrated drives from fNIRS measurements using multivariate logistic ridge regression analysis for all participants.

| Participant       | P1   | P2   | P3   | P4   | P5   | P6   | P7   | P8   | P9   | P10  | P11  | P12  |
|-------------------|------|------|------|------|------|------|------|------|------|------|------|------|
| Mean accuracy (%) | 76   | 75   | 73   | 88   | 65   | 95   | 78   | 95   | 64   | 76   | 88   | 64   |
| F1-score          | 0.76 | 0.74 | 0.73 | 0.88 | 0.64 | 0.95 | 0.77 | 0.95 | 0.63 | 0.76 | 0.87 | 0.63 |





in the Frustr than in the noFrustr condition and its probability additionally correlated with the subjective frustration rating. Therefore, this pattern subsuming activity from muscles from the mouth region (chin raiser, lip pucker and lip pressor) and

**TABLE 7 |** Brain areas showing increased activation in the Frustr compared to the noFrustr condition.

| Brain areas             | Putative Brodmann Area (BA) | X   | Y   | Z  | <i>t</i> -value |
|-------------------------|-----------------------------|-----|-----|----|-----------------|
| <b>For HbR</b>          |                             |     |     |    |                 |
| Right inferior frontal  | 45                          | 46  | 48  | 24 | 2.0             |
| Left inferior frontal   | 45                          | -49 | 41  | 24 | 3.6*            |
| Right ventral motor     | 6                           | 62  | 7   | 36 | 3.5*            |
| Left ventral motor      | 6                           | -58 | 9   | 37 | 3.6*            |
| Right inferior parietal | 22                          | 63  | -46 | 11 | 4.2*            |
| Left temporo-occipital  | 21                          | -63 | -51 | 1  | 3.4             |
| <b>For HbO</b>          |                             |     |     |    |                 |
| Right inferior frontal  | 45                          | 51  | 31  | 16 | 2.2             |
| Left inferior frontal   | 45                          | -53 | 35  | 20 | 4.1*            |
| Right ventral motor     | 6                           | 61  | 10  | 29 | 3.0             |
| Left ventral motor      | 6                           | -59 | 10  | 36 | 3.4             |
| Left temporo-occipital  | 21                          | -66 | -47 | 7  | 2.3             |

The MNI coordinates of activation and their *t*-values are shown. \*Indicates statistically significant differences which survived the Bonferroni-corrected thresholding of  $p < 0.05$ .

traces from the nose wrinkle can likely be seen as comprising the frustrated facial expression. Interestingly, similar patterns of AU activation have been associated to frustration in previous research (D'Mello et al., 2005; Grafsgaard et al., 2013; Ihme et al., in press). In contrast, the Cluster 5 was activated more often in the non-frustrating drives by the participants. Because it also has no activated AUs involved, we consider it as referring to a neutral facial expression. None of the remaining clusters differed in frequency of occurrence between the two conditions. Presumably, Cluster 1 can also be seen as neutral, because it included only little facial muscle activity. With highest activation in cheek raiser and lip corner puller (and some activation in the nose wrinkler and the upper lip raiser), Cluster 2 likely represents a smiling face (Ekman and Friesen, 2003; Ekman, 2004; Hoque et al., 2012). Finally, Cluster 3 showed a pattern with high activity in action units around the eyes (brow lowerer and nose wrinkler), which could be a frowning as a sign of anger or concentration (Ekman and Friesen, 2003). One interesting issue is that the nose wrinkler (AU 9) occurred frequently and, according to our analysis, is part of Cluster 2 (smiling), 3 (frowning) and 4 (frustration), although most previous research has associated it predominantly with disgust (e.g., Ekman et al., 1980; Lucey et al., 2010), which was likely not induced through our experimental paradigm and set-up. We speculate that two aspects may explain this frequent occurrence of AU9. First, it could be that the software which we used misclassified movements of the eyebrows and attributed these to the nose wrinkler. This is possible and poses a disadvantage of automated techniques to extract facial muscle activity compared to manual coding approaches. Second, it could be that the nose wrinkler is not a particular sign of disgust, but rather a sign of one factor of a dimensional model of emotions. For example, Boukricha et al. (2009) have shown a correlation between AU9 and low pleasure as well as high dominance. We would like to stress here that although the frustrated facial expression (Cluster 4) occurred most often in the frustrated drives, the results indicate that it was not the only facial expression that has been shown by the participants (as already speculated above). Therefore, the approach to cluster time-resolved AU activations into patterns of co-activation in order to gain information about the shown facial expressions appears promising to better understand which facial expressions are shown by the drivers when they experience frustration or other emotions. Future studies should evaluate whether the results from the clustering can be utilized to generate labels that not only indicate the emotion induction phase from which a sample stems, but the facial expression that was actually shown by the participant. This could improve the training data set for the classification as well as classification accuracy. To sum up, the detailed cluster analysis revealed that the facial expression of frustration is mainly linked to the facial muscle activity in the mouth region.

To investigate the frustration predictive features from the fNIRS brain recordings, we performed univariate regression analyses separately for each channel using GLM to determine the localization of brain areas most predictive to frustration

while driving. Our group level results indicate that fNIRS brain activation patterns of frustrated drivers were clearly discernible from non-frustrated drivers. Frustration during driving was reflected with stronger HbR and HbO activation bilaterally in the inferior frontal areas (putative BA 45) and the ventral motor cortex (putative BA 6) in the group level analysis. The fNIRS channels close to the right inferior parietal areas (putative BA 22) also show increased activation to frustrated driving in the HbR t-value maps. Additionally, both HbR and HbO t-value maps show some channels in the left temporo-occipital areas (putative BA 21) to be predictive for frustrated driving although the average linear trend is not as strong there as it is in the frontal areas. Overall, fNIRS revealed brain areas displaying higher activity in the frustrating drives which are in line with the literature on frustration-related neuroimaging lab studies. These areas have been reported to be related with cognitive appraisal, impulse control and emotion regulation processes. Previous research has shown the lateral frontal cortices as a neural correlate for frustration (Siegrist et al., 2005; Hirshfield et al., 2014; Perlman et al., 2014; Yu et al., 2014; Bierzyńska et al., 2016). BA 45 and BA 6 are thought to play an important role in modulating emotional response (Olejarczyk, 2007), regulation of negative affect (Ochsner and Gross, 2005; Phillips et al., 2008; Erk et al., 2010), processing emotions (Deppe et al., 2005) and inhibition control (Rubia et al., 2003). BA 22 has been shown to play a crucial role in attributing intention to others (Brunet et al., 2000), and in social perception e.g., processing of non-verbal cues to assess mental states of others (Jou et al., 2010).

The current study has a few limitations that need to be mentioned. First of all, for obtaining the multivariate predictions, the entire Frust condition had been labeled as “frustrated,” while the complete noFrust condition had been labeled as “non-frustrated.” However, it is very likely that the subjectively experienced level of frustration was not constant across the entire drives, because blocking events can temporally increase the level of frustration that also could build up over time (for instance with increasing number of blocking events). We have not considered these factors for our analysis in order to reduce the complexity. In future studies, a more fine-grained analysis of the current frustration level and its development over time could improve the ground truth where the decoding model could discriminate the different levels of frustration (similar to what Unni et al., 2017 achieved for working memory load). Second, stressful cognitive tasks as in the case of frustrated driving may elicit task-related changes in the physiological parameters such as heart rate, respiration rate, blood pressure etc. (Tachtsidis and Scholkmann, 2016). These global components represent a source of noise in the fNIRS data. There are different approaches to monitor these parameters and use them as additional regressors in the GLM e.g., using short-separation fNIRS channels to capture the effects of these physiological signals (Saager and Berger, 2005) or using principle component spatial filtering to separate the global and local components in fNIRS (Zhang et al., 2016). These approaches have been reviewed by Scholkmann et al. (2014). In our study, for the fNIRS analysis, we did not separate the influence of these

global components from the intracerebral neural components. However, the localized predictive activation we find renders it unlikely that global physiological effects contribute significantly to our results.

Third, due to the study design with two different participant cohorts, we could not combine the decoding models from the two experiments into one single prediction model. We separated the two experiments because we wanted to have a free view on participants' face, which is not covered (partly) by the fNIRS cap. Since the results revealed that facial expression of frustration primarily includes activity in the mouth region, we assume that a combination of both measures is feasible, so that future studies should investigate the potential for frustration detection using a combination of facial expressions and brain activity.

Another minor limitation is that we did not use the same subjective questionnaires in the two experiments, so we did not explicitly ask the participants to report the frustration level in the second experiment. Still, the valence and frustration ratings in the first experiment were highly correlated. Moreover, the valence and arousal ratings in were comparable in both experiments and in line with the classification of frustration according to dimensional theories of emotion (Russell, 1980), so that a successful induction of frustration in both experiments is likely.

## CONCLUSION AND OUTLOOK

This study demonstrated the potential of video recordings from the face and whole head fNIRS brain activation measurements for the automated recognition of driver frustration. Although the results of this study are relatively promising, future research is needed to further validate the revealed facial muscle and brain activation patterns. In addition, a combination of both measures (potentially even together with further informative parameters such as peripheral physiology) appears auspicious for improving our models of driver frustration thereby boosting the classification accuracy. The availability of wireless and portable fNIRS devices could make it possible to assess driver frustration *in situ* in real driving in the future. Overall, our results pave the way for an automated recognition of driver frustration for usage in adaptive systems for increasing traffic safety and comfort.

## AUTHOR CONTRIBUTIONS

KI, AU, JR and MJ planned the research. Data collection was done by KI and AU. Data analysis was carried out by KI, AU, MZ and JR. KI, AU, MZ, JR and MJ prepared the manuscript. KI and AU contributed equally.

## FUNDING

This work was supported by the funding initiative Niedersächsisches Vorab of the Volkswagen Foundation and the Niedersächsische Ministerium für Wissenschaft und

Kultur (Ministry of Science and Culture of Lower Saxony) as a part of the Interdisciplinary Research Centre on Critical Systems Engineering for Socio-Technical Systems and a Deutsche Forschungsgemeinschaft (DFG)-grant RI1511/2-1 to JR.

## ACKNOWLEDGMENTS

We thank Andrew Koerner, Dirk Assmann and Henrik Surm for their technical assistance as well as Christina Dömeland and Helena Schmidt for helping out with the data collection.

## REFERENCES

- Abler, B., Walter, H., and Erk, S. (2005). Neural correlates of frustration. *Neuroreport* 16, 669–672. doi: 10.1097/00001756-200505120-00003
- Akaike, H. (1998). “Information theory and an extension of the maximum likelihood principle,” in *Selected Papers of Hirotugu Akaike. Springer Series in Statistics (Perspectives in Statistics)*, eds E. Parzen, K. Tanabe and G. Kitagawa (New York, NY: Springer), 199–213.
- Barker, J. W., Aarabi, A., and Huppert, T. J. (2013). Autoregressive model based algorithm for correcting motion and serially correlated errors in fNIRS. *Biomed. Opt. Express* 4, 1366–1379. doi: 10.1364/BOE.4.001366
- Bartlett, M., Littlewort, G., Wu, T., and Movellan, J. (2008). “Computer expression recognition toolbox,” in *Proceedings of the 8th IEEE International Conference on Automatic Face and Gesture Recognition (FG)* (Amsterdam), 1–2.
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Bierzynska, M., Bielecki, M., Marchewka, A., Debowska, W., Duszyk, A., Zajkowski, W., et al. (2016). Effect of frustration on brain activation pattern in subjects with different temperament. *Front. Psychol.* 6:669. doi: 10.3389/fpsyg.2015.01989
- Boukricha, H., Wachsmuth, I., Hofstatter, A., and Grammer, K. (2009). “Pleasure-arousal-dominance driven facial expression simulation,” in *Proceedings of the 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops (ACII 2009)* (Amsterdam), 1–7.
- Bradley, M. M., and Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *J. Behav. Ther. Exp. Psychiatry* 25, 49–59. doi: 10.1016/0005-7916(94)90063-9
- Brass, M., and Haggard, P. (2007). To do or not to do: the neural signature of self-control. *J. Neurosci.* 27, 9141–9145. doi: 10.1523/JNEUROSCI.0924-07.2007
- Bruce, V. (1992). “What the human face tells the human mind: some challenges for the robot-human interface,” in *Proceedings IEEE International Workshop on Robot and Human Communication* (Tokyo: IEEE Press), 44–51.
- Brunet, E., Sarfati, Y., Hardy-Baylé, M.-C., and Decety, J. (2000). A PET investigation of the attribution of intentions with a nonverbal task. *Neuroimage* 11, 157–166. doi: 10.1006/nimg.1999.0525
- Deppe, M., Schwindt, W., Kugel, H., Plassmann, H., and Kenning, P. (2005). Nonlinear responses within the medial prefrontal cortex reveal when specific implicit information influences economic decision making. *J. Neuroimage* 15, 171–182. doi: 10.1177/1051228405275074
- D’Mello, S., Craig, S., Gholson, B., Franklin, S., Picard, R., and Graesser, A. (2005). “Integrating affect sensors in an intelligent tutoring system,” in *Affective Interactions: The Computer in the Affective Loop Workshop at 2005 International Conference on Intelligent User Interfaces* (New York, NY: AMC Press), 7–13.
- Dolcinar, S., Grün, B., Leisch, F., and Schmidt, K. (2014). Required sample sizes for data-driven market segmentation analyses in tourism. *J. Travel Res.* 53, 296–306. doi: 10.1177/0047287513496475
- Ekman, P. (2004). *Emotions Revealed: Understanding Faces and Feelings*. London: Phoenix.
- Ekman, P., and Friesen, W. V. (2003). *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*. Cambridge, MA: Malor Books.
- Ekman, P., Friesen, W. V., and Ancoli, S. (1980). Facial signs of emotional experience. *J. Pers. Soc. Psychol.* 39, 1125–1134. doi: 10.1037/h0077722

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2018.00327/full#supplementary-material>

**FIGURE S1 |** Bonferroni-corrected *t*-value maps obtained from channel-wise linear regression of (A) HbR and (B) HbO fNIRS data for the group-level analyses for the (Frustr-noFrustr) condition using Generalized Linear Model (GLM). The red and yellow dots represent the sources and detectors, respectively.

- Ekman, P., Friesen, W. V., and Hager, J. (2002). *Investigator’s Guide to the Facial Action Coding System (FACS)*. Salt Lake City, UT: A Human Face.
- Ekman, P. (1993). Facial expression and emotion. *Am. Psychol.* 48, 376–379. doi: 10.1037/0003-066X.48.4.384
- Erickson, K., and Schulkin, J. (2003). Facial expressions of emotion: a cognitive neuroscience perspective. *Brain Cogn.* 52, 52–60. doi: 10.1016/s0278-2626(03)00008-3
- Erk, S., Mikschl, A., Stier, S., Ciaramidaro, A., Gapp, V., Weber, B., et al. (2010). Acute and sustained effects of cognitive emotion regulation in major depression. *J. Neurosci.* 30, 15726–15734. doi: 10.1523/JNEUROSCI.1856-10.2010
- European Commission. (2011). *White Paper ‘Roadmap to a Single European Transport Area—Towards a Competitive and Resource-Efficient Transport System’*. Brussels: COM(2011) 144 final.
- Farnsworth, B. (2016). *Facial Action Coding System (FACS)—A Visual Guidebook*. Available online at: <https://imotions.com/blog/facial-action-coding-system/>. Accessed on July 2, 2018.
- Fischer, M., Richter, A., Schindler, J., Plättner, J., Temme, G., Kelsch, J., et al. (2014). “Modular and scalable driving simulator hardware and software for the development of future driver assistance and automation systems,” in *Proceedings of the New Developments in Driving Simulation Design and Experiments: Driving Simulation Conference* (Paris), 223–229.
- Gao, H., Yu, A., and Thiran, J.-P. (2014). “Detecting emotional stress from facial expressions for driving safety,” in *Proceedings of the IEEE International Conference on Image Processing (ICIP)* (Paris), 5961–5965.
- Grabell, A. S., Li, Y., Barker, J. W., Wakschlag, L. S., Huppert, T. J., and Perlman, S. B. (2018). Evidence of non-linear associations between frustration-related prefrontal cortex activation and the normal: abnormal spectrum of irritability in young children. *J. Abnorm. Child Psychol.* 46, 137–147. doi: 10.1007/s10802-017-0286-5
- Grafsgaard, J. F., Wiggins, J. B., Boyer, K. E., Wiebe, E. N., and Lester, J. C. (2013). “Automatically recognizing facial indicators of frustration: a learning-centric analysis,” in *Proceedings of the Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII)* (Geneva: IEEE Press), 159–165.
- Hamm, J., Kohler, C. G., Gur, R. C., and Verma, R. (2011). Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders. *J. Neurosci. Methods* 200, 237–256. doi: 10.1016/j.jneumeth.2011.06.023
- Hart, S. G., and Staveland, L. (1988). “Development of the NASA-TLX (Task Load Index): results of empirical and theoretical research,” in *Human Mental Workload*, eds P. A. Hancock and N. Meshkati (Amsterdam: North Holland Press), 139–183.
- Hastie, T., Tibshirani, R., and Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd Edn. New York, NY: Springer.
- Hirshfield, L. M., Bobko, P., Barelka, A., Hirshfield, S. H., Farrington, M. T., Gulbronson, S., et al. (2014). Using noninvasive brain measurement to explore the psychological effects of computer malfunctions on users during human-computer interactions. *Adv. Hum. Comput. Interact.* 2014, 1–13. doi: 10.1155/2014/101038
- Holdgraf, C. R., Rieger, J. W., Micheli, C., Martin, S., Knight, R. T., and Theunissen, F. E. (2017). Encoding and decoding models in cognitive electrophysiology. *Front. Syst. Neurosci.* 11:61. doi: 10.3389/fnsys.2017.00061



- Hoque, M. E., McDuff, D. J., and Picard, R. W. (2012). Exploring temporal patterns in classifying frustrated and delighted smiles. *IEEE Trans. Affective Comput.* 3, 323–334. doi: 10.1109/t-affc.2012.11
- Ihme, K., Dömeland, C., Freese, M., and Jipp, M. (in press). Frustration in the face of the driver: a simulator study on facial muscle activity during frustrated driving. *Interaction Studies*
- Jeon, M. (2012). “A systematic approach to using music for mitigating affective effects on driving performance and safety,” in *Proceedings of the 14th ACM Conference on Ubiquitous Computing* (New York, NY: ACM Press), 1127–1132.
- Jeon, M. (2015). Towards affect-integrated driving behaviour research. *Theor. Issues Ergon. Sci.* 16, 553–585. doi: 10.1080/1463922x.2015.1067934
- Jöbsis, F. F. (1977). Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science* 198, 1264–1267. doi: 10.1126/science.929199
- Jou, R. J., Minshew, N. J., Keshavan, M. S., Vitale, M. P., and Hardan, A. Y. (2010). Enlarged right superior temporal gyrus in children and adolescents with autism. *Brain Res.* 1360, 205–212. doi: 10.1016/j.brainres.2010.09.005
- Lazarus, R. S. (1991). Progress on a cognitive-motivational-relational theory of emotion. *Am. Psychol.* 46, 819–834. doi: 10.1037//0003-066x.46.8.819
- Lee, Y.-C. (2010). Measuring drivers' frustration in a driving simulator. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 54, 1531–1535. doi: 10.1177/154193121005401937
- Lee, Y.-C., and LaVoie, N. (2014). Relationship between frustration justification and vehicle control behaviors—A simulator study. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 58, 2235–2239. doi: 10.1177/1541931214581429
- Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z., and Matthews, I. (2010). “The extended cohn-kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)* (San Francisco, CA), 94–101.
- Malta, L., Miyajima, C., Kitaoka, N., and Takeda, K. (2011). Analysis of real-world driver's frustration. *IEEE Trans. Intell. Transport. Syst.* 12, 109–118. doi: 10.1109/tits.2010.2070839
- Ochsner, K., and Gross, J. (2005). The cognitive control of emotion. *Trends Cogn. Sci.* 9, 242–249. doi: 10.1016/j.tics.2005.03.010
- Olejarczyk, E. (2007). “Application of fractal dimension method of functional MRI time-series to limbic dysregulation in anxiety study,” in *Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Lyon), 3408–3410.
- Perlman, S. B., Luna, B., Hein, T. C., and Huppert, T. J. (2014). fNIRS evidence of prefrontal regulation of frustration in early childhood. *Neuroimage* 85, 326–334. doi: 10.1016/j.neuroimage.2013.04.057
- Phillips, M. L., Ladouceur, C. D., and Drevets, W. C. (2008). A neural model of voluntary and automatic emotion regulation: implications for understanding the pathophysiology and neurodevelopment of bipolar disorder. *Mol. Psychiatry* 13, 833–857. doi: 10.1038/mp.2008.65
- Picard, R. W., and Klein, J. (2002). Computers that recognise and respond to user emotion: theoretical and practical implications. *Interact. Comput.* 14, 141–169. doi: 10.1016/s0953-5438(01)00055-8
- Qian, J., Hastie, T., Friedman, J., Tibshirani, R., and Simon, N. (2013). *Glmnet for Matlab*. Available online at: [http://www.stanford.edu/~hastie/glmnet\\_matlab/](http://www.stanford.edu/~hastie/glmnet_matlab/)
- Reichert, C., Fendrich, R., Bernarding, J., Tempelmann, C., Hinrichs, H., and Rieger, J. W. (2014). Online tracking of the contents of conscious perception using real-time fMRI. *Front. Neurosci.* 8:116. doi: 10.3389/fnins.2014.00116
- Rendon-Velez, E., van Leeuwen, P. M., Happee, R., Horváth, I., van der Vegte, W., and de Winter, J. (2016). The effects of time pressure on driver performance and physiological activity: a driving simulator study. *Transp. Res. Part F Traffic Psychol. Behav.* 41, 150–169. doi: 10.1016/j.trf.2016.06.013
- Rieger, J. W., Reichert, C., Gegenfurtner, K. R., Noesselt, T., Braun, C., Heinze, H.-J., et al. (2008). Predicting the recognition of natural scenes from single trial MEG recordings of brain activity. *Neuroimage* 42, 1056–1068. doi: 10.1016/j.neuroimage.2008.06.014
- Rubia, K., Smith, A. B., Brammer, M. J., and Taylor, E. (2003). Right inferior prefrontal cortex mediates response inhibition while mesial prefrontal cortex is responsible for error detection. *Neuroimage* 20, 351–358. doi: 10.1016/s1053-8119(03)00275-1
- Russell, J. A. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol.* 39, 1161–1178. doi: 10.1037/h0077714
- Saager, R. B., and Berger, A. J. (2005). Direct characterization and removal of interfering absorption trends in two-layer turbid media. *J. Opt. Soc. Am. A* 22, 1874–1882. doi: 10.1364/josaa.22.001874
- Sassaroli, A., and Fantini, S. (2004). Comment on the modified Beer-Lambert law for scattering media. *Phys. Med. Biol.* 49, N255–N257. doi: 10.1088/0031-9155/49/14/n07
- Scherer, K. R. (2005). What are emotions? And how can they be measured? *Soc. Sci. Inf.* 44, 695–729. doi: 10.1177/0539018405058216
- Schmitz, C. H., Klemer, D. P., Hardin, R., Katz, M. S., Pei, Y., Graber, H. L., et al. (2005). Design and implementation of dynamic near-infrared optical tomographic imaging instrumentation for simultaneous dual-breast measurements. *Appl. Opt.* 44, 2140–2153. doi: 10.1364/ao.44.002140
- Schneider, P., Piper, S., Schmitz, C., Schreiter, N., Volkwein, N., Lüdemann, L., et al. (2011). Fast 3D near-infrared breast imaging using indocyanine green for detection and characterization of breast lesions. *Rofo* 183, 956–963. doi: 10.1055/s-0031-1281726
- Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Mata Pavia, J., Wolf, U., et al. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *Neuroimage* 85, 6–27. doi: 10.1016/j.neuroimage.2013.05.004
- Shalev-Shwartz, S., and Ben-David, S. (2016). *Understanding Machine Learning: From Theory to Algorithms*. New York, NY: Cambridge University Press.
- Siegrist, J., Menrath, I., Stöcker, T., Klein, M., Kellermann, T., Shah, N. J., et al. (2005). Differential brain activation according to chronic social reward frustration. *Neuroreport* 16, 1899–1903. doi: 10.1097/01.WNR.0000186601.50996.f7
- Tachtsidis, I., and Scholkmann, F. (2016). False positives and false negatives in functional near-infrared spectroscopy: issues, challenges, and the way forward. *Neurophotonics* 3:31405. doi: 10.1117/1.NPh.3.3.031405
- Unni, A., Ihme, K., Jipp, M., and Rieger, J. W. (2017). Assessing the driver's current level of working memory load with high density functional near-infrared spectroscopy: a realistic driving simulator study. *Front. Hum. Neurosci.* 11:167. doi: 10.3389/fnhum.2017.00167
- Villringer, A., Planck, J., Hock, C., Schleinkofer, L., and Dirnagl, U. (1993). Near infrared spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults. *Neurosci. Lett.* 154, 101–104. doi: 10.1016/0304-3940(93)90181-j
- Weichwald, S., Meyer, T., Özdenizci, O., Schölkopf, B., Ball, T., and Grosse-Wentrup, M. (2015). Causal interpretation rules for encoding and decoding models in neuroimaging. *Neuroimage* 110, 48–59. doi: 10.1016/j.neuroimage.2015.01.036
- Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. *arXiv:1308.5499* [Preprint].
- Xu, Y., Graber, H. L., and Barbour, R. L. (2014). “nirsLAB: a computing environment for fNIRS neuroimaging data analysis,” in *Poster BM3A.1 at Biomedical Optics* (Miami, FL).
- Yu, R., Mobbs, D., Seymour, B., Rowe, J. B., and Calder, A. J. (2014). The neural signature of escalating frustration in humans. *Cortex* 54, 165–178. doi: 10.1016/j.cortex.2014.02.013
- Zhang, X., Noah, J. A., and Hirsch, J. (2016). Separation of the global and local components in functional near-infrared spectroscopy signals using principal component spatial filtering. *Neurophotonics* 3:15004. doi: 10.1117/1.NPH.3.1.015004

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Ihme, Unni, Zhang, Rieger and Jipp. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Using Smartbands, Pupillometry and Body Motion to Detect Discomfort in Automated Driving

Matthias Beggiato\*, Franziska Hartwich and Josef Krems

Department of Psychology, Cognitive and Engineering Psychology, Chemnitz University of Technology, Chemnitz, Germany

## OPEN ACCESS

### Edited by:

Karel Brookhuis,  
University of Groningen, Netherlands

### Reviewed by:

Dick De Waard,  
University of Groningen, Netherlands  
Benjamin Cowley,  
University of Helsinki, Finland

### \*Correspondence:

Matthias Beggiato  
matthias.beggiato@psychologie.tu-chemnitz.de

**Received:** 30 April 2018

**Accepted:** 07 August 2018

**Published:** 24 September 2018

### Citation:

Beggiato M, Hartwich F and Krems J  
(2018) Using Smartbands,  
Pupillometry and Body Motion to  
Detect Discomfort in Automated  
Driving.  
Front. Hum. Neurosci. 12:338.  
doi: 10.3389/fnhum.2018.00338

As technological advances lead to rapid progress in driving automation, human-machine interaction (HMI) issues such as comfort in automated driving gain increasing attention. The research project KomfoPilot at Chemnitz University of Technology aims to assess discomfort in automated driving using physiological parameters from commercially available smartbands, pupillometry and body motion. Detected discomfort should subsequently be used to adapt driving parameters as well as information presentation and prevent potentially safety-critical take-over situations. In an empirical driving simulator study, 40 participants from 25 years to 84 years old experienced two highly automated drives with three potentially critical and discomfort-inducing approaching situations in each trip. The ego car drove in a highly automated mode at 100 km/h and approached a truck driving ahead with a constant speed of 80 km/h. Automated braking started very late at a distance of 9 m, reaching a minimum of 4.2 m. Perceived discomfort was assessed continuously using a handset control. Physiological parameters were measured by the smartband Microsoft Band 2 and included heart rate (HR), heart rate variability (HRV) and skin conductance level (SCL). Eye tracking glasses recorded pupil diameter and eye blink frequency; body motion was captured by a motion tracking system and a seat pressure mat. Trends of all parameters were analyzed 10 s before, during and 10 s after reported discomfort to check for overall parameter relevance, direction and strength of effects; timings of increase/decrease; variability as well as filtering, standardization and artifact removal strategies to increase the signal-to-noise ratio. Results showed a reduced eye blink rate during discomfort as well as pupil dilation, also after correcting for ambient light influence. Contrary to expectations, HR decreased significantly during discomfort periods, whereas HRV diminished as expected. No effects could be observed for SCL. Body motion showed the expected pushback movement during the close approach situation. Overall, besides SCL, all other parameters showed changes associated with discomfort indicated by the handset control. The results serve as a basis for designing and configuring a real-time discomfort detection algorithm that will be implemented in the driving simulator and validated in subsequent studies.

**Keywords:** discomfort, automated driving, smartband, pupillometry, psychophysiology, motion tracking

## INTRODUCTION

Automated driving is expected to bring several mobility benefits such as improved traffic safety, reduced congestions and emissions, social inclusion, accessibility and more comfort (ERTRAC, 2017). As technological advances have enabled the rapid progression in driving automation, human-machine interaction (HMI) issues gain more attention and are considered a key question for broad public acceptance (Banks and Stanton, 2016; Riener et al., 2016; ERTRAC, 2017). One central HMI issue involves the question of how comfortable automated driving can be implemented to ensure a positive driving experience (Elbanhawi et al., 2015; ERTRAC, 2017; Bellem et al., 2018). Having a positive driving experience is a main factor for deciding to purchase and use a vehicle or in-vehicle system (Engelbrecht, 2013). In automated driving, discomfort could additionally lead to potential safety-critical situations, for example, due to (non-necessary) takeover with all associated risks such as reduced situation awareness (Hergeth et al., 2017). As the human role in automated driving changes from active driver to passenger, new and additional determinants of driving comfort are discussed, such as motion sickness, apparent safety, trust in the system, feelings of control, familiarity of driving maneuvers, and information about system states and actions (Beggiato et al., 2015; Elbanhawi et al., 2015; Bellem et al., 2016). There is no agreed-upon definition for comfort in the scientific community (Hartwich et al., 2018); however, existing comfort definitions share some central assumptions: comfort (a) is a subjective construct and, therefore, differs between individuals; (b) is affected by physical, physiological, and psychological factors; and (c) results from interaction with the environment (de Looze et al., 2003). Thus, comfort is hereby understood as a subjective, pleasant state of relaxation expressed through confidence and apparently safe vehicle operation (Constantin et al., 2014), “which is achieved by the removal or absence of uneasiness and distress” (Bellem et al., 2016, p. 45).

The research project KomfoPilot at Chemnitz University of Technology aims to investigate factors that influence comfort in automated driving. One objective is to find parameters that affect comfort on a general level, for example, situations and driving parameters such as speed, longitudinal/lateral distance, driving style familiarity, or personal characteristics (Hartwich et al., 2015, 2018). A second objective is the development of an algorithm for real-time discomfort detection to adapt driving style and information presentation at each moment once discomfort begins. The underlying idea is the metaphor of a vehicle–driver–team that knows each other’s strengths, limitations, and current states, and is able to react accordingly (Klein et al., 2004). The algorithm will be developed by project partners who specialize in data fusion (FusionSystems GmbH and Communication Engineering Department at Chemnitz University of Technology) and should combine data from different sensors such as in-car sensors (2D and 3D cameras, motion tracking), physiological sensors (smartband Microsoft Band 2, eye tracking), vehicle data and environment sensors. As a basis for developing the algorithm, the present article reports the results of the psychophysiological parameters

pupil diameter, eye blink frequency, heart rate (HR), heart rate variability (HRV), electrodermal activity (EDA) and body motion with regard to discomfort during automated driving in a driving simulator. Driving simulators offer an optimal environment for creating standardized situations under experimental control and applying sensors for measuring physiological parameters (Brookhuis and de Waard, 2011), although with limited external validity. The presented analyses aim to provide information about the potential of each parameter for detecting discomfort in an approaching automated situation, such as overall relevance, variability, direction and strength of effects, timing such as increase and decrease before and after discomfort as well as filtering and artifact removal strategies.

The use of these physiological parameters to infer mental states has a long research tradition. Despite results that are often contradictory, the main findings for these parameters are summarized subsequently and hypotheses regarding discomfort are derived. Pupil diameter has been studied largely as an indicator for mental effort, cognitive workload, stress, fatigue, information processing, affective processing and attention (Andreassi, 2000; Cowley et al., 2016). One of the major challenges in interpreting pupil size changes out of controlled lab studies is the heavy dependence on ambient light (Palinko and Kun, 2012). Despite these problems in separating the effects of ambient factors and mental states, a central finding is that pupil diameter increases with task difficulty, mental workload, emotionality of stimuli, and information-processing demands (Andreassi, 2000; Bacs and Boucsein, 2000; Cowley et al., 2016). Thus, an increase in pupil diameter is expected during uncomfortable situations. Eye blink rate is considered a sensitive indicator for mental workload, mood states, fatigue and task demands (Andreassi, 2000; Cowley et al., 2016). A decrease in blink rate in complex situations requiring visual attention has been found for car driving in complex situations as well as for fighter pilots (Bacs and Boucsein, 2000). Thus, a decrease in eye blink rate is expected during discomfort situations in automated driving, which are visually monitored by the driver.

The cardiovascular parameters HR and HRV are often used in driving simulation and on-road driving studies as indicators of mental effort, stress, workload, and task demands (see the overview of studies in Bacs and Boucsein, 2000; Mulder et al., 2005; Brookhuis and de Waard, 2011; Mehler et al., 2012; Ahonen et al., 2016; Schmidt et al., 2016). A common finding is that with higher invested effort and stress, HR increases and HRV decreases. The discomfort-inducing close approach situation investigated in this study could be seen as analogous to stress situations, including the uncertainty about the capability of a system to successfully complete a task. Thus, an increase in HR and a decrease in HRV during uncomfortable situations are expected. Similar to HR and HRV, EDA has a long tradition in psychophysiological research. Common findings include an increase of skin conductance level (SCL) with higher arousal, alertness, mental effort, workload, emotional load, stress, and task difficulty (Dawson et al., 2017). However, as EDA is sensitive to a wide variety of stimuli, it is not a clearly interpretable measure of any particular

psychological process and must be interpreted by including the stimulus conditions (Cowley et al., 2016; Dawson et al., 2017). For discomfort, an increase in SCL is expected due to a prediction of higher alertness and arousal. The HR, HRV, and EDA were measured using the smartband Microsoft (MS) Band 2. The use of a commercially available smartband was an explicit project goal to estimate the potential and problems of such a psychophysiological sensor. On the one hand, the market for smartbands is growing (Wade, 2017); thus, smartbands connected to vehicles could be an option for assessing psychophysiological parameters inside cars. On the other hand, the MS Band 2 has already been used in research for assessing mental workload in different environments (Binsch et al., 2016; Cropley et al., 2017; Reinerman-Jones et al., 2017; Schmalfuß et al., 2018), activity recognition in a home setting (Filippoupolitis et al., 2016), and for predicting and regulating personal thermal comfort in buildings (Laftchiev and Nikovski, 2016; Li et al., 2017).

Body motion during driving has mainly been investigated with regard to head movements for predicting driver intentions (Pech et al., 2014), hand movements for estimating driver distraction (Tran and Trivedi, 2009), trapezius muscle tension as an indicator for stress (Morris et al., 2017), or facial features for monitoring driver states (Baker et al., 2004). Moreover, the whole 3D driver posture is considered potentially useful for extracting information related to intentions, affective states, and distraction (Tran and Trivedi, 2010). However, posture dynamics are strongly related to situations and should, therefore, be combined with other contextual information (Tran and Trivedi, 2010). In the specific approach situation with the danger of a potential rear-end collision, a pushback movement is expected that should be reflected in motion tracking and seat pressure mat data. **Table 1** provides a summary of the expected effects during discomfort periods for all parameters.

## MATERIALS AND METHODS

### Study Design and Route

The driving simulator study was composed of two separate driving sessions with an approximate 2-month delay in between. Every driving session was composed of a 3-min highly automated trip on a straight, single carriageway, rural road. The trip was prerecorded and was exactly the same for all participants; there was no possibility to intervene by pedals or steering wheel. In every session, participants experienced three identical and potentially discomfort-inducing approach situations with the

danger of a potential rear-end collision (**Figure 1**). A white truck drove in front of the ego car with a constant speed of 80 km/h, whereas the ego car approached in a fully automated mode at 100 km/h. Automated braking was initialized very late at a distance of 9 m, which resulted in a minimum distance of 4.2 m and minimum time to contact of 1.1 s. After the approach, the ego car fell back at a distance of 100 m, and the approach started again. Participants were not informed about the situation and were instructed to press the lever of the handset control (**Figure 2A**) according to the extent of perceived discomfort. Thus, every participant experienced six approach situations in total, which resulted in 240 situations for all 40 participants and both sessions. The main reasons for inviting the participants twice were to: (a) obtain a higher overall number of discomfort situations per person; and (b) assess habituation effects within subjects over short and longer time periods (3 min vs. 2 months). Evaluation of habituation effects resulted in small to almost no effects, both for short- and long-term periods. Thus, all situations were included in the subsequent analyses.

### Participants

A total of 40 participants (15 females, 25 males) took part in both sessions of the study. Ages ranged from 25 years to 84 years with two distinct age groups, one from 25 years to 45 years (younger group,  $N = 21$ ,  $M = 30$  years,  $SD = 4.3$ ) and the other over 65 years (older group,  $N = 19$ ,  $M = 72$  years,  $SD = 6.0$ ). All subjects were required to currently hold a valid driver's license, and none of them had had previous experience of highly automated driving in the driving simulator. Participants were compensated with 20 euros for participation. This study was carried out in accordance with the recommendations, regulations and consent templates of the TU Chemnitz ethics commission. The protocol was approved by the TU Chemnitz ethics commission. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

### Material and Sensors

The study took place in a fixed-base driving simulator (SILAB 5.1 Software) with a fully equipped interior, a rear-view mirror, two side mirrors and a 180° horizontal field of view. Fully automated trips were prerecorded and replayed, while the participants sat in the driver's seat. Pedals and steering wheel were inoperative during these trips. Perceived discomfort was assessed during the whole trip by a handset control integrated into the driving simulator (Hartwich et al., 2015, 2018; **Figure 2A**). Participants could press the lever gradually in accordance with the extent of perceived discomfort. The

**TABLE 1** | Overview of expected effects of different sensor parameters during discomfort.

| Sensor          | Parameter               | Expected trend during discomfort  |
|-----------------|-------------------------|---|
| Eye tracking    | Pupil diameter          | Dilation (increase of diameter during mental effort/stress/attention/task difficulty) |
|                 | Blink rate              | Decrease (attention/arousal/alertness)  |
| MS Band 2       | Heart rate              | Increase (mental workload/stress)   |
|                 | Heart rate variability  | Decrease (mental workload/stress)   |
|                 | Skin conductance level  | Increase (mental workload/arousal/alertness/emotional response/stress)                |
| Motion tracking | Shoulder/head movements | Push-back/lean-back (decrease on z-axis)  |
| Pressure mat    | Pressure                | Push-back/lean-back (pressure increase at back position sensor)                       |



**FIGURE 1 |** Setup of the driving simulator study during one approach situation to the truck driving ahead. Left side: motion capture markers at head, right shoulder, and hands; handset control for reporting discomfort held in the right hand. Right side: driver camera view (top); front scenery camera view recorded from the roof of the mock-up (middle); reported discomfort as well as driving parameters at that particular moment such as TTC, ego speed/truck speed/speed difference and distance to the truck (bottom). Written informed consent was obtained from the individual for the publication of this image.

smartband Microsoft Band 2 (**Figure 2B**) was used to record the physiological parameters of HR, HRV and SCL via a Bluetooth connection. Accelerometer and gyroscope data were recorded as well from the band sensors to identify and correct for hand movements. The MS Band 2 was provided with a Software Development Kit that allowed for programming a dedicated logging application. Eye tracking data were recorded by SMI Eye Tracking Glasses 2 (SMI ETG 2, **Figure 2C**) and included pupil diameter, fixations, saccades and blinks. Participants already wearing eyeglasses ( $N = 10$ ) could not wear the SMI ETG 2, which resulted in less eye tracking data. In addition, the SMI ETG 2 were not applied in the whole second driving session because of testing camera-based, facial-feature recognition algorithms. Body motion was simultaneously captured by two sensor systems. The first device was a marker-based motion tracking system from OptiTrack composed of four Flex 13 infrared cameras recording with 120 fps (**Figure 2D**). A total of four distinct rigid bodies were tracked (left and right hand, right shoulder and head; see **Figures 1, 2D**). Rigid bodies are a collection of three or more markers on an undeformable object. These rigid bodies can be attached to tracked objects (e.g., clothes, gloves, headbands) and allow for recording position and orientation in six degrees of freedom. Participants with eyeglasses wore a headband with the rigid body attached (as in **Figure 1**), whereas the SMI ETG 2 allowed for directly attaching rigid bodies (**Figure 2C**). The second sensor system for body

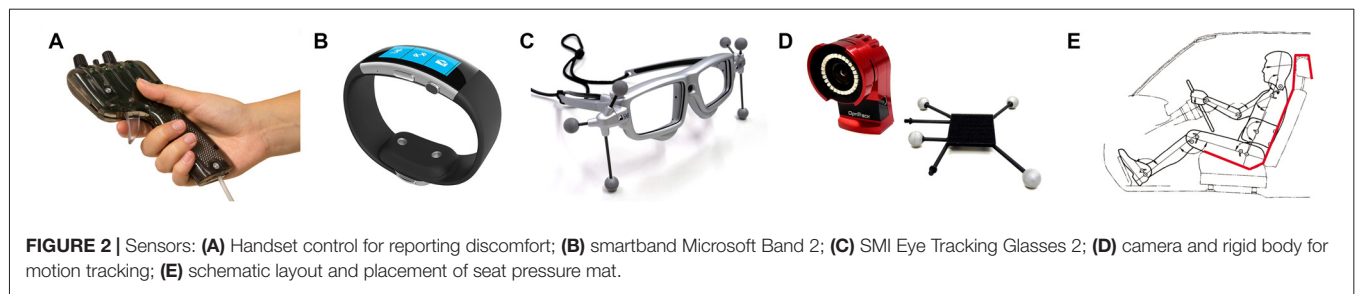
motion was a seat pressure mat developed by the project partner FusionSystems GmbH (**Figure 2E**). The mat can easily be placed on top of the seat and includes eight pressure sensors at different positions.

## Data Recording and Sequence Extraction

Data were recorded by several independent data loggers for each sensor with different recording frequencies. System time for all recording devices was continuously synchronized with a software tool based on the network time protocol (Meinberg NTP Software). Recording frequencies were 60 Hz for the driving simulator data, including handset control, 10 Hz for the MS Band 2, 60 Hz for the SMI ETG 2 eye tracking data, 120 Hz for motion tracking, and 10 Hz for the seat pressure mat. Raw data for each recorder were imported into a storage and analysis framework based on the relational open-source database management system PostgreSQL (Beggiato, 2015). The synchronization procedure was based on the timestamps of the driving simulator data (60 Hz) by adding the current value of all other sensor systems at this specific moment. To analyze changes in the sensor data with regard to perceived discomfort, data during reported discomfort by the handset control were compared with 10-s time intervals prior and after (**Figure 3**).

Discomfort intervals were extracted from the start of pressing the handset control lever until releasing, independent of the





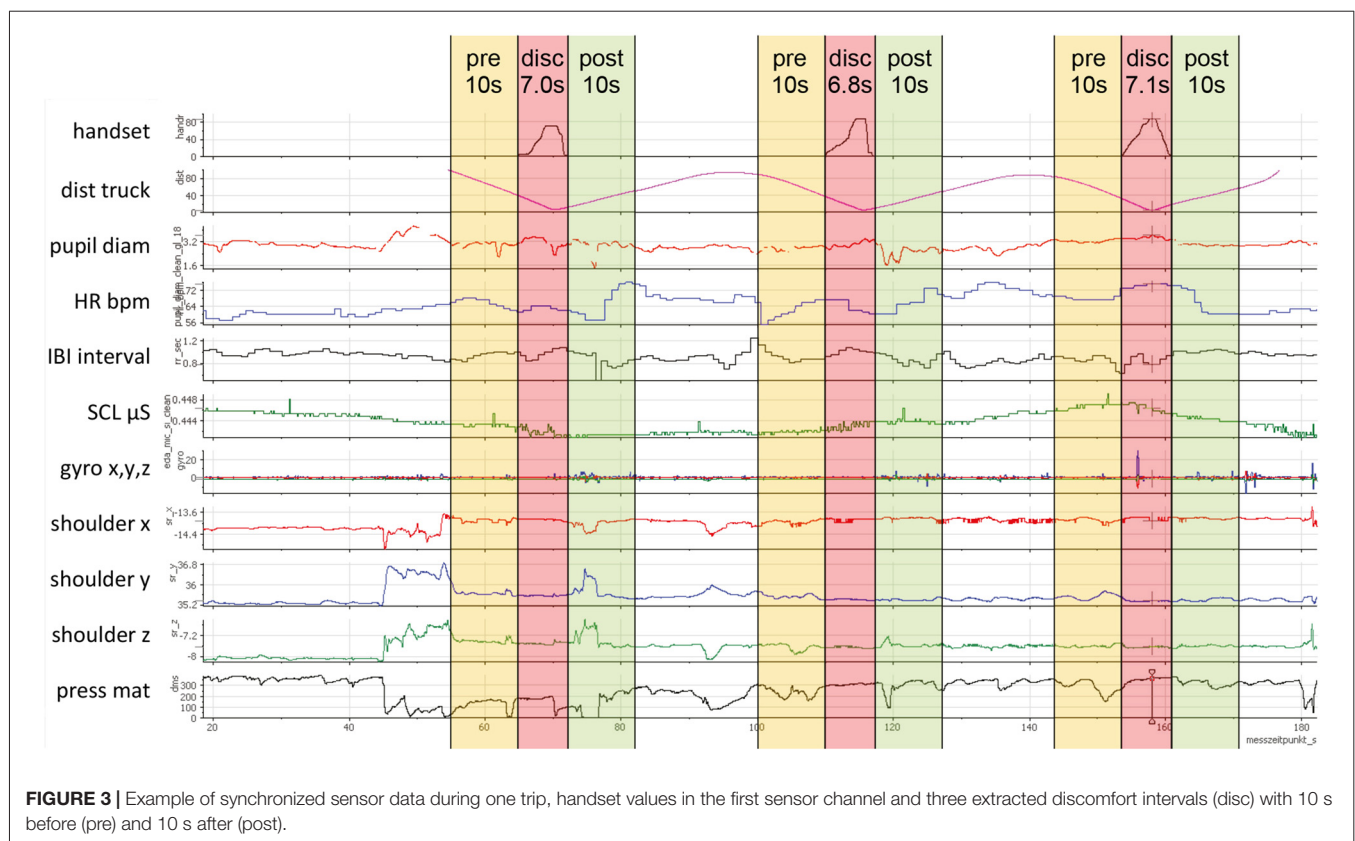
magnitude. However, the handset control was only pressed in 208 of the 240 approach situations. The distribution and descriptive statistics of the 208 extracted discomfort intervals are presented in **Figure 4**. In addition, single sensor channels were not recorded in some situations (e.g., no SMI ETG 2 for subjects already wearing eyeglasses or technical problems). Thus, all charts in the results section contain the respective number and mean duration of discomfort intervals that were included in the analysis. For the subsequent results section, the term “sequence” refers to the whole time period including the discomfort interval as well as the 10 s beforehand and afterwards.

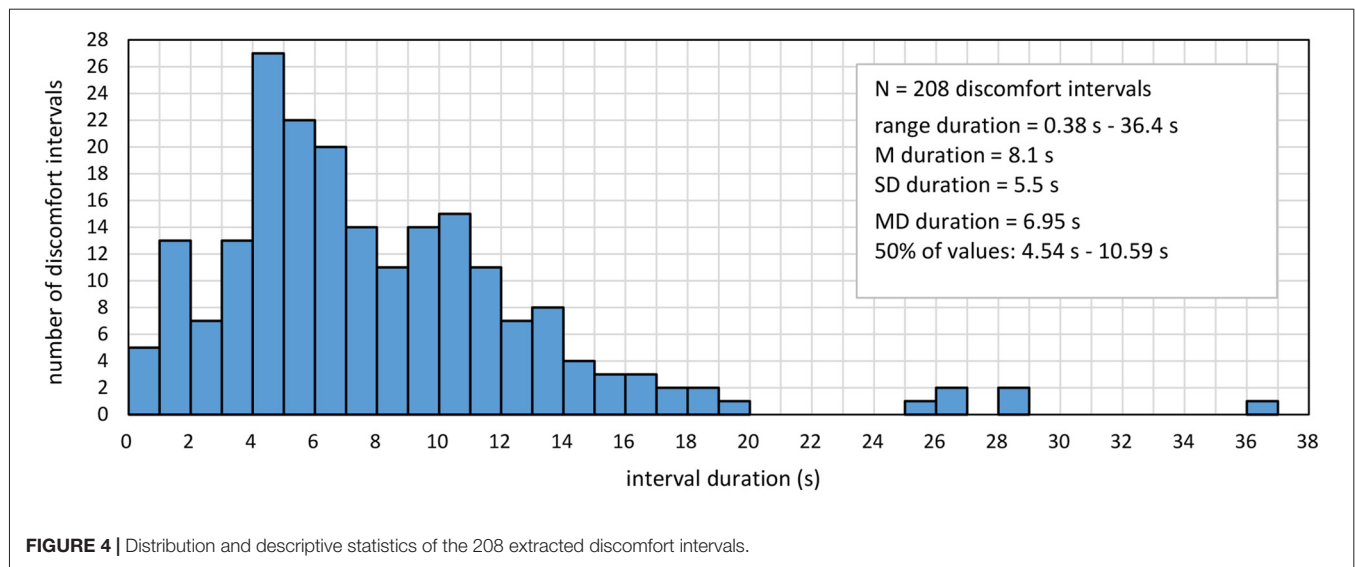
## Data Preparation

### Common X-Axis

To show the development of all assessed parameters before, during, and after the discomfort interval, a common time axis

was created for the charts in the results section (**Figure 5**). As the discomfort intervals varied in duration (**Figure 4**), a percent scale from 0% to 300% over the whole sequence was used to allow for displaying all values in the same scale. Periods before and after the discomfort interval were always 10 s long; thus, 1% corresponds to 0.1 s. Each discomfort interval was divided into percent slices, and the mean of each parameter was calculated for the specific time period of the respective percent slice. Finally, each percentage section before, during, and after reported discomfort was combined into one chart to show the progress of values over time. As not every sensor was active during the trips, each chart contains the number of sequences with mean duration and standard deviation of the included discomfort intervals in the caption. The main reason for using the percentage scale was to strictly respect the subjective aspect of discomfort mentioned in the definition. Thus, the different durations of reported discomfort intervals





**FIGURE 4 |** Distribution and descriptive statistics of the 208 extracted discomfort intervals.

should enter with the same weight in all analyses, which can be obtained by the percentage scale. In addition, the analysis method should also be applicable in less standardized situations, which requires a reliance on the reported handset values. However, using the percentage scale also has some drawbacks. It is not possible to give precise time-related indications, as it is not time, but the subjectively reported intervals that represent the unit of measurement. However, descriptive statistics about the intervals presented in **Figure 4** provide an indication of temporal dimensions. A second drawback is in regard to short sequences of a few seconds, in which some physiological processes such as changes in HR and SCL could hardly take effect. Similar concerns could be raised regarding longer sequences in terms of outliers, such as the six sequences over 20 s (**Figure 4**). However, as the percentage scale assigns the same weight to all sequences, excluding these six sequences does not change the results (tested for all analyses). Thus, despite these mentioned potential drawbacks, all sequences were included to present the overall picture.

### Z-Standardization and 95% Confidence Intervals

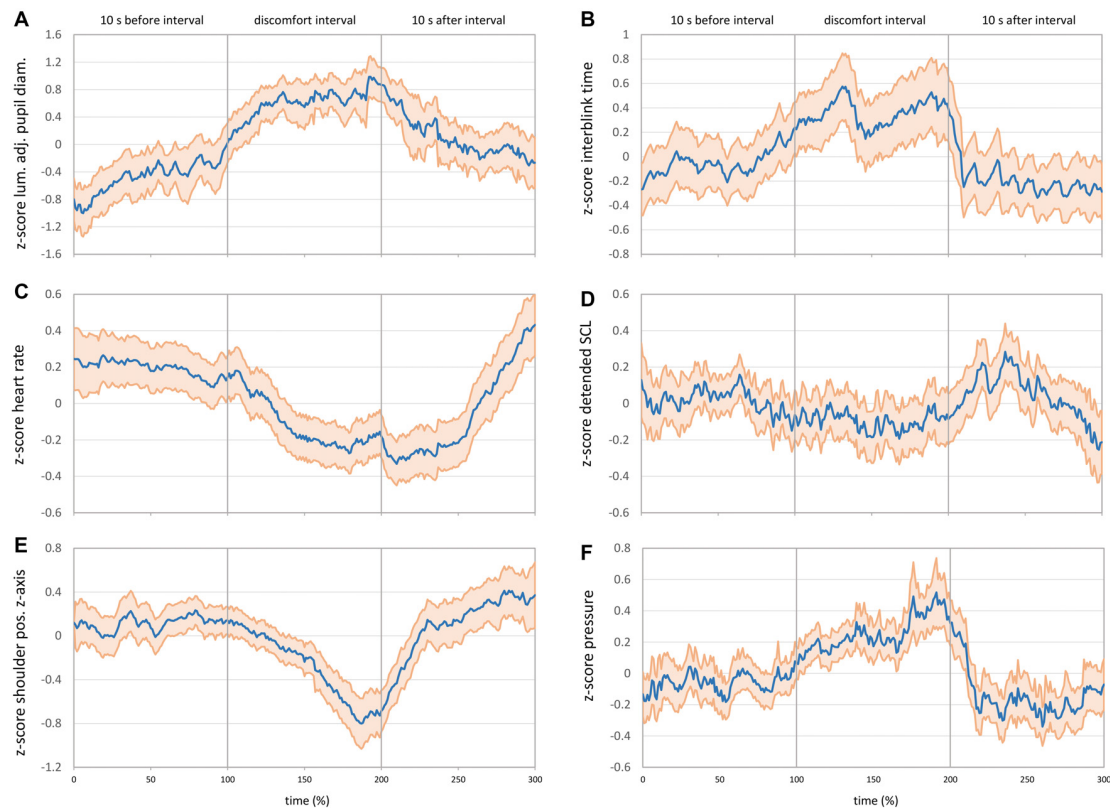
An important issue in processing psychophysiological data is distinguishing the signal of interest from noise (Gratton and Fabiani, 2017). Most of the physiological parameters such as HR or EDA have a strong individual component, which means that absolute values can hardly be compared between subjects. Thus, relative changes within one person provide better signal-to-noise ratio, for example, comparing changes of HR or EDA before, during and after discomfort intervals. However, these changes need to be transformed into a common scale to be compared between subjects. One of the common and best-performing transformations is the z-score, which expresses all values as the distance to the mean in units of standard deviations with a total mean of zero and a standard deviation of one (Jennings and Allen, 2017). Z-transformation was applied for each sequence, resulting in the relative changes over time in units of standard deviations. Resulting z-values were averaged

over all sequences at each single percent level from 0% to 300% and displayed as a blue line in the results charts (**Figure 5**). Beside these general transformations, some parameter-specific data correction methods and transformations were applied and are described for each parameter in the subsequent results sections. To obtain a quick estimation about the statistical significance of changes over time, the 95% confidence interval (CI) of each of these means was calculated pointwise and plotted as a light red area around the blue means. If the 95% CI does not overlap between two points in time, these two means differ in a statistically significant manner at  $p < 0.01$  (Field, 2013). The pointwise CI does not include multiplicity correction as would be the case for simultaneous confidence bands. Simultaneous CI bands control for the familywise error in autocorrelated time series by estimating the simultaneous coverage probability of the whole curve (Korpela et al., 2014; Francisco-Fernández and Quintela-del-Río, 2016; Ahonen et al., 2018). As the aim of the present analyses is not to fit a curve, but allow for visual comparison of single points in time, pointwise CIs were used. Pointwise CIs are narrower than a simultaneous CI band would be, and pointwise CIs allow only for comparing single points (as an ANOVA would do), but do not appropriately reflect the CI for the curve as a whole.

## RESULTS

### Pupil Diameter and Eye Blinks

Raw pupil diameters for the left and right eye (mm) from the SMI ETG 2 were averaged to get a single diameter from both eyes. To correct for signal fluctuations (especially close to blinks), a moving average over  $\pm 300$  ms was calculated, and a z-transformation of these values was applied for each sequence. As pupil diameter is not only dependent on mental states, but primarily on ambient light (Watson and Yellott, 2012), the metric could potentially be confounded during the white



**FIGURE 5 | (A)** Mean luminance-adjusted z-score of pupil diameter before, during and after discomfort intervals ( $N = 65$  intervals,  $M = 7.65$  s,  $SD = 4.74$  s). **(B)** Mean z-score of interblink interval time before, during and after discomfort intervals ( $N = 67$  intervals,  $M = 7.73$  s,  $SD = 4.70$  s). **(C)** Mean z-score of HR before, during and after discomfort intervals ( $N = 206$  intervals,  $M = 8.10$  s,  $SD = 5.52$  s). **(D)** Mean z-score of detrended SCL before, during and after discomfort intervals ( $N = 203$  intervals,  $M = 8.16$  s,  $SD = 5.51$  s). **(E)** Mean z-score of right shoulder movements on the z-axis before, during and after discomfort intervals ( $N = 114$  intervals,  $M = 7.85$  s,  $SD = 4.79$  s). **(F)** Mean z-score of pressure mat sensor at the back position before, during and after discomfort intervals ( $N = 202$  intervals,  $M = 8.04$  s,  $SD = 5.41$  s). The bold blue line shows the mean values, and the light red area shows the 95% pointwise confidence interval (CI) in all charts.

truck approach situation. Thus, the mean luminance value of all pixels (HSL color model) was calculated for each video frame of the SMI ETG 2 front camera video. A z-transformation of this mean luminance was applied for the whole trip in order to subtract these luminance z-scores from the z-scores of pupil diameter. The resulting luminance-adjusted z-values of pupil diameter are shown in **Figure 5A**. In line with the hypotheses, pupil diameter increased significantly during the discomfort interval and decreased steadily after reported discomfort. About 5 s after the end of the discomfort interval (approx. 250%), the 95% CI does not overlap anymore with the 95% CI during the discomfort interval (side note: without correcting for ambient luminance, the effects are the same but more pronounced).

Eye blink rate recorded by the SMI ETG 2 was computed in two different ways: first, blinks per second were calculated for each whole interval before, during, and after reported discomfort. **Figure 6A** shows the expected decrease in blink rate from 0.25 blinks per second before discomfort to 0.17 blinks per second during discomfort and the increase afterwards to 0.37 blinks per second ( $F_{(1.37,118.57)} = 26.37$ ,  $p < 0.001$ ,

$\eta_p^2 = 0.285$ ). However, this representation of blink rate does not allow for judging timings of increase/decrease as well as significance levels over time. Thus, it does not provide information for parameterizing an online detection algorithm. Therefore, a second way of obtaining a continuous blink rate was applied by calculating a running “interblink interval time.” This timer is set to zero every time a new blink is detected by the eye tracker and increases until the subsequent eye blink start is detected. Blink duration is not excluded and enters the running time. Z-values of this running interblink interval time were calculated for each sequence and averaged for each percent of time. **Figure 5B** shows the progress of interblink interval time z-scores with a noticeable increase during discomfort intervals (meaning less blinks) and the return to the prior level after the discomfort interval.

### Heart Rate and Heart Rate Variability

Raw HR values in beats per minute recorded by the MS Band 2 were transformed into z-values for each of the 206 sequences. **Figure 5C** shows the mean z-scores for HR over time. In contrast to the hypothesis, HR decreased steadily at the beginning of the

discomfort interval. The bottom HR plateau was reached at about the middle of the discomfort interval (150%) and kept until about 5 s after reported discomfort (250%). Afterward, HR rapidly rose up to approximately the prior level.

The HRV was computed using the interbeat interval times (IBI) in s from the MS Band 2. The HR and IBI are not exact reciprocal values in the case of the MS Band 2, but IBI is recommended for HRV calculations (Cropley et al., 2017). The time-domain metric root mean square successive difference (RMSSD) was calculated for each interval and averaged over all 202 sequences. The RMSSD is recommended for measuring high-frequency HRV and when time intervals to compare are not equally long (Berntson et al., 2017). Frequency domain and nonlinear HRV measures were not applied due to the relatively short time periods investigated. In line with the hypothesis, **Figure 6B** shows the expected u-shaped pattern with a decrease of HRV during reported discomfort ( $\chi^2_{(2)} = 40.05$ ,  $p < 0.001$ ; nonparametric Friedman's ANOVA).

## Skin Conductance Level

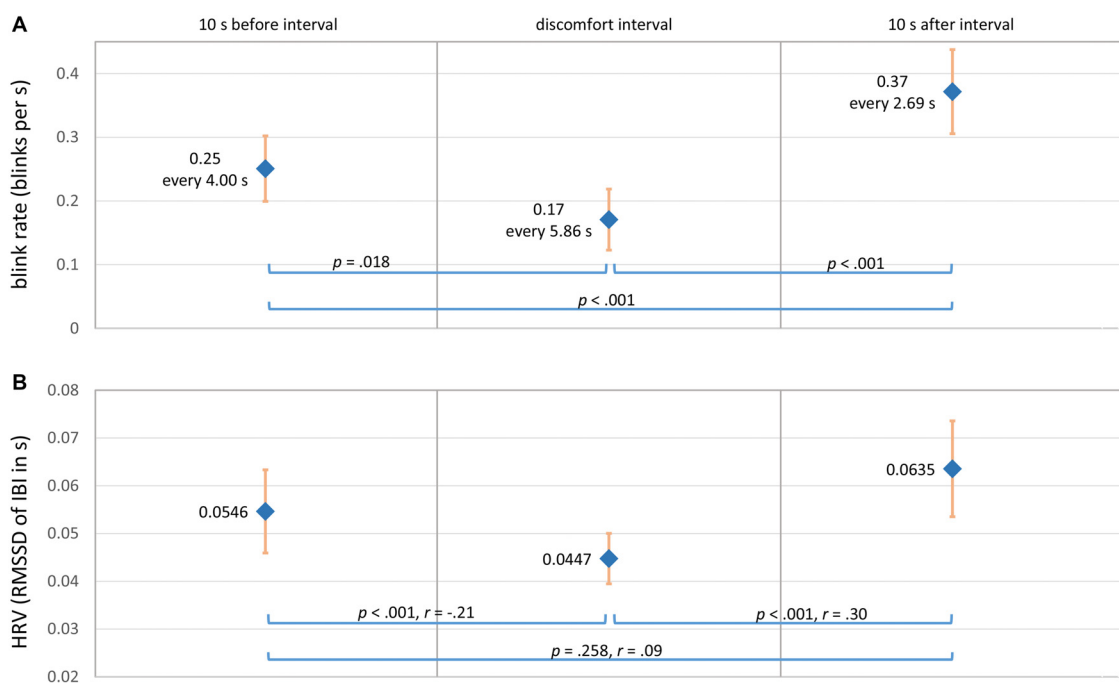
Two electrodes on the opposite side of the MS Band 2 display (**Figure 2B**) measured skin resistance level in kilo ohm. These values were inverted and multiplied by 1,000 to obtain the SCL in micro Siemens. The SCL values were very sensitive to changes in the hand/arm position such as placing a hand on the knees. Thus, SCL values were excluded (missing data) during high-movement episodes on the basis of the MS Band 2 accelerometer and gyroscope data. The remaining values were

z-standardized for each sequence. Results showed a continuous linear increase of SCL over time, independent of the situation. As this linear growing trend was probably related to the fact that subjects simply got warm during driving, a detrending algorithm was applied. Thus, a linear regression was calculated for each sequence. The SCL z-scores were subtracted from the regression values in order to obtain detrended z-scores, which are shown in **Figure 5D**. Detrended SCL showed almost no changes during the discomfort interval compared with the interval before and after.

## Body Movements

To assess body movements, data from the marker-based motion tracking system as well as the seat pressure mat were evaluated. The position of the right shoulder (mm) was captured by the motion tracking system. As the absolute marker position in the 3D space differed for each individual subject and each drive, differences on the z-axis position were computed for each sequence starting with zero at the beginning of the sequence. These value changes were transformed into z-scores. **Figure 5E** shows the mean z-scores of shoulder movement on the z-axis. As expected, the pushback of the body was represented by the u-shaped decrease of the shoulder z-position during the discomfort interval. Shoulder movements on the x- and y-axis showed similar but weaker effects; the main movement was backwards.

The pushback movement should also be represented in the data of the seat pressure mat, which would potentially allow for an easier movement measurement than motion tracking. To



**FIGURE 6 | (A)** Mean eye blink rate before, during and after discomfort intervals ( $N = 67$  intervals,  $M = 7.73$  s,  $SD = 4.70$  s). **(B)** Mean HRV (RMSSD) before, during and after discomfort intervals ( $N = 202$  intervals,  $M = 8.10$  s,  $SD = 5.53$  s). The bold blue dots show the mean values, and the light red bars show the 95% pointwise CI.



analyze the seat pressure mat data, the sensor at the back position was taken into account. Pressure values were *z*-transformed for each sequence. The *z*-scores of the pressure sensor (**Figure 5F**) showed the corresponding pattern to the motion-tracking results with an increase of pressure during the discomfort interval (pushback movement).

## DISCUSSION

The present study aimed at detecting discomfort in automated driving by physiological parameters from smartbands, pupillometry and body motion. Discomfort is considered an important issue for broad public acceptance of automated vehicles as well as for safety issues such as critical and not-necessary take-over situations. Considering the metaphor of a vehicle-driver-team that knows each other, automated systems could react to detected discomfort by changing driving style parameters and information presentation. An important basis for a real-time discomfort detection algorithm is information about physiological sensor parameters associated with reported discomfort, such as overall relevance, direction and strength of effects, timings, variability as well as filtering and artifact removal strategies.

Overall, besides SCL, all other assessed parameters like pupil diameter, eye blink rate, HR, HRV and body motion showed changes associated with discomfort indicated by the handset control. However, filtering and standardization procedures are required to increase the signal-to-noise ratio and remove bias caused by individual differences. In addition, every parameter has its own specificities, which are subsequently discussed.

Pupil diameter showed the expected inverse u-shaped pattern with a dilation during discomfort and recovery afterward, analogous to results regarding workload (Andreassi, 2000; Cowley et al., 2016). However, pupil diameter is not only dependent on mental states, but also primarily on ambient light conditions. Despite the fact that light conditions in the driving simulator do not change as much as on-road, a correction algorithm was applied by subtracting the *z*-standardized mean pixel luminance from the *z*-values of pupil diameter at every front camera video frame. Even with this adjustment, the effects are still observable. However, this quite simple adjustment procedure has some limitations. First, the exact association between ambient light and pupil diameter is much more complex than a simple linear relationship (Watson and Yellott, 2012). Second, cameras themselves adapt to ambient light, which does not allow to exactly measure luminance out of a video image. Third, eye tracking with the front camera can be used for lab experiments; in automated vehicles, luminance must be measured by other sensors. Despite these limitations, the applied adjustment procedure is real-time capable and will again be tested in subsequent studies within the project.

Eye blink rate showed the expected u-shaped pattern with fewer blinks during the discomfort interval (i.e., participants kept their eyes open in this situation). However, as the baseline blink rate is about one blink every 4 s, eye blinks are a “rare event” in relation to the duration of

discomfort intervals. Thus, the low frequency of eye blinks lowers the potential to serve as real-time predictor for discomfort.

Contrary to the expected trend, HR decreased during discomfort periods and returned to the prior level approximately 5 s after reported discomfort. A possible explanation for the unexpected decrease could be the effect of “preparation for action,” which means an anticipatory deceleration of HR prior to planned actions (Schandry, 1998; Cooke et al., 2014). The effect was reported for sport actions such as golf putting, but also for simpler reaction time (RT) paradigms: “It is well established that HR deceleration occurs during the fixed foreperiod of an RT task” (Andreassi, 2000, p. 270). The HRV measured by the RMSSD showed the expected u-shaped pattern with a decrease during the discomfort intervals.

The SCL showed a linear increasing trend over time, which could probably be explained by the effect that participants got warm during driving. After correcting for this linear trend using a regression approach, SCL showed almost no situation-related changes during discomfort intervals. The missing effects could be related to measurement procedures associated with the smartband. First, absolute SCL values were highly dependent on how tightly the band was closed. These differences could be corrected by the *z*-transformation; however, some bias could remain (e.g., when the band was worn very loosely). Second, SCL measures were taken from the outer side of the wrist, which is considered a much less sensitive place for SCL-changes compared with the fingers (Andreassi, 2000). Third, hand movements partly caused strong offsets in EDA values. The simple correction method of excluding these parts from the analysis could potentially be improved by more sophisticated algorithms.

The mentioned problems such as limited control on how tight the band is closed are to some extent related to the use of smartbands instead of more sophisticated measurement devices for physiological parameters. However, the aim of the KomfoPilot project was and is to estimate the potential of existing wearable devices with all the real-world usage challenges. Even with these problems, effects associated with discomfort could be identified in the data. One of the major challenges for using these devices will be the use of adequate signal analysis methods for gaining maximum signal-to-noise ratio.

Body movements captured by the pressure seat mat and the motion tracking of the right shoulder showed the expected pushback during the close approach to the truck. As posture dynamics are strongly related to specific situations (Tran and Trivedi, 2010), these movement patterns cannot automatically be generalized across different discomfort situations. However, discomfort associated with gaps that are too close or potential rear-end collisions could be detected involving body motion. A potential approach for data fusion algorithms could be the inclusion of environment sensor information such as time headway (Leonhardt et al., 2017) and to consider the pushback motion pattern only in these situations.

To sum up, the predicted mechanism of the monitored physiological signals for designing and configuring the real-time detection algorithm includes the following aspects. Most relevant parameters with the highest discomfort-specific changes resulted in ambient light-corrected pupil diameter, HR and the pushback movement. Interblink interval time and HRV measures showed changes, but could be unstable due to the short time intervals. The SCL from the MS Band 2 did not show specific changes and is, therefore, not recommended for inclusion in the algorithm. Regarding variability and filtering, relative changes within one person need to be assessed due to the strong individual component of all parameters. This could be achieved in real-time by e.g., performing individual  $z$ -standardization in sliding time windows and comparing the current signal value with these scores. This comparison could include several time windows of different lengths and different onsets, e.g., with 10 s and 5 s duration and an onset 3 s and 5 s before the current moment. This procedure would allow one to keep trace of the individual parameter variability by offering, at the same time, the application of standardized thresholds (such as a decrease in HR by 0.3  $SD$ -units compared to the sliding window). Threshold values as well as timings can be obtained from the results in **Figure 5** and can be adjusted to configure the sensitivity of the detection algorithm. To combine these predictions of each single parameter into one discomfort-score, probabilistic data fusion methods such as Bayesian Networks could be used. The nodes of such a network allow for integration of environment information (such as presence of a vehicle driving ahead) as well as for “inverting” the algorithm, once discomfort was detected, in order to return to the baseline. This method has already been applied by the Communication Engineering Department at Chemnitz University of Technology for real-time prediction of lane change maneuvers, combining parameters from the driver, the vehicle and the environment (Leonhardt et al., 2017).

In conclusion, the assessed parameters from smartbands, eye tracking and motion tracking showed potential for detecting discomfort in this approach situation. Despite commercially available smartbands providing less precise measures as dedicated lab devices, effects associated with discomfort could be identified. However, wearable devices also pose new challenges such as less control on how users apply

them. A limitation of this study is of course that only this specific truck approach situation has been investigated. The findings must be validated in other potentially discomfort-inducing situations, which are the next steps in the project. However, the use of this highly standardized approach situation also provides some advantages: (a) a distance that is experienced as too close is one of the most mentioned issues for discomfort as a codriver (dpa, 2013); (b) comfortable adjustment of headway distance and approach situations are not only relevant in conditional and high automation (SAE Levels 3 and above), but also for driver assistance systems such as adaptive cruise control and partial automation (SAE levels 1 and 2); and (c) the high standardization of the situation allowed for estimating the potential of different sensors as well as testing data filtering and artifact-removal strategies. Thus, the results serve as a basis for designing and configuring the real-time detection algorithm that is in development by the project partners who specialize in data fusion (FusionSystems GmbH and Communication Engineering Department at Chemnitz University of Technology). The algorithm will be implemented in the driving simulation software and tested in subsequent studies.

## AUTHOR CONTRIBUTIONS

MB, FH and JK contributed to conception and design of the study and also contributed to manuscript revision, read and approved the submitted version. MB performed sensor data preparation, setup of the PostgreSQL database and statistical analyses of sensor data. FH prepared and analyzed participant and questionnaire data. MB drafted major parts of “Introduction, Results and Discussion” sections. FH contributed to the “Introduction” section and drafted main parts of the “Materials and Methods” section. JK contributed to the discussion section.

## FUNDING

The research project KomfoPilot (2017–2019) is funded by the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung) under grant no. 16SV7690K. More information can be found at <https://www.tu-chemnitz.de/hsw/psychologie/professuren/allpsy1/english/traffic.php>.

## REFERENCES

- Ahonen, L., Cowley, B., Hellas, A., and Puolamäki, K. (2018). Biosignals reflect pair-dynamics in collaborative work: EDA and ECG study of pair-programming in a classroom environment. *Sci. Rep.* 8:3138. doi: 10.1038/s41598-018-21518-3
- Ahonen, L., Cowley, B., Torniaainen, J., Ukkonen, A., Vihavainen, A., Puolamäki, K., et al. (2016). Cognitive collaboration found in cardiac physiology: study in classroom environment. *PLoS One* 11:e0159178. doi: 10.1371/journal.pone.0159178
- Andreassi, J. L. (2000). *Psychophysiology: Human Behavior and Physiological Response*. 4th Edn. Mahwah, NJ: L. Erlbaum.
- Backs, R. W., and Boucsein, W. (2000). *Engineering Psychophysiology: Issues and Applications*. Mahwah, NJ: L. Erlbaum.
- Baker, S., Matthews, I., Xiao, J., Gross, R., Ishikawa, T., and Kanade, T. (2004). Real-time non-rigid driver head tracking for driver mental state estimation (No. CMU-RI-TR-04-10). Pittsburgh, PA: Carnegie Mellon University. Available online at: [http://ri.cmu.edu/pub\\_files/pub4/baker\\_simon\\_2004\\_1/baker\\_simon\\_2004\\_1.pdf](http://ri.cmu.edu/pub_files/pub4/baker_simon_2004_1/baker_simon_2004_1.pdf)
- Banks, V. A., and Stanton, N. A. (2016). Keep the driver in control: automating automobiles of the future. *Appl. Ergon.* 53, 389–395. doi: 10.1016/j.apergo.2015.06.020
- Beggiato, M. (2015). Changes in motivational and higher level cognitive processes when interacting with in-vehicle automation (Doctoral Dissertation). University of Technology Chemnitz. Available online at: <http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-167333> [Accessed 12 April 2017].
- Beggiato, M., Hartwich, F., Schleinitz, K., Krems, J. F., Othersen, I., and Petermann-Stock, I. (2015). What would drivers like to know during automated driving? Information needs at different levels of automation, in *Proceedings of the 7th Conference on Driver Assistance* (Munich: Technische Universität München (TUM)). doi: 10.13140/RG.2.1.2462.6007

- Bellem, H., Schöenberg, T., Krems, J. F., and Schrauf, M. (2016). Objective metrics of comfort: developing a driving style for highly automated vehicles. *Transp. Res. Part F Traffic Psychol. Behav.* 41, 45–54. doi: 10.1016/j.trf.2016.05.005
- Bellem, H., Thiel, B., Schrauf, M., and Krems, J. F. (2018). Comfort in automated driving: an analysis of preferences for different automated driving styles and their dependence on personality traits. *Transp. Res. Part F Traffic Psychol. Behav.* 55, 90–100. doi: 10.1016/j.trf.2018.02.036
- Berntson, G. G., Quigley, K. S., Norman, G. J., and Lozano, D. L. (2017). “Cardiovascular psychophysiology,” in *Handbook of Psychophysiology*, eds J. T. Cacioppo, L. G. Tassinary and G. G. Berntson, 4th Edn. (Cambridge: Cambridge University Press), 183–216.
- Binsch, O., Wabeke, T., and Valk, P. (2016). “Comparison of three different physiological wristband sensor systems and their applicability for resilience- and work load monitoring,” in *Proceedings of the IEEE 13th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, (San Francisco, CA), 272–276.
- Brookhuis, K. A., and de Waard, D. (2011). “Measuring physiology in simulators,” in *Handbook of Driving Simulation for Engineering, Medicine and Psychology*, eds D. L. Fisher, M. Rizzo, J. K. Caird and J. D. Lee (Boca Raton, FL: CRC Press), 17.1–17.10.
- Constantin, D., Nagi, M., and Mazilescu, C.-A. (2014). Elements of discomfort in vehicles. *Procedia Soc. Behav. Sci.* 143, 1120–1125. doi: 10.1016/j.sbspro.2014.07.564
- Cooke, A., Kavussanu, M., Gallicchio, G., Willoughby, A., McIntyre, D., and Ring, C. (2014). Preparation for action: Psychophysiological activity preceding a motor skill as a function of expertise, performance outcome, and psychological pressure. *Psychophysiology* 51, 374–384. doi: 10.1111/psyp.12182
- Cowley, B., Filetti, M., Lukander, K., Torniaainen, J., Henelius, A., and Ahonen, L. (2016). The psychophysiology primer: a guide to methods and a broad review with a focus on human-computer interaction. *Found. Trends Hum. Comp. Inter.* 9, 151–308. doi: 10.1561/1100000065
- Cropley, M., Plans, D., Morelli, D., Sütterlin, S., Inceoglu, I., Thomas, G., et al. (2017). The association between work-related rumination and heart rate variability: a field study. *Front. Hum. Neurosci.* 11:27. doi: 10.3389/fnhum.2017.00027
- Dawson, M. E., Schell, A. M., and Filion, D. L. (2017). “The electrodermal system,” in *Handbook of Psychophysiology*, eds J. T. Cacioppo, L. G. Tassinary and G. G. Berntson, 4th Edn. (Cambridge: Cambridge University Press), 217–243.
- de Looze, M. P., Kuijt-Evers, L. F. M., and van Dieën, J. (2003). Sitting comfort and discomfort and the relationships with objective measures. *Ergonomics* 46, 985–997. doi: 10.1080/0014013031000121977
- dpa. (2013). *Viele verängstigte Beifahrer Leiden im Stillen*. Welt.de. Available online at: [https://www.welt.de/newsticker/dpa\\_nt/serviceline/auto\\_verkehr/article113657517/Viele-veraengstigte-Beifahrer-leiden-im-Stillen.html](https://www.welt.de/newsticker/dpa_nt/serviceline/auto_verkehr/article113657517/Viele-veraengstigte-Beifahrer-leiden-im-Stillen.html)
- Elbanhawi, M., Simic, M., and Jazar, R. (2015). In the passenger seat: investigating ride comfort measures in autonomous cars. *IEEE Intell. Transp. Syst. Mag.* 7, 4–17. doi: 10.1109/itsmag.2015.2405571
- Engelbrecht, A. (2013). *Fahrkomfort und Fahrspaß beim Einsatz von Fahrerassistenzsystemen*. Hamburg: Disserta Verlag.
- ERTRAC. (2017). *Automated Driving Roadmap. European Road Transport Research Advisory Council*. Available online at: [http://www.ertrac.org/uploads/documentsearch/id48/ERTRAC\\_Automated\\_Driving\\_2017.pdf](http://www.ertrac.org/uploads/documentsearch/id48/ERTRAC_Automated_Driving_2017.pdf)
- Field, A. P. (2013). *Discovering Statistics Using IBM SPSS Statistics*. 4th Edn. Los Angeles, CA: SAGE.
- Filippoupolitis, A., Takand, B., and Loukas, G. (2016). “Activity recognition in a home setting using off the shelf smart watch technology,” in *Proceedings of the 15th International Conference on Ubiquitous Computing and Communications and 2016 International Symposium on Cyberspace and Security (IUCC-CSS)* (Granada: Spain), 39–44. doi: 10.1109/IUCC-CSS.2016.014
- Francisco-Fernández, M., and Quintela-del-Río, A. (2016). Comparing simultaneous and pointwise confidence intervals for hydrological processes. *PLoS One* 11:e0147505. doi: 10.1371/journal.pone.0147505
- Gratton, G., and Fabiani, M. (2017). “Biosignal processing in psychophysiology: principles and current developments,” in *Handbook of Psychophysiology*, eds J. T. Cacioppo, L. G. Tassinary and G. G. Berntson, 4th Edn. (Cambridge: Cambridge University Press), 628–661.
- Hartwich, F., Beggiato, M., Dettmann, A., and Krems, J. F. (2015). “Drive me comfortable: individual customized automated driving styles for younger and older drivers,” in *VDI (Eds.). Der Fahrer im 21. Jahrhundert. VDI-Berichte 2264*, (Düsseldorf: VDI-Verlag), 271–283.
- Hartwich, F., Beggiato, M., and Krems, J. F. (2018). Driving comfort, enjoyment and acceptance of automated driving—effects of drivers’ age and driving—style familiarity. *Ergonomics* 61, 1017–1032. doi: 10.1080/00140139.2018.1441448
- Hergeth, S., Lorenz, L., and Krems, J. F. (2017). Prior familiarization with takeover requests affects drivers’ takeover performance and automation trust. *Hum. Factors* 59, 457–470. doi: 10.1177/0018720816678714
- Jennings, J. R., and Allen, B. (2017). “Methodology,” in *Handbook of Psychophysiology*, 4th Edn, eds J. T. Cacioppo, L. G. Tassinary and G. G. Berntson (Cambridge: Cambridge University Press), 583–611.
- Klein, G., Woods, D., Bradshaw, J., Hoffman, R., and Feltovich, P. (2004). Ten challenges for making automation a “team player” in joint human-agent activity. *IEEE Intell. Syst.* 19, 91–95. doi: 10.1109/mis.2004.74
- Korpela, J., Puolamäki, K., and Gionis, A. (2014). Confidence bands for time series data. *Data Min. Knowl. Discov.* 28, 1530–1553. doi: 10.1007/s10618-014-0371-0
- Laftchiev, E., and Nikovski, D. (2016). “An IoT system to estimate personal thermal comfort,” in *Proceedings of the IEEE 3rd World Forum on Internet of Things (WF-IoT)* (Reston, VA: IEEE), 672–677.
- Leonhardt, V., Pech, T., and Wanielik, G. (2017). “Fusion of driver behaviour analysis and situation assessment for probabilistic driving manoeuvre prediction,” in *URBAN Human Factors in Traffic. Approaches for Safe, Efficient and Stressfree Urban Traffic*, eds K. Bengler, S. Hoffmann, D. Manstetten and A. Neukum (Wiesbaden: Springer Vieweg Verlag), 223–244.
- Li, D., Menassa, C. C., and Kamat, V. R. (2017). Personalized human comfort in indoor building environments under diverse conditioning modes. *Build Environ.* 126, 304–317. doi: 10.1016/j.buildenv.2017.10.004
- 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. *Hum. Factors* 54, 396–412. doi: 10.1177/0018720812442086
- Morris, D. M., Erno, J. M., and Pilcher, J. J. (2017). “Electrodermal response and automation trust during simulated self-driving car use,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Clemson, SC: Clemson University), 1759–1762. doi: 10.1177/1541931213601921
- Mulder, L., de Waard, D., and Brookhuis, K. A. (2005). “Estimating mental effort using heart rate and heart rate variability,” in *Handbook of Human Factors and Ergonomics Methods*, eds N. A. Stanton, A. Hedge, K. A. Brookhuis, E. Salas and H. Hendrick (Boca Raton, FL: CRC Press), 21–28.
- Palinko, O., and Kun, A. L. (2012). “Exploring the effects of visual cognitive load and illumination on pupil diameter in driving simulators,” in *ETRA’12, Proceedings of the Symposium on Eye Tracking Research and Applications* (New York, NY: ACM), 413–416.
- Pech, T., Lindner, P., and Wanielik, G. (2014). “Head tracking based glance area estimation for driver behaviour modelling during lane change execution,” in *Proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems (ITSC14)* (New York, NY: United States), 655–660.
- Reinerman-Jones, L., Harris, J., and Watson, A. (2017). “Considerations for using fitness trackers in psychophysiology research,” in *Human Interface and the Management of Information: Information, Knowledge and Interaction Design (LNCS 10273)*, ed. S. Yamamoto (Cham: Springer International Publishing), 598–606. Available online at: [https://link.springer.com/chapter/10.1007/978-3-319-58521-5\\_47](https://link.springer.com/chapter/10.1007/978-3-319-58521-5_47).
- Riener, A., Boll, S., and Kun, A. L. (2016). *Automotive User Interfaces in the Age of Automation (Dagstuhl Reports)*, 6, 111–159. doi: 10.4230/DagRep.6.6.111
- Schandry, R. (1998). *Lehrbuch Psychophysiologie: Körperliche Indikatoren Psychischen Geschehens (Studienausg.)*. Weinheim: Beltz, Psychologie Verlags Union.
- Schmalfuß, F., Mach, S., Klüber, K., Habelt, B., Beggiato, M., Körner, A., et al. (2018). “Potential of wearable devices for mental workload detection in different physiological activity conditions,” in *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2017 Annual Conference*, eds D. de Waard, F. Nocera, D. Coelho, J. Edworthy, K. Brookhuis and F. Ferlazzo (Netherlands: University of

- Groningen), 179–191. Available online at: <https://www.hfes-europe.org/wp-content/uploads/2017/10/Schmalfuss2017.pdf>
- Schmidt, E., Decke, R., and Rasshofer, R. (2016). “Correlation between subjective driver state measures and psychophysiological and vehicular data in simulated driving,” in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, (Gothenburg, Sweden), 1380–1385.
- Tran, C., and Trivedi, M. M. (2009). “Driver assistance for ‘Keeping hands on the wheel and eyes on the road,’” in *Proceedings of the IEEE International Conference on Vehicular Electronics and Safety (ICVES)*, (Pune), 97–101.
- Tran, C., and Trivedi, M. M. (2010). “Towards a vision-based system exploring 3D driver posture dynamics for driver assistance: issues and possibilities,” in *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, (San Diego, CA), 179–184.
- Wade. (2017). Wearable technology statistics and trends 2018. Available online at: <https://www.smartinsights.com/digital-marketing-strategy/wearables-statistics-2017/>
- Watson, A. B., and Yellott, J. I. (2012). A unified formula for light-adapted pupil size. *J. Vis.* 12:12. doi: 10.1167/12.10.12
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- The reviewer DW and handling Editor declared their shared affiliation.
- Copyright © 2018 Beggiato, Hartwich and Krems. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# A Novel Method for Classifying Driver Mental Workload Under Naturalistic Conditions With Information From Near-Infrared Spectroscopy

Anh Son Le<sup>1,2\*</sup>, Hirofumi Aoki<sup>1</sup>, Fumihiko Murase<sup>3</sup> and Kenji Ishida<sup>3</sup>

<sup>1</sup> Human Factors and Aging Laboratory, Institutes of Innovation for Future Society, Nagoya University, Nagoya, Japan,

<sup>2</sup> Department of Power Engineering, Faculty of Engineering, Vietnam National University of Agriculture, Hanoi, Vietnam,

<sup>3</sup> Denso Corporation, Nisshin, Japan

## OPEN ACCESS

### Edited by:

Bruce Mehler,  
Massachusetts Institute of  
Technology, United States

### Reviewed by:

Noman Naseer,  
Air University, Pakistan  
Sung-Phil Kim,  
Ulsan National Institute of Science and  
Technology, South Korea  
Toshinori Kato,  
KatoBrain Co., Ltd., Japan  
Mauricio Muñoz,  
Bosch Center for Artificial Intelligence,  
Germany

### \*Correspondence:

Anh Son Le  
leanhsonvn@gmail.com

**Received:** 12 April 2018

**Accepted:** 02 October 2018

**Published:** 26 October 2018

### Citation:

Le AS, Aoki H, Murase F and Ishida K  
(2018) A Novel Method for Classifying  
Driver Mental Workload Under  
Naturalistic Conditions With  
Information From Near-Infrared  
Spectroscopy.  
*Front. Hum. Neurosci.* 12:431.  
doi: 10.3389/fnhum.2018.00431

Driver cognitive distraction is a critical factor in road safety, and its evaluation, especially under real conditions, presents challenges to researchers and engineers. In this study, we considered mental workload from a secondary task as a potential source of cognitive distraction and aimed to estimate the increased cognitive load on the driver with a four-channel near-infrared spectroscopy (NIRS) device by introducing a machine-learning method for hemodynamic data. To produce added cognitive workload in a driver beyond just driving, two levels of an auditory presentation n-back task were used. A total of 60 experimental data sets from the NIRS device during two driving tasks were obtained and analyzed by machine-learning algorithms. We used two techniques to prevent overfitting of the classification models: (1) *k*-fold cross-validation and principal-component analysis, and (2) retaining 25% of the data (testing data) for testing of the model after classification. Six types of classifier were trained and tested: decision tree, discriminant analysis, logistic regression, the support vector machine, the nearest neighbor classifier, and the ensemble classifier. Cognitive workload levels were well classified from the NIRS data in the cases of subject-dependent classification (the accuracy of classification increased from 81.30 to 95.40%, and the accuracy of prediction of the testing data was 82.18 to 96.08%), subject 26 independent classification (the accuracy of classification increased from 84.90 to 89.50%, and the accuracy of prediction of the testing data increased from 84.08 to 89.91%), and channel-independent classification (classification 82.90%, prediction 82.74%). NIRS data in conjunction with an artificial intelligence method can therefore be used to classify mental workload as a source of potential cognitive distraction in real time under naturalistic conditions; this information may be utilized in driver assistance systems to prevent road accidents.

**Keywords:** near-infrared spectroscopy, cognitive distraction, classification, driver attention, mental workload, artificial intelligence

## INTRODUCTION

Driver distraction is a major cause of traffic accidents (NHTSA, 2015). An analysis by the US Highway Traffic Safety Administration (NHTSA) showed that driver distraction can be categorized into three types: visual distraction, manual distraction, and cognitive distraction (NHTSA, 2012). Among these, cognitive distraction is the most difficult type to address, because it occurs within the driver's brain (Rizzo and Hurtig, 1987; Engström et al., 2005; Angell et al., 2006). Cognitive distraction is defined as the mental workload associated with a task that involves thinking about something other than driving. The detection of cognitive distraction imposed by a secondary task while driving might play an important role in creating a new driver-assistance system to reduce the incidence of traffic accidents.

Dong et al. (2011) categorized techniques for measuring mental workload while driving into five groups: (1) subjective metrics, (2) biological metrics, (3) physical metrics, (4) performance metrics, and (5) combinations of these metrics. Because the central goal of our research was to identify and improve a metric that might permit the detection of mental workload in real time and which could operate under real conditions in the presence of, for example, vibration from the vehicle, we examined only physical metrics in the present study.

One potential physical metric involves the use of eye-movement information. Many researchers have previously attempted to identify a relationship between mental workload and various items of information on the eye, such as blink (Tsai et al., 2007; Benedetto et al., 2011), pupil diameter (Backs and Walrath, 1992; Klingner et al., 2008; Schwalm et al., 2008; Klingner, 2010), saccades (Tsai et al., 2007; Pierce, 2009; Tokuda et al., 2009), gaze concentration (Wang et al., 2014), or eye fixation (Klingner, 2010). Each of these methods has its advantages and disadvantages. For example, pupil diameter has a strong relationship to the level of cognitive load but it is also highly sensitive to the frequent changes in light that occur while driving (Palinko and Kun, 2012). Another potential method is to use the involuntary eye movements based on the vestibulo-ocular reflex model that are simulated by head movements or by vibrations from the moving vehicle. In this method, differences between predicted and actual eye simulation are assessed as a measure of mental workload (Obinata et al., 2008, 2009, 2010; Aoki et al., 2015; Anh Son et al., 2016, 2017a,b,c,d,e, 2018; Le and Aoki, 2018; Son and Hirofumi, 2018; Son et al., 2018). However, the use of eye information to measure mental workload still has some limitations, such as oversensitivity to light, vibration, noise, and visual information.

In terms of a physical metric, monitoring of brain activity by electroencephalography or the use of the heart rate as an indicator of mental workload have been confirmed to be effective (Meshkati, 1988; Lee and Park, 1990; Jorna, 1992; Porges and Byrne, 1992; Veltman and Gaillard, 1996; Ryu and Myung, 2005; Henelius et al., 2009; Mehler et al., 2012; Cinaz et al., 2013; Angell and Perez, 2015). However, these techniques require physical attachment of the monitoring equipment and are highly sensitive to the driver age, body position, and muscle activity.

One method with a high potential for application is the use of information from near-infrared spectroscopy (NIRS) to classify mental workload (Kopton and Kenning, 2014). NIRS has been used in various fields; for example, in agriculture to check the quality of crops and in medicine to assess oxygenation and microvascular function. In terms of classifying mental workload, a relationship between mental workload and activity of the central nervous system has been confirmed by McBride and Schmorow (2005). Since their work, other researchers have attempted to classify levels of mental workload by applying artificial-intelligence analyses (Tsunashima and Yanagisawa, 2009; Herff et al., 2014; Ichikawa et al., 2014; Aghajani et al., 2017). All of these researchers showed that NIRS has considerable potential in quantifying mental workload while driving, especially in naturalistic cases (Kopton and Kenning, 2014; Liu et al., 2016).

Furthermore, Toshinori Kato and his group have done various investigations on NIRS data, especially how to filter the signals and map it (Kohri et al., 2002; Yoshino et al., 2013, 2015; Orino et al., 2015). In actual driving, his group pointed out that there was a relationship of the brain activity with the vehicle speed. Their research also confirmed that fNIRS data is one of a good solution for monitoring the driver status while driving especially in actual condition. Further, Liu et al. (2017) confirmed that the cognitive workload has a relationship with the hemodynamic activity level (Liu et al., 2017). His team also mentioned that the effective association can be weak in case of driving with subtasks. However, none of them investigated in applying machine learning with raw data to detect mental workload by secondary task while driving.

The central goal of our study was to examine whether or not it is possible to classify driver mental workload by using supervised learning with NIRS data obtained in a real vehicle. In this report, we initially point out the importance of detecting cognitive load while driving. We then review and summarize methods for evaluating cognitive workload that have been reported in the literature. We also discuss the differences in mental workload between doing one task and driving with secondary task. We then review the use of NIRS information to classify cognitive load, and we describe our experimental design and methods for analyzing NIRS data. The result of the classification are reported and, finally, we discuss our conclusions and any challenges that remain.

## MATERIALS AND METHODS

### Experimental Design

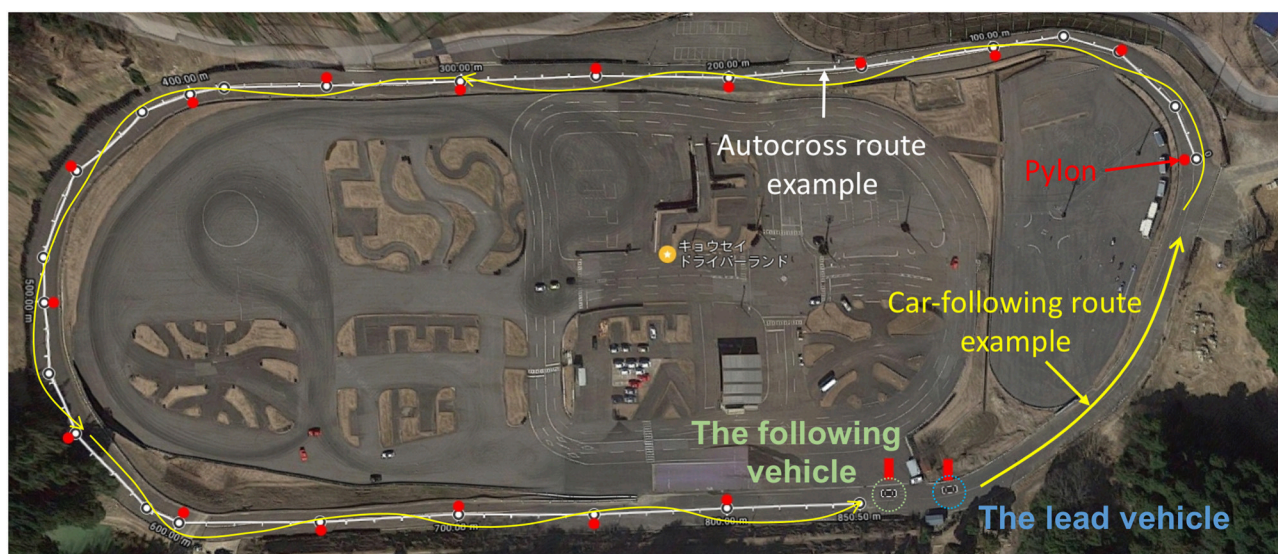
One female and four male subjects, who each held a driver's license (mean age:  $38 \pm 10$ ; two professional drivers, and three newly qualified drivers), were recruited for this test. A total of 60 experiments were performed involving navigating a defined course alone (autocross) or following another vehicle (car-following) on a test course (**Figure 1**). The experiments were approved by the Ethical Review Board of Nagoya University's Institute of Innovation for Future Society. All subjects were provided with explanations regarding the

experimental procedure, and all gave their written informed consent.

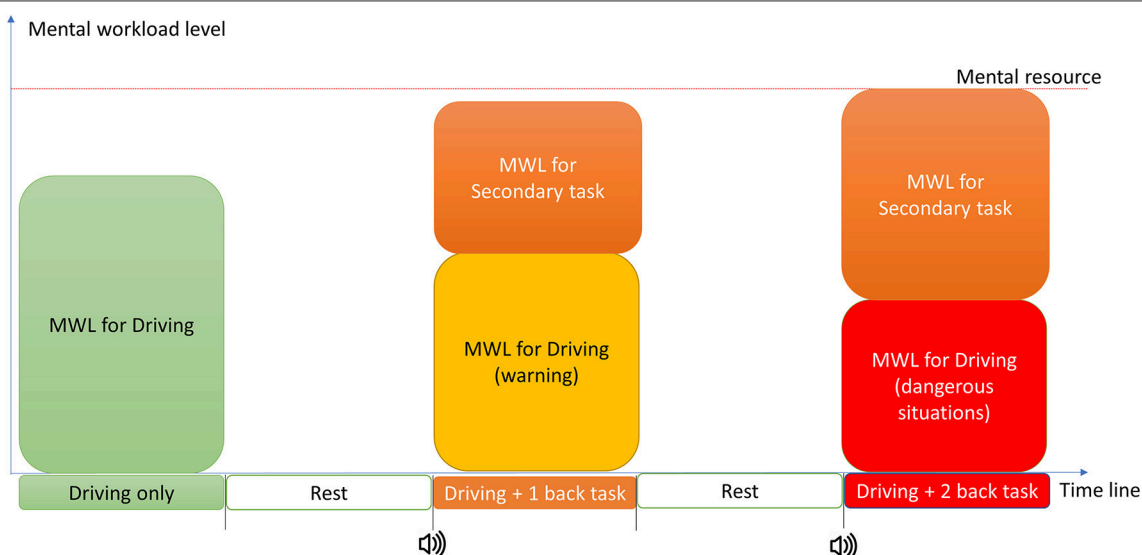
The experiment procedure is shown in **Figure 2**. The subjects performed a series of tests involving a driving task (autocross or car following) during which drivers were asked to drive around 40 km/h. For portions of these drives, additional mental workload was introduced by asking subjects to engage in two levels of an n-back auditory digit recall task in which a number was verbally presented to the subject every 2 s. In the 1-back test, the subject was asked to press the “Yes” button when the number heard was the same as the previous one or the “No” button when it was different. In the 2-back task, the subject similarly had to remember the number preceding the previous one. The “Yes” and

“No” buttons were installed on the driving wheel so they could be easily pressed.

As our main aim is to create an algorithm for an advanced driver-assistance system to help prevent traffic accidents by identifying driver cognitive distraction, the classification needed to be reliable, quick, and easy. To achieve this, we used a commercial four-channel NIRS system (Astem Corp., Fukuoka, Japan) which was placed on the forehead of the subject, where the signals from the four channels are almost the same (**Figure 3**). This device can measure blood oxyhemoglobin (oxy-Hb) and deoxyhemoglobin (deoxy-Hb) levels at wavelengths of 770 nm (probe distance 35 mm) and 830 nm (probe distance is 40 mm), and oxygen saturation at 35 mm.

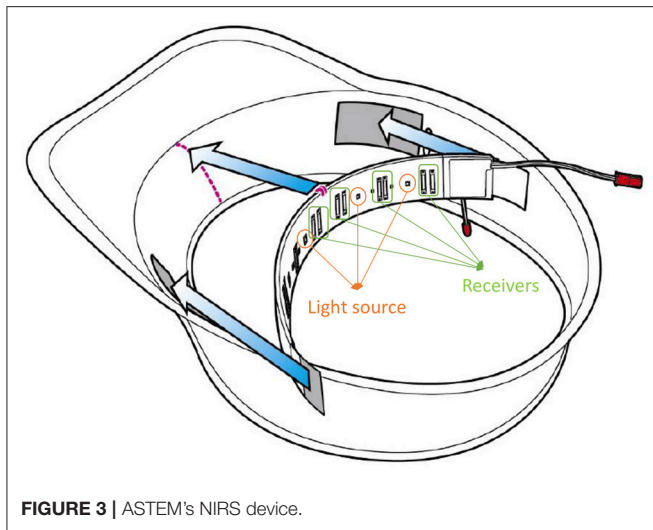


**FIGURE 1 |** Test course.



**FIGURE 2 |** Experiment procedure and hypothetical image of MWL.





Furthermore, to keep the class-balance data set for machine learning, the subjects were asked to drive around the course for each task (driving task only, driving plus 1-back task, and driving plus 2-back task), and repeat it twice (average time for each trial was  $1 \text{ min } 54 \pm 16 \text{ s}$  depending upon the speed actually driven). Therefore, the sample data input to machine learning step was class-balance.

## Data Processing

**Figure 4** provides an overview of the method we used to preprocess the NIRS data. Because of noise arising from movement artifacts (Cooper et al., 2012; Kirilina et al., 2013), the raw NIRS data from each channel were preprocessed by using a modified form of the Beer–Lambert Law (Huppert et al., 2016) with removal of lost data (time shift) (Kirilina et al., 2013), bandpass filtration (0.02–1 Hz) (Ichikawa et al., 2014), and Kalman filtration (Abdelnour and Huppert, 2009), before finally being transformed into features.

In our experiment, the raw data of each channel includes (**Figure 5**): oxyhemoglobin at 35 mm (OxHb35), deoxyhemoglobin at 35 mm (DoxHb35), total oxyhemoglobin (ToxHb35), absolute tissue saturation (StO<sub>2</sub>), oxyhemoglobin at 40 mm (OxHb40), deoxyhemoglobin at 40 mm (DoxHb40). After filtering, all of the information from NIRS will be transform to 6 features namely OxHb35, DoxHb35, ToxHb35, StO<sub>2</sub>, OxyHb40, and DoxHb40 for preparing to input for machine learning step.

These features were then processed to create training data and testing data for subject-dependent, subject-independent, channel-independent, and subject-independent plus channel-independent cases. After taking all of the data, they were divided into 75% (training data) and 25% (testing data, which is not used to improve the model, but to measure its predictive performance; **Figure 6**). Because we used four channels for the forehead, the data combinations were obtained merely by combining the data together. To prevent overfitting during machine learning, a fivefold cross-validation and a principal-component analysis were applied before the data were used to train the system. The fivefold cross-validation was conducted in the following three

steps. First, the training data (75% of all data) was split into 5-fold. Second, a model for each fold using all the data outside the fold (75% of the training data) was trained and validated. After that, the features were transformed with PCA to reduce the dimensionality of the predictor space (we applied 5 principal components).

Definition:

- Subject-dependent (Subject-dependent + channel-dependent; **Figure 7**): data of each channel for each subject in all task was combined as the input of the machine learning step. Total 20 datasets were prepared for running machine learning.
- Subject-independent (Subject-independent + channel-dependent; **Figure 8**): data of each channel for all subjects in all task was combined as the input of the machine learning step. Total 4 datasets were prepared for running machine learning.
- Channel-independent (Channel-independent + Subject-dependent; **Figure 7**): data of all channel for each subject in all task were combined as the input of machine learning step. Total 5 datasets were prepared for running machine learning.
- Subject-independent + Channel-independent (**Figure 8**): data of all channel for all subject in all task were combined as the input of machine learning step. Only one data set was prepared for running machine learning.

## The Classification Method

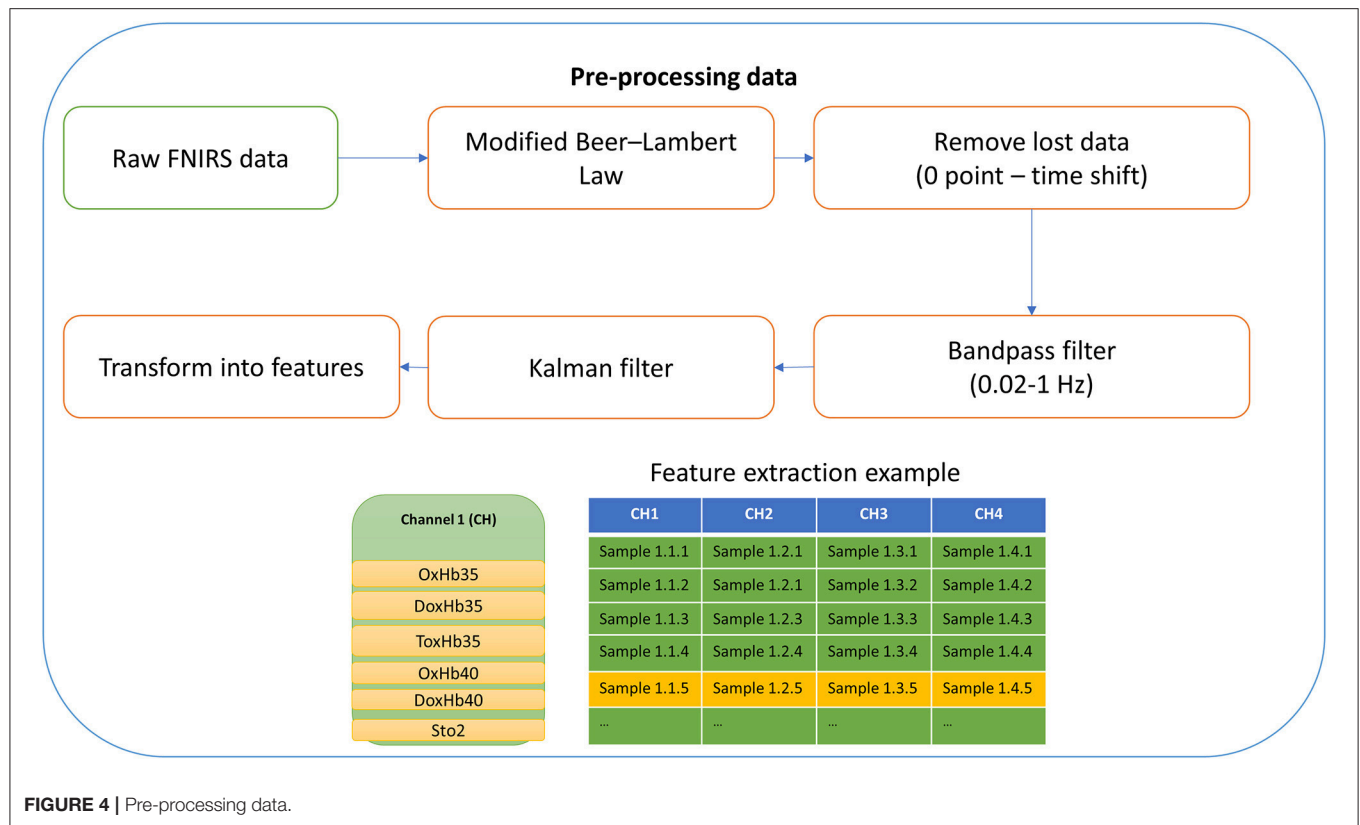
Previous studies on the classification of mental workload from NIRS data have used an SVM (Devos et al., 2009; Ichikawa et al., 2014; Aghajani et al., 2017), linear discriminant analysis (Luu and Chau, 2009), the hidden Markov model (Sitaram et al., 2007; Zimmermann et al., 2013), or artificial neural networks (Chan et al., 2012; Thanh Hai et al., 2013). However, most of these studies involved complicated multichannel NIRS systems. In our study, because of the large number of samples and the low number of channels, we applied supervised learning in MATLAB 2017b (MathWorks Inc., Natick, MA, United States) (**Figure 5**). We used the 75% of the data to train several well-known models, including the decision-tree model, the discriminant-analysis model, the logistic-regression model, SVMs, nearest-neighbor classifiers, and ensemble classifiers. The performance of these classifiers was determined from the accuracy, as calculated by using the equation shown below:

$$\text{Accuracy } (A_{cc}) = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

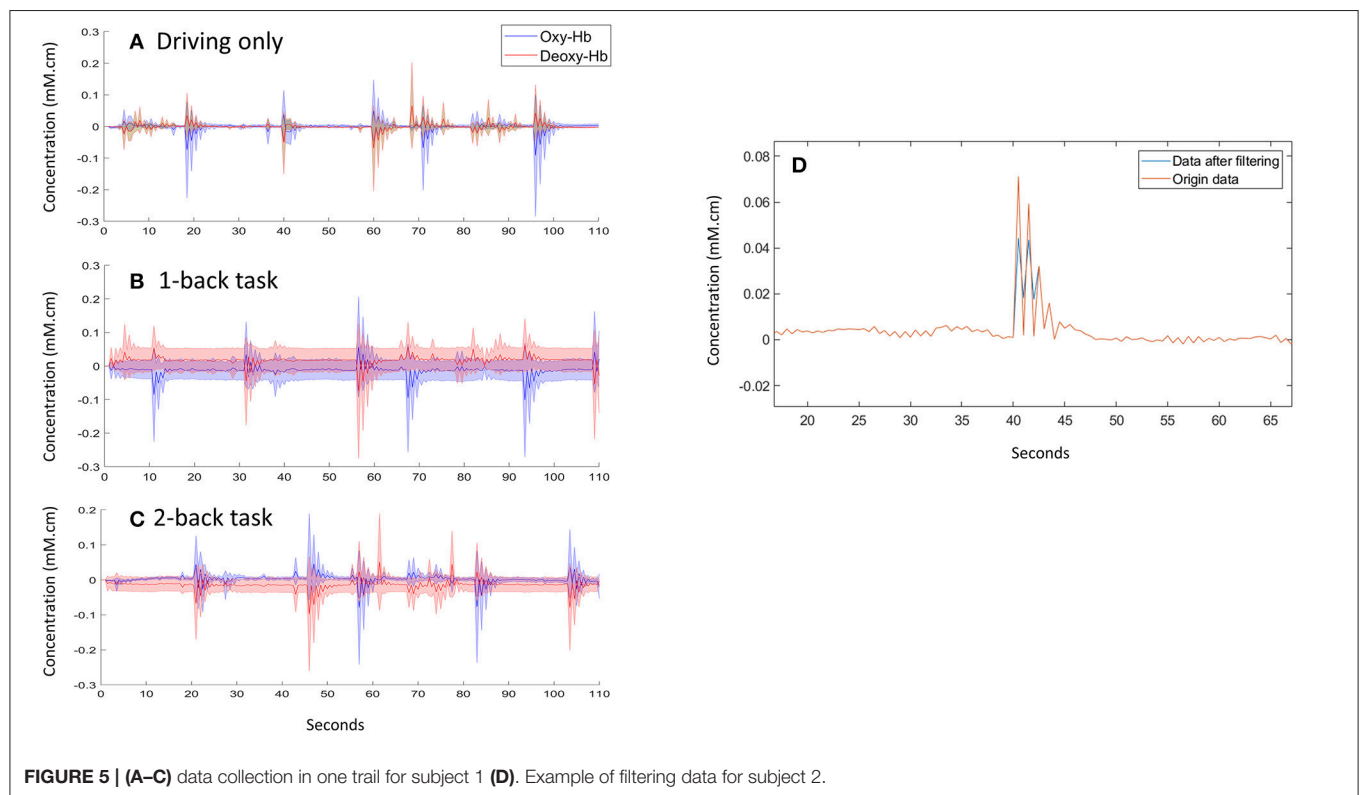
where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

In addition, in case of applying SVMs to perform multiclass classification (just driving vs. 1-back vs. 2-back), the transformation technique was applied to reduce the multiclass classification problem to a set of binary classification subproblems, with one SVM learner for each subproblem. One-vs.-All trains one learner for each class. It learns to distinguish one class from all others will be applied in our case.

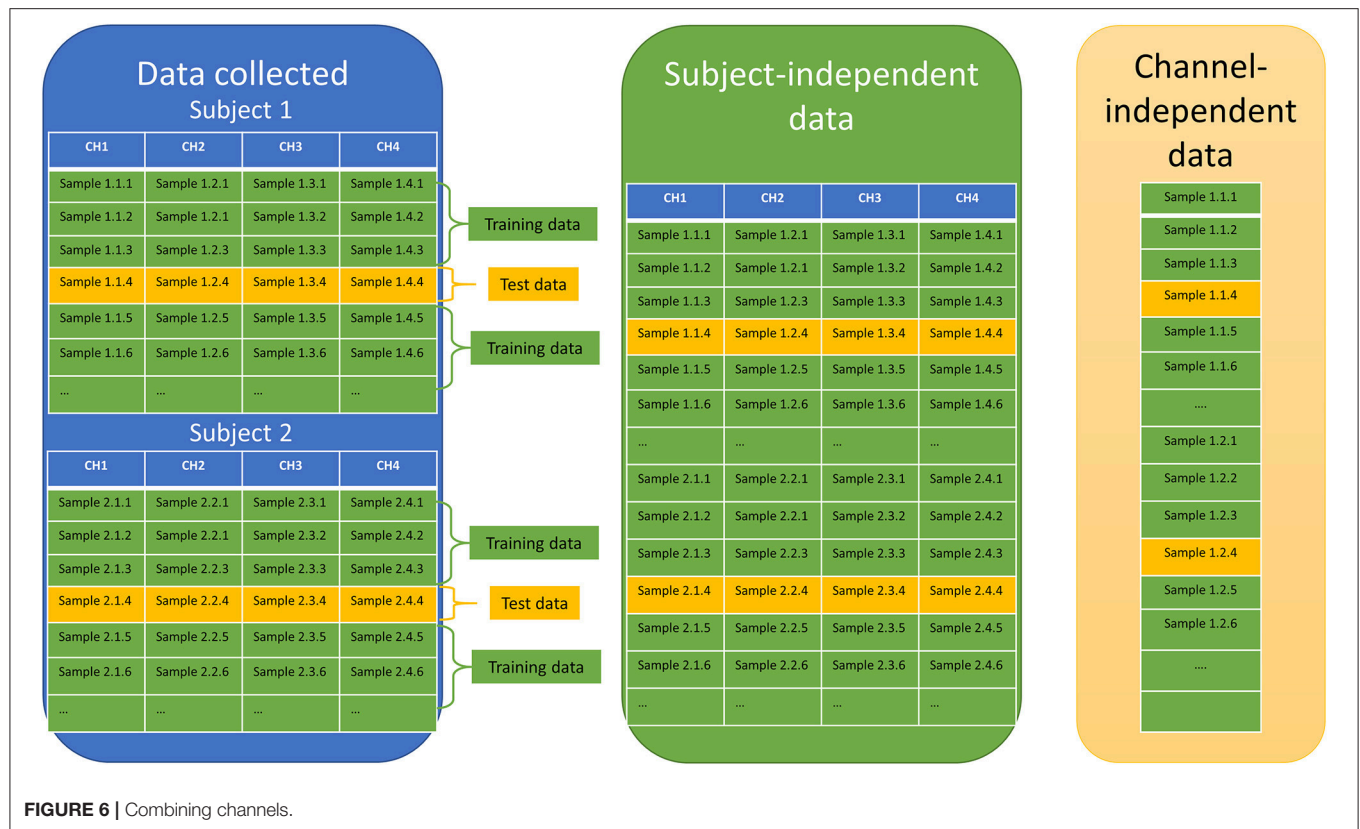


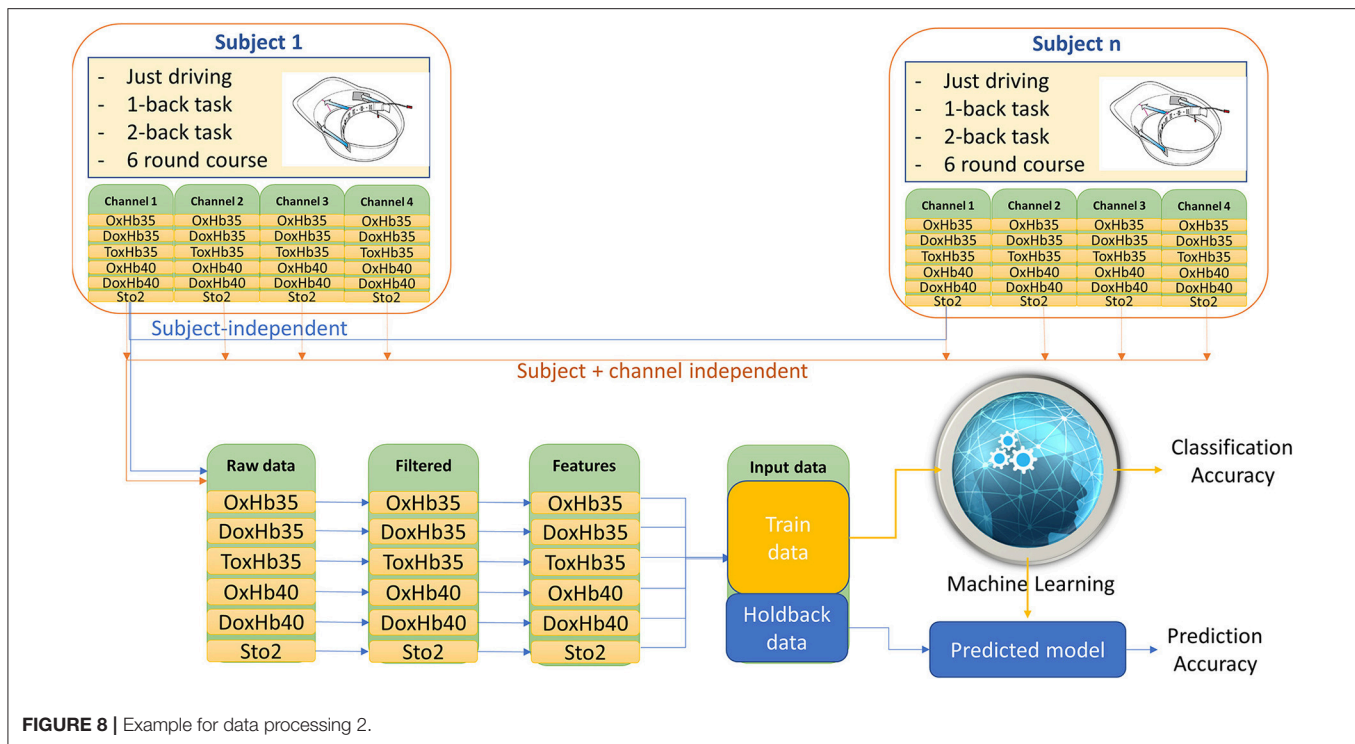


**FIGURE 4 |** Pre-processing data.



**FIGURE 5 | (A–C)** data collection in one trail for subject 1 **(D)**. Example of filtering data for subject 2.





The most accurate model for classification was selected and used in the prediction step with the testing data. All parameters of any model were kept the same in both classification and prediction step.

Definition (**Figure 9**):

- Classification accuracy: the accuracy of the classification using 75% of the data for train step. The model of highest accuracy was selected to use in the prediction step.
- Accuracy on the testing data: the accuracy of prediction using 25% of the data for the prediction step. In this step, the prediction model was exported from the selected model in the classification step and apply it for the new data (25% testing data).

## RESULTS

### Subject-Dependent Classification Analysis

Data for each channel for each subject were separated for the purposes of training and prediction. The classification between driving only, driving with a 1-back task, and driving with a 2-back task showed a good performance and a high accuracy (the classification accuracy increased from 81.30 to 95.40%, and the accuracy for the testing data increased from 82.18 to 96.08%) (**Video 1**). Details of the accuracy are shown in **Table 1**.

### Subject-Independent Classification Analysis

With the main aim of comparing the effects of individual characteristics on the accuracy of classification, we also performed a classification with the data for each channel for all subjects. The results are shown in **Figure 10**. The accuracies

in classifying the driver's mental workload from each channel were found to be in the range 84.9 to 89.5%, and the accuracy in predicting testing data increased from 84.08 to 89.91%. These results indicated that individual characteristics affected the accuracy of classification of the mental workload.

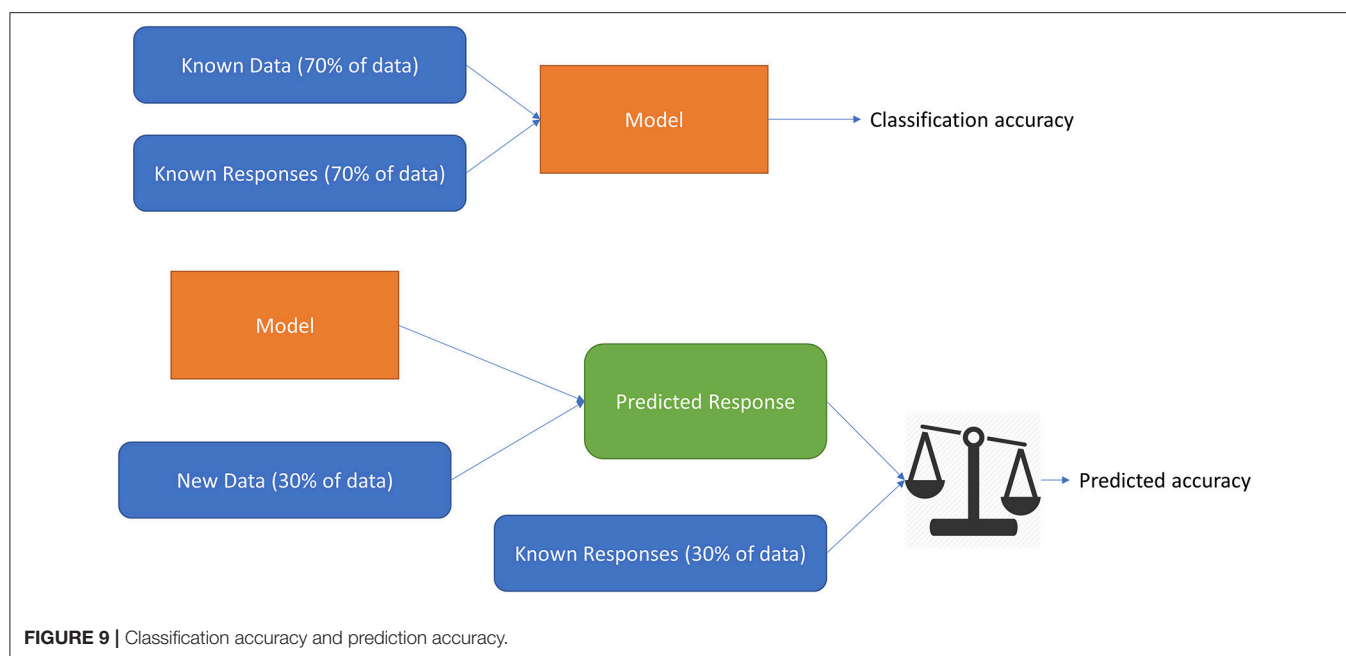
### Channel-Independent Classification Analysis

Before combining the data from all channels from the NIRS together, we performed a multiple comparison of oxy-HB and deoxy-HB levels for all the channel data from each subject in the same task by means of a Bonferroni test. The results showed that there was no significant difference between the data from the various channels in the same task (the  $p$ -value in all cases was  $>0.05$ ). According to the result, we decided to combine the data from all the channels together when checking the accuracy of classification.

First, the data from the four channels for each subject were combined to test the accuracy of classification. The results of this classification are shown in **Figure 11**. The accuracy of classification increased from 80.8 to 88.6%, and the accuracy on the testing data increased from 83.4 to 88.2%. This shows that acceptably accurate results of classification can be obtained simply by combining the data for the various channels.

### Subject-Independent + Channel-Independent Classification Analysis

Finally, to examine the potential for creating a system real-time classification of driver cognitive load to prevent accidents, we



**TABLE 1 |** The classification accuracy (the accuracy on the testing data) (%).

|           | Subject 1   | Subject 2   | Subject 3   | Subject 4   | Subject 5   |
|-----------|-------------|-------------|-------------|-------------|-------------|
| Channel 1 | 94.0 (93.3) | 95.4 (96.1) | 91.4 (91.3) | 92.1 (91.3) | 89.7 (91.1) |
| Channel 2 | 88.9 (87.9) | 92.1 (88.3) | 89.7 (89.9) | 92.6 (92.3) | 87.3 (91.0) |
| Channel 3 | 92.9 (90.4) | 89.8 (90.6) | 91.9 (88.7) | 85.2 (82.3) | 83.0 (86.4) |
| Channel 4 | 94.4 (92.5) | 89.9 (90.0) | 87.8 (89.5) | 85.7 (85.4) | 81.3 (82.2) |

combined the data from all the channels and all the subjects and we used the combined data to train the classification. We then examined the effect of this combination on the accuracy of classification and on the accuracy of prediction. As expected, the accuracy of classification was 82.9% and the accuracy of prediction was 82.71%, which were similar values to those previously obtained. This showed that the position of the channel on the forehead did not have a significant effect on the accuracy, and it confirmed that a compact NIRS device can capture the cognitive distraction of a driver, even under naturalistic conditions.

Compare with the result done by Naseer and Hong (2015) and Hong et al. (2015), which was used fNIRS signal, and then show the possibility of the hybrid feature extraction to classification with motor imagery tasks. The highest classification accuracy was around 77.5% with multi-class LDA. Furthermore, the classification of the right—and left—wrist motor imageries also done by Naseer and Hong (2013) using fNIRS information. By reducing the time span within the task period to 2–7 s, the accuracy for classification was increased to 77.56 and 87.28%. Here, we performed the classification with a machine learning algorithm and get better results with accuracy around 82.9% (in the case of subject-independent + channel-independent). The different of the accuracy may depend on the difference in mental workload level, different experiment condition, and so on.

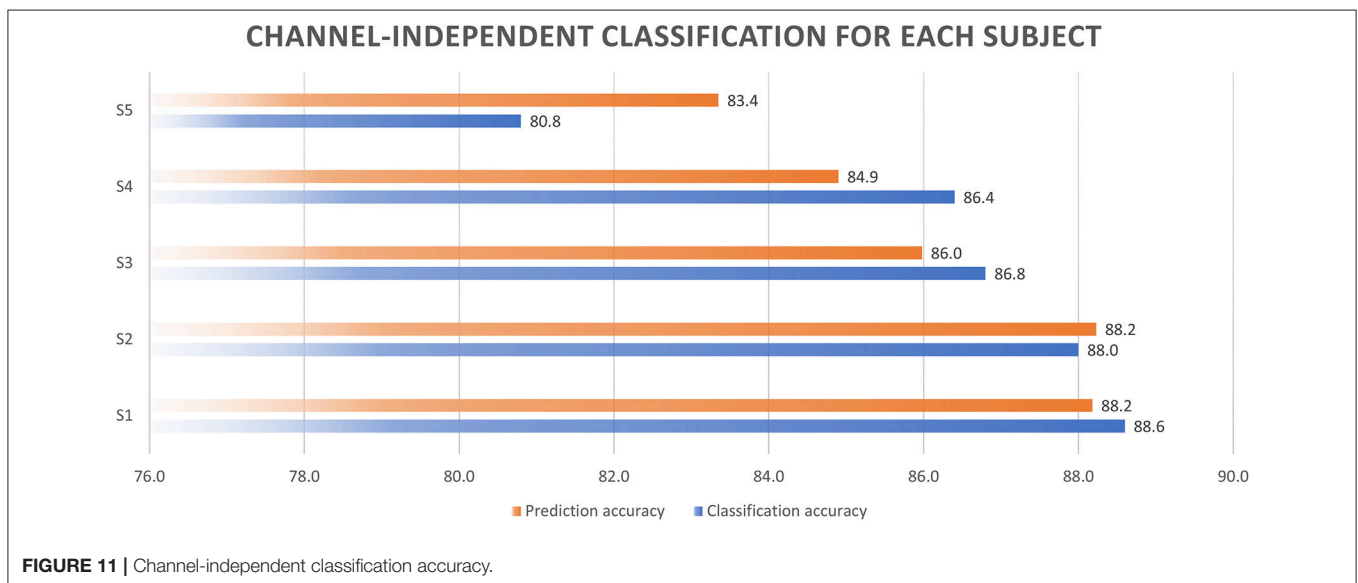
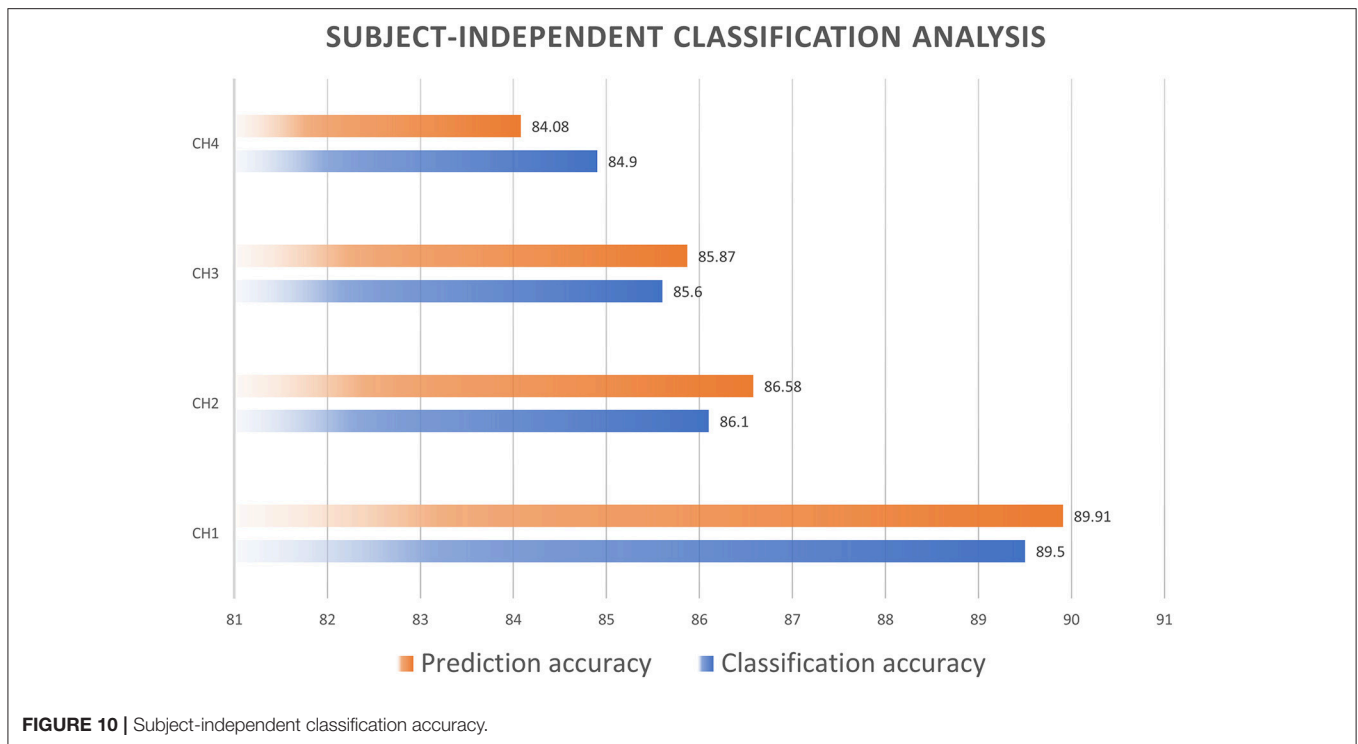
## DISCUSSION

### Model Selection for Classification of NIRS Data

As we have mentioned above, most previous researchers have used an SVM (Devos et al., 2009; Ichikawa et al., 2014; Aghajani et al., 2017), linear discriminant analysis (Luu and Chau, 2009), the hidden Markov model (Sitaram et al., 2007; Zimmermann et al., 2013), or artificial neural networks (Chan et al., 2012; Thanh Hai et al., 2013) to classify mental workload. In this study, we trained the data with some new models, such as the k-nearest neighbors model (k-NN) and the bagged tree (random forests) model, depending on the number of samples. Our results showed that the random forests model provided the highest accuracy, even with large numbers of samples, whereas the cubic SVM showed the worst performance (The average accuracies of each model in all previous analyses are shown in Figure 12). In addition, the k-NN model is also suitable for classifying mental workloads by using NIRS data because of its ability to maintain similar levels of accuracy even when the sample size changes markedly.

The SVM, a well-known method that has been previously applied in various classifications of NIRS data, showed very good performance with small numbers of samples. However, when there were large numbers of samples, the SVM was very slow and its accuracy was low compared with other methods. For example, in the case of a channel-independent test for Subject 1, where the sample size was over 15,000 samples, the accuracy of the SVM was 68.2%, compared with 87.3% for the random forests method and 88.6% for the k-NN classifier. In addition, the SVM took 1,341 s to perform the classification, whereas the k-NN classifier required only 88.6 s. Similar effects were observed in subject-dependent and subject-independent classifications. The sudden reduction in the accuracy of the SVM might arise from





differences in the data after data combination, as well as from effects of individual characteristics.

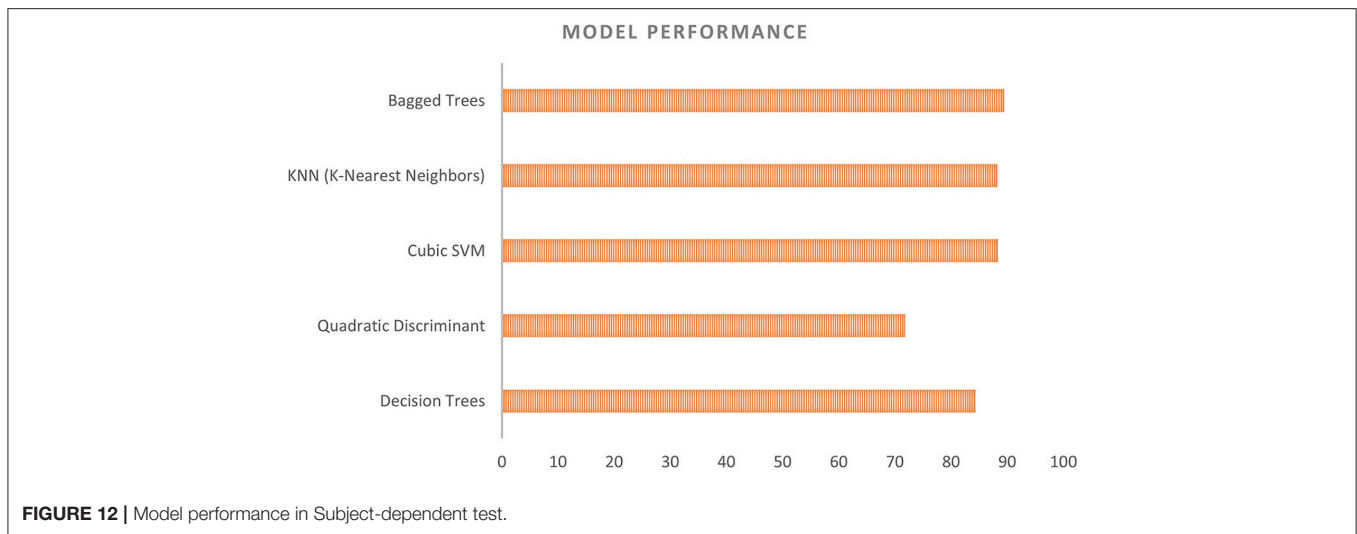
On the other hand, the way to select testing sample also plays a very importance role in the classification. In our case, we selected the testing sample following  $X \rightarrow X \rightarrow X \rightarrow Y \rightarrow X \rightarrow X \rightarrow Y \rightarrow X \dots$ , where  $X$  is a sample taken for the 75% and  $Y$  is a testing sample. It may make the nearest neighbor classifier that will perform very well, probably because the variation from  $Y$  to its neighbors in time (the  $X$  before and after  $Y$ ) will be very low.

We also believe that for higher numbers of subjects, a lower accuracy is attained for subject-independent classification.

Consequently, for large numbers of subjects, the machine-learning algorithm should be changed to a deep learning or convolution neural-network algorithm, which can still show good performance with large quantities of data.

### The Potential For Using NIRS Data to Evaluate Levels of Driver Mental Workload

This study is one example of the application of machine learning in classifying driver mental workload from data obtained with a simple commercial NIRS device, which has a high potential



for routine use with drivers because of its acceptable price. However, we believe that the use of a combination of channels for NIRS is necessary because signal losses tended to occur often under naturalistic conditions. In some cases, however, one or two channels provided sufficient signals due to the activities of the driver.

The ability to unobtrusively detect changes in mental workload is relevant since high levels of cognitive load can reduce a driver's ability to anticipate and respond to emergent dangers in the driving environment. Broadly considered, these findings suggest various lines of potential research related to the development of advanced driver assistance systems (e.g., a new method to prevent accidents by detecting levels of mental workload that may lead to cognitive distraction), basic human factors insight (exploring the relationship between individual characteristics and objective indicators of mental workload), and mathematical modeling (combining channel, improve accuracy by applying different technical).

In conclusion, as previously suggested (Kopton and Kenning, 2014; Unni et al., 2017), simple NIRS has considerable potential for capturing driver mental workload, especially under naturalistic conditions.

## LIMITATIONS

The relatively small sample size used in this study (a total of 5 subjects including one female and four males) could be considered a limitation. While we believe that the NIRS signals were found to be predictive for this small sample under our specific set of conditions, it would be worthwhile to repeat the experiment with a larger sample and a wider range of conditions (e.g., driving track, time of day, gender balance, driver skill level, age, etc.).

## CONCLUSIONS

Our study suggested that it is possible to use NIRS data to classify levels of driver mental workload, even in a naturalistic situation. Furthermore, a simple combination of forehead channels was shown to provide acceptably high accuracies of classification. While the fNIRS sensors employed in this study required contact with the participants' skin, the lightweight ball cap configuration was much less intrusive than more traditional electrophysiological measures used in related work. We also confirmed the potential of using machine learning (channel- and subject-independent) to predict possible driver cognitive distraction, a critical factor in road safety.

## AUTHOR CONTRIBUTIONS

AL, HA, FM, and KI generated the idea. AL and HA planned the experiments, collected and analyzed the data, and prepared the manuscript for this study.

## ACKNOWLEDGMENTS

This study was partially supported by DENSO CORPORATION and by the Center of Innovation Program (Nagoya University COI: Mobility Innovation Center) from Japan Science and Technology Agency.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2018.00431/full#supplementary-material>

**Video 1 |** NIRS toolbox for classifying driver mental workload.

## REFERENCES

- Abdelnour, A. F., and Huppert, T. (2009). Real-time imaging of human brain function by near-infrared spectroscopy using an adaptive general linear model. *Neuroimage* 46, 133–143. doi: 10.1016/j.neuroimage.2009.01.033
- Aghajani, H., Garbey, M., and Omurtag, A. (2017). Measuring mental workload with EEG+fNIRS. *Front. Hum. Neurosci.* 11:359. doi: 10.3389/fnhum.2017.00359
- Angell, L., Auflick, J., Austria, P. A., Kochhar, D., Tijerina, L., Biever, W., et al. (2006). *Driver Workload Metrics Project*. Task 2 final Report. Security, 460.
- Angell, L., and Perez, M. (2015). *Identification of Cognitive Load in Naturalistic Driving*. Blacksburg, VA: Virginia Tech Transportation Institute.
- Anh Son, L., Aoki, H., and Suzuki, T. (2017a). Evaluation of driver distraction with changes in gaze direction based on a vestibulo-ocular reflex model. *J. Transp. Technol.* 7, 336–350. doi: 10.4236/jtts.2017.73022
- Anh Son, L., Hamada, H., Inagami, M., Suzuki, T., and Aoki, H. (2017d). “Effect of mental workload and aging on driver distraction based on the involuntary eye movement,” in *Advances in Human Aspects of Transportation, Vol. 484, Advances in Intelligent Systems and Computing*, eds N. Stanton, S. Landry, G. Di Bucchianico, and A. Vallicelli (Cham: Springer), 349–359. doi: 10.1007/978-3-319-41682-3\_30
- Anh Son, L., Hiroto, H., Tatsuya, S., and Hirofumi, A. (2017b). “Driver distraction evaluation using reflex eye movement simulation,” in *Proceedings of the Future Active Safety Technology Towards Zero Traffic Accidents Conference* (Nara).
- Anh Son, L., Inagami, M., Hiroto, H., Tatsuya, S., and Hirofumi, A. (2018). Towards online detection of driver distraction: Eye-movement simulation based on a combination of vestibulo-ocular reflex and optokinetic reflex models. *Transp. Res. Part F Traffic Psychol. Behav.* doi: 10.1016/j.trf.2018.02.019
- Anh Son, L., Inagami, M., Suzuki, T., Aoki, H., and Obinata, G. (2016). “Driver’s internal state estimation based on the eye movement,” in *Proceedings of the 2016 International Conference University Village* (Nagoya).
- Anh Son, L., Makoto, I., Hiroto, H., Tatsuya, S., and Hirofumi, A. (2017c). “Evaluation of cognitive distraction in a real vehicle based on the reflex eye movement,” in *Proceedings of the 2017 Japan Society Automotive Engineers Annual Congress* (Yokohama).
- Anh Son, L., Makoto, I., Hiroto, H., Tatsuya, S., and Hirofumi, A. (2017e). The effect of visual stimulus on voluntary eye movement based on a VOR / OKR Model. *Int. J. Automot. Eng.* 8, 37–44. doi: 10.20485/jsaeijae.8.2\_37
- Aoki, H., Anh Son, L., Hamada, H., Suzuki, T., and Obinata, G. (2015). “Quantitative evaluation of mental workload with voluntary eye movements by means of the vestibulo-ocular reflex,” in *Proceedings of the 2015 Japan Society Automotive Engineers Annual Congress* (Sapporo), 780–785.
- Backs, R. W., and Walrath, L. C. (1992). Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. *Appl. Ergon.* 23, 243–254.
- Benedetto, S., Pedrotti, M., Minin, L., Baccino, T., Re, A., and Montanari, R. (2011). Driver workload and eye blink duration. *Transp. Res. Part F Traffic Psychol. Behav.* 14, 199–208. doi: 10.1016/j.trf.2010.12.001
- Chan, J., Power, S., and Chau, T. (2012). Investigating the need for modelling temporal dependencies in a brain-computer interface with real-time feedback based on near infrared spectra. *J. Near Infrared Spectrosc.* 20, 107–116. doi: 10.1255/jnirs.971
- Cinaz, B., Arnrich, B., La Marca, R., and Tröster, G. (2013). Monitoring of mental workload levels during an everyday life office-work scenario. *Pers. Ubiquitous Comput.* 17, 229–239. doi: 10.1007/s00779-011-0466-1
- Cooper, R. J., Selb, J., Gagnon, L., Phillip, D., Schytz, H. W., Iversen, H. K., et al. (2012). A systematic comparison of motion artifact correction techniques for functional near-infrared spectroscopy. *Front. Neurosci.* 6:147. doi: 10.3389/fnins.2012.00147
- Devos, O., Ruckebusch, C., Durand, A., Duponchel, L., and Huvenne, J. P. (2009). Support vector machines (SVM) in near infrared (NIR) spectroscopy: Focus on parameters optimization and model interpretation. *Chemom. Intell. Lab. Syst.* 96, 27–33. doi: 10.1016/j.chemolab.2008.11.005
- Dong, Y., Hu, Z., Uchimura, K., and Murayama, N. (2011). Driver inattention monitoring system for intelligent vehicles: a review. *IEEE Trans. Intell. Transp. Syst.* 12, 596–614. doi: 10.1109/TITS.2010.2092770
- Engström, J., Johansson, E., and Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transp. Res. Part F Traffic Psychol. Behav.* 8, 97–120. doi: 10.1016/j.trf.2005.04.012
- Henelius, A., Hirvonen, K., Holm, A., Korpela, J., and Muller, K. (2009). Mental workload classification using heart rate metrics. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2009, 1836–1839. doi: 10.1109/IEMBS.2009.5332602
- Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., and Schultz, T. (2014). Mental workload during n-back task—quantified in the prefrontal cortex using fNIRS. *Front. Hum. Neurosci.* 7:935. doi: 10.3389/fnhum.2013.00935
- Hong, K. S., Naseer, N., and Kim, Y. H. (2015). Classification of prefrontal and motor cortex signals for three-class fNIRS-BCI. *Neurosci. Lett.* 587, 87–92. doi: 10.1016/j.neulet.2014.12.029
- Huppert, T. J., Diamond, S. G., Franceschini, M. A., and Boas, D. A. (2016). HomER: a review of time-series analysis methods for nearinfrared spectroscopy of the brain. *Appl. Opt.* 48, 280–298. doi: 10.1016/j.drugalcdep.2008.02.002.A
- Ichikawa, H., Kitazono, J., Nagata, K., Manda, A., Shimamura, K., Sakuta, R., et al. (2014). Novel method to classify hemodynamic response obtained using multi-channel fNIRS measurements into two groups: exploring the combinations of channels. *Front. Hum. Neurosci.* 8:480. doi: 10.3389/fnhum.2014.00480
- Jorna, P. G. (1992). Spectral analysis of heart rate and psychological state: a review of its validity as a workload index. *Biol. Psychol.* 34, 237–257. doi: 10.1016/0301-0511(92)90017-O
- Kirilina, E., Yu, N., Jelzow, A., Wabnitz, H., Jacobs, A. M., and Tachtsidis, I. (2013). Identifying and quantifying main components of physiological noise in functional near infrared spectroscopy on the prefrontal cortex. *Front. Hum. Neurosci.* 7:864. doi: 10.3389/fnhum.2013.00864
- Klingner, J. (2010). “Fixation-aligned pupillary response averaging,” in *Proceedings of the 2010 Symposium Eye-Tracking Research & Applications* (Austin, TX), 275–282.
- Klingner, J., Kumar, R., and Hanrahan, P. (2008). “Measuring the task-evoked pupillary response with a remote eye tracker,” in *Proceedings of the 2008 Symposium Eye Tracking Research & Applications-ETRA’08* 1, 69 (Savannah, GA).
- Kohri, S., Hoshi, Y., Tamura, M., Kato, C., Kuge, Y., and Tamaki, N. (2002). Quantitative evaluation of the relative contribution ratio of cerebral tissue to near-infrared signals in the adult human head: a preliminary study. *Physiol. Meas.* 23, 301–312. doi: 10.1088/0967-3334/23/2/306
- Kopton, I. M., and Kenning, P. (2014). Near-infrared spectroscopy (NIRS) as a new tool for neuroeconomic research. *Front. Hum. Neurosci.* 8:549. doi: 10.3389/fnhum.2014.00549
- Le, A. S., and Aoki, H. (2018). Smart eye and polhemus data for vestibulo-ocular reflex and optokinetic reflex model. *Data Br.* 18, 882–885. doi: 10.1016/j.dib.2018.03.097
- Lee, D. H., and Park, K. S. (1990). Multivariate analysis of mental and physical load components in sinus arrhythmia scores. *Ergonomics* 33, 35–47. doi: 10.1080/00140139008927092
- Liu, T., Pelowski, M., Pang, C., Zhou, Y., and Cai, J. (2016). Near-infrared spectroscopy as a tool for driving research. *Ergonomics* 59, 368–379. doi: 10.1080/00140139.2015.1076057
- Liu, Z., Zhang, M., Xu, G., Huo, C., Tan, Q., Li, Z., et al. (2017). Effective connectivity analysis of the brain network in drivers during actual driving using near-infrared spectroscopy. *Front. Behav. Neurosci.* 11:11. doi: 10.3389/fnbeh.2017.00211
- Luu, S., and Chau, T. (2009). Decoding subjective preference from single-trial near-infrared spectroscopy signals. *J. Neural Eng.* 6:016003. doi: 10.1088/1741-2560/6/1/016003
- McBride, D. K., and Schmorow, D. (eds.). (2005). *Quantifying Human Information Processing*. Lanham, MA: Lexington Books.
- 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. *Hum. Factors* 54, 396–412. doi: 10.1177/0018720812442086
- Meshkati, N. (1988). Heart rate variability and mental workload assessment. *Hum. Ment. Workload* 52, 101–115. doi: 10.1016/S0166-4115(08)62384-5
- Naseer, N., and Hong, K.-S. (2015). fNIRS-based brain-computer interfaces: a review. *Front. Hum. Neurosci.* 9:3. doi: 10.3389/fnhum.2015.00003

- Naseer, N., and Hong, K. S. (2013). Classification of functional near-infrared spectroscopy signals corresponding to the right- and left-wrist motor imagery for development of a brain-computer interface. *Neurosci. Lett.* 553, 84–89. doi: 10.1016/j.neulet.2013.08.021
- NHTSA (2012). Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices. *Dep. Transp.* 77, 1–34.
- NHTSA (2015). *Research Note Distracted Driving 2013*. Dot Hs 812 132, 1–8. Available online at: <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812381>.
- Obinata, G., Fukuda, K., Moriyama, A., Tokuda, S., and Kim, Y. W. (2010). Evaluating the influence of distractions to drivers based on reflex eye movement model. *IFAC* doi: 10.3182/20100831-4-FR-2021.00045
- Obinata, G., Tokuda, S., Fukuda, K., and Hamada, H. (2009). “Quantitative evaluation of mental workload by using model of involuntary eye movement,” in *Engineering Psychology and Cognitive Ergonomics SE - 24 Lecture Notes in Computer Science*, ed D. Harris (Berlin; Heidelberg: Springer), 223–232. doi: 10.1007/978-3-642-02728-4\_24
- Obinata, G., Usui, T., and Shibata, N. (2008). On-line method for evaluating driver distraction of memory-decision workload based on dynamics of vestibulo-ocular reflex. *Rev. Automot. Eng.* 29, 627–632. doi: 10.11351/jsaereview.29.627
- Orino, Y., Yoshino, K., Oka, N., Yamamoto, K., Takahashi, H., and Kato, T. (2015). Brain activity involved in vehicle velocity changes in a sag vertical curve on an expressway: vector-based functional near-infrared spectroscopy study. *J. Transport. Res. Board* 2518, 18–26. doi: 10.3141/2518-03
- Palinko, O., and Kun, A. (2012). “Exploring the influence of light and cognitive load on pupil diameter in driving simulator studies,” in *Proceedings Symposium Eye Tracking Research & Application*, 413–416.
- Pierce, E. T. (2009). *Mental Workload Measurement Using the Intersaccadic Interval*. ProQuest Diss. Theses, 160. Available online at: [http://ezproxy.net.ucf.edu/login?url=http://search.proquest.com/docview/858993124?accountid=10003%5Chttp://sfx.fcla.edu/ucf?url\\_ver=Z39.88-2004&rft\\_val\\_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+\\$&\\$&\\$theses&sid=ProQ:ProQuest+\\$Dissertations+\\$&\\$&\\$T](http://ezproxy.net.ucf.edu/login?url=http://search.proquest.com/docview/858993124?accountid=10003%5Chttp://sfx.fcla.edu/ucf?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+$&$&$theses&sid=ProQ:ProQuest+$Dissertations+$&$&$T).
- Porges, S. W., and Byrne, E. A. (1992). Research methods for measurement of heart rate and respiration. *Biol. Psychol.* 34, 93–130. doi: 10.1016/0301-0511(92)90012-J
- Rizzo, M., and Hurtig, R. (1987). Looking but not seeing: attention, perception, and eye movements in simultanagnosia. *Neurology* 37, 1642–1648.
- Ryu, K., and Myung, R. (2005). Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. *Int. J. Ind. Ergon.* 35, 991–1009. doi: 10.1016/j.ergon.2005.04.005
- Schwalm, M., Keinath, A., and Zimmer, H. D. (2008). Pupillometry as a method for measuring mental workload within a simulated driving task. *Hum. Factors Assist. Autom.* 1–13.
- Sitaram, R., Zhang, H., Guan, C., Thulasidas, M., Hoshi, Y., Ishikawa, A., et al. (2007). Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain-computer interface. *Neuroimage* 34, 1416–1427. doi: 10.1016/j.neuroimage.2006.11.005
- Son, L. A., and Hirofumi, A. (2018). *Evaluating Driver Mental Workload in a Realistic Condition Using Eye Movement Sensor*. Nagoya: Tokai Engineering Complex. doi: 10.1299/jsmetokai.2018.67.408
- Son, L. A., Suzuki, T., and Aoki, H. (2018). Evaluation of Cognitive Distraction in a Real Vehicle Based on the Reflex Eye Movement. *Int. J. Automot. Eng.* 9, 1–8. doi: 10.20485/jsaiejae.9.1\_1
- Thanh Hai, N., Cuong, N. Q., Dang Khoa, T. Q., and Van Toi, V. (2013). Temporal hemodynamic classification of two hands tapping using functional near-infrared spectroscopy. *Front. Hum. Neurosci.* 7:516. doi: 10.3389/fnhum.2013.00516
- Tokuda, S., Palmer, E., Merkle, E., and Chaparro, A. (2009). Using saccadic intrusions to quantify mental workload. *Hum. Factors Ergon. Soc. Annu. Meet. Proc.* 53, 809–813. doi: 10.1518/107118109X12524442637660
- Tsai, Y. F., Viirre, E., Strychacz, C., Chase, B., and Jung, T.-P. (2007). Task performance and eye activity: predicting behavior relating to cognitive workload. *Aviat. Space. Environ. Med.* 78, 176–185.
- Tsunashima, H., and Yanagisawa, K. (2009). Measurement of brain function of car driver using functional near-infrared spectroscopy (fNIRS). *Comput. Intell. Neurosci.* 2009:164958. doi: 10.1155/2009/164958
- Unni, A., Ihme, K., Jipp, M., and Rieger, J. W. (2017). Assessing the driver's current level of working memory load with high density functional near-infrared spectroscopy: a realistic driving simulator study. *Front. Hum. Neurosci.* 11:167. doi: 10.3389/fnhum.2017.00167
- Veltman, J. A., and Gaillard, A. W. (1996). Physiological indices of workload in a simulated flight task. *Biol. Psychol.* 42, 323–342. doi: 10.1016/0301-0511(95)05165-1
- Wang, Y., Reimer, B., Dobres, J., and Mehler, B. (2014). The sensitivity of different methodologies for characterizing drivers' gaze concentration under increased cognitive demand. *Transport. Res. Part F* 26, 227–237. doi: 10.1016/j.trf.2014.08.003
- Yoshino, K., Oka, N., Yamamoto, K., Takahashi, H., and Kato, T. (2013). Functional brain imaging using near-infrared spectroscopy during actual driving on an expressway. *Front. Hum. Neurosci.* 7:882. doi: 10.3389/fnhum.2013.00882
- Yoshino, K., Oka, N., Yamamoto, K., Takahashi, H., and Kato, T. (2015). Correlation of prefrontal cortical activation with changing vehicle speeds in actual driving: a vector-based functional near-infrared spectroscopy study. *Front. Hum. Neurosci.* 9:665. doi: 10.3389/fnhum.2015.00665
- Zimmermann, R., Marchal-Crespo, L., Edelmann, J., Lamberg, O., Fluet, M. C., Riener, R., et al. (2013). Detection of motor execution using a hybrid fNIRS-biosignal BCI: A feasibility study. *J. Neuroeng. Rehabil.* 10:4. doi: 10.1186/1743-0003-10-4

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Le, Aoki, Murase and Ishida. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Brain Network Changes in Fatigued Drivers: A Longitudinal Study in a Real-World Environment Based on the Effective Connectivity Analysis and Actigraphy Data

André Fonseca<sup>1,2\*</sup>, Scott Kerick<sup>3</sup>, Jung-Tai King<sup>4</sup>, Chin-Teng Lin<sup>5</sup> and Tzyy-Ping Jung<sup>2</sup>

<sup>1</sup> Center of Mathematics, Computation and Cognition, Federal University of ABC, São Paulo, Brazil, <sup>2</sup> Swartz Center for Computational Neuroscience, University of California, San Diego, La Jolla, CA, United States, <sup>3</sup> US Army Research Laboratory, Aberdeen, MD, United States, <sup>4</sup> Brain Research Center, National Chiao Tung University, Hsinchu, Taiwan, <sup>5</sup> Faculty of Engineering and Information Technology, Centre for Artificial Intelligence, University of Technology Sydney, Sydney, NSW, Australia

## OPEN ACCESS

### Edited by:

Gianluca Borghini,  
Università degli Studi di Roma La  
Sapienza, Italy

### Reviewed by:

Dionissios Hristopoulos,  
Technical University of Crete, Greece  
Yuri Antonacci,  
Università degli Studi di Roma La  
Sapienza, Italy

### \*Correspondence:

André Fonseca  
andre.fonseca@ufabc.edu.br

**Received:** 21 June 2018

**Accepted:** 27 September 2018

**Published:** 12 November 2018

### Citation:

Fonseca A, Kerick S, King J-T, Lin C-T and Jung T-P (2018) Brain Network Changes in Fatigued Drivers: A Longitudinal Study in a Real-World Environment Based on the Effective Connectivity Analysis and Actigraphy Data. *Front. Hum. Neurosci.* 12:418. doi: 10.3389/fnhum.2018.00418

The analysis of neurophysiological changes during driving can clarify the mechanisms of fatigue, considered an important cause of vehicle accidents. The fluctuations in alertness can be investigated as changes in the brain network connections, reflected in the direction and magnitude of the information transferred. Those changes are induced not only by the time on task but also by the quality of sleep. In an unprecedented 5-month longitudinal study, daily sampling actigraphy and EEG data were collected during a sustained-attention driving task within a near-real-world environment. Using a performance index associated with the subjects' reaction times and a predictive score related to the sleep quality, we identify fatigue levels in drivers and investigate the shifts in their effective connectivity in different frequency bands, through the analysis of the dynamical coupling between brain areas. Study results support the hypothesis that combining EEG, behavioral and actigraphy data can reveal new features of the decline in alertness. In addition, the use of directed measures such as the Convergent Cross Mapping can contribute to the development of fatigue countermeasure devices.

**Keywords:** drivers, fatigue, sleep, actigraphy, EEG, effective connectivity, Convergent Cross Mapping

## 1. INTRODUCTION

Fatigue is a complex, dynamic, multidimensional construct involving subjective, behavioral, neural, and physiological processes that interact over varying timescales across a milieu of tasks and environmental contexts, making it difficult to operationally define and measure in a consistent or unitary way for scientific investigation. This study considers two different sources of fatigue operating on different timescales that interact in complex ways and vary both across individuals and within individuals over time. The first source of fatigue (or sleepiness) is related to circadian rhythms or sleep-wake cycles (sleep-related, e.g., acute or chronic sleep deprivation leading to sleep pressure) and the second source is related to the nature, complexity, and duration of the current task one is performing (task-related, e.g., task difficulty or demand, time-on-task which may lead to one's disinclination to continue performing a particular task). The importance of

distinguishing between sleep- and task-related fatigue is that they reflect two conceptually distinct and separable sources of potential variations in performance and underlying brain mechanisms and require different mitigation strategies (May and Baldwin, 2009; Balkin and Wesensten, 2011). However, these underlying processes may interact in complex ways. Fatigue may lead to the decline of cognitive functioning and lapses in attention. It has cumulative and persistent effects in daytime performance (Belenky et al., 2003) and is considered a major factor in traffic accidents caused by human errors (Inoue and Komada, 2014).

Fatigue diminishes road safety, accounting for approximately 25% of car accidents (Brown, 1994) and 57% of commercial truck accidents (Bonnet and Arand, 1995). Young people around 20 years old are particularly vulnerable to fatigue-related accidents (Pack et al., 1995). Generally speaking, fatigue is also associated with increased stress and impaired cognitive performance at work (Härmä et al., 2006). The effects of fatigue can vary over various timescales depending on task and context, but are generally classified as acute (sudden onset, relieved by rest) or chronic (persistent, lasting from days to years) which vary from poor accomplishments to health and security problems (Spurgeon et al., 1997).

Understanding antecedents and consequences of fatigue and having a capability to predict fatigue-related performance decrements is a matter of public safety and wellness. When there is a risk of error or accident, the individual alertness and cognitive performance can be measured and the attention lapses can be putative. Specifically about drivers, biomathematical models have been developed to associate fatigue levels with working patterns. For instance, the circadian information, which is linked to task performance (Harrison et al., 2007), can be recorded from activity and rest periods and then processed by those models to estimate sleep quality and to infer sleep-related fatigue. From several biomathematical approaches we choose the *Sleep, Activity, Fatigue, and Task Effectiveness* (SAFTE) (Hursh et al., 2004), which records data of circadian rhythm, homeostatic drive, and sleep inertia, to characterize the sleep-awake history of drivers. The results of SAFTE has been validated as a neurobehavioral performance predictor in laboratory and real-world environments (Dawson et al., 2011).

Another important resource to investigate and predict drivers' fatigue is the qualitative and quantitative EEG analysis, which has been used to unveil the relation between brain activity or brain network changes and the decline in alertness (Huang et al., 2015, 2016; Lin et al., 2016). The findings link behavioral performance with changes in EEG power spectrum and in the default mode network, suggesting that significant neural circuits must be activated to sustain performance and prevent attentional lapses. The investigation of those correlates is based on the brain connectivity theory and considering the alert-drowsiness transitions as an emergent effect of a complex system.

To analyze the underlying brain circuitry in the fatigue phenomenon, concepts of functional and effective connectivity can be applied. The first one refers to the statistical dependence in the neuronal activity and the second quantifies the influence that one brain area exerts over another (see Friston, 2011 and

Goldenberg and Galvn, 2015 for definitions and techniques). Those concepts allow different interpretations and can be complimentary (Friston et al., 2013).

Functional connectivity can be undirected as in correlation and coherence measures, or directed as in *Granger Causality* (GC) (Granger, 1969) and transfer entropy (Schreiber, 2000). Multivariate extensions of GC such as directed transfer function (Kaminski and Blinowska, 1991) and partial directed coherence (Baccalá and Sameshima, 2001) allow time-varying and frequency-selective analysis (see Barnett and Seth, 2014 for theoretical basis and numerical simulation of several brain connectivity estimators based on GC). Effective connectivity measures consider the directed integration in neuronal macrocircuits as in the dynamic causal modeling (Friston et al., 2003). The methodology choice relies on the assumptions of the underlying mechanism.

In our analysis, we considered dynamic emergent effects from coupling variables and the effective connectivity approach was selected. The study was performed using the *Convergent Cross Mapping* (CCM) (Sugihara et al., 2012). CCM quantifies the directed interactions considering non-linear and linear components, stationary and non-stationary features in bivariate or multivariate systems (McCracken and Weigel, 2014; Hirata et al., 2016; Jiang et al., 2016). CCM detects the causal relation strength and information exchanged between signals, assessing the synchronization features through the correspondence of the reconstructed phase-spaces, obtained from time-delay embedding coordinates. CCM has provided new insights into physiological states by considering the brain as a complex network system (McBride et al., 2015; Schiecke et al., 2017).

This work analyzed the brain network changes of drivers by the shifts in the effective connectivity expressed in the CCM oscillations. Moreover, this work investigated the modulation of the power spectra by those shifts. To assess possible CCM-power correlations, we first decomposed the EEG signals into different frequency bands prior to evaluating causal relations, providing information about effective connectivity changes for each neural rhythm. Using this procedure and the properties of dynamical coupling, it is plausible to assume that the CCM from the source signal to the target signal can modify dynamically phase and amplitude of the target observation. This principle is supported by fMRI studies such as Baechinger et al. (2017).

This methodology aims to detect changes in brain dynamics associated with the task-positive network of drivers to characterize alert and fatigue states during the simulated driving task. We combine EEG and non-EEG (subjective and behavioral data) recordings in the context of non-stationary data. For EEG signals, we choose to explore causal features in the reconstructed phase space considering the sources near the Frontal Midline and Parietal Midline brain areas. Our approach was based on the importance of dominating brain regions during driving to detect fluctuations in attention (Lin et al., 2016) and on the evidence of specific connectivity patterns in cortical regions related to behavioral microsleeps, a inherently non-stationary phenomena (Toppi et al., 2016).

The present study begins with a description of the subjects, the actigraphy data used to define levels of sleep-related fatigue

and of the realistic sustained-attention experiment, detailing the EEG data, and *Reaction Time* (RT) acquisitions. The information of those sources was combined to test the hypothesis that driving performance impairment in fatigued drivers is associated with effective connectivity shifts. Second, we define a *Driver Performance* (DP) index, explained the phase-space reconstruction procedure (needed for the CCM evaluation) and presented a simulation study to test CCM efficiency in a brain connectivity model. Next, we describe the statically analysis of DP, CCM, and power values over the sleep-related fatigue levels. Finally, we show the results and conclude our work with a discussion of the different brain network patterns detected in the sleep-related normal and fatigued levels.

## 2. MATERIALS AND METHODS

### 2.1. Subjects

Seventeen healthy university students, 13 males and 4 females, with normal or corrected to normal eyesight, no neurological, or psychiatric disorders, aged  $22.4 \pm 1.5$  years, all right-handed, from National Chiao Tung University (NCTU) in Taiwan participated in this study. The experimental protocol of the sustained-attention task was approved by the Institutional Review Board and written informed consent was obtained from each participant after a full explanation of the study.

### 2.2. Actigraphy Data Acquisition and Fatigue Level

As a part of a *Daily Sampling System* (DSS), the subjects used a wrist-worn device (Fatigue Science Readiband™), which records circadian, homeostatic and sleep inertia processes on a minute basis. This device incorporates the information collected in the last 3 days and, applying the biomathematical model SAFTE, provides a putative performance level called *Effectiveness Score* (ES), which can be easily read from the device. Based on this score, we classified the subjects into three levels of sleep-related fatigue, *Normal* (NO) for ES greater than 90%, *Reduced Risk* (RR) for ES between 70 and 90%, and *High Risk* (HR) for ES smaller than 70%. The HR level of sleep-related fatigue represents a putative performance comparable to subjects with 0.08 blood alcohol level or awake for 21 h. For more information about the SAFTE model and ES use/validation see Hursh et al. (2004), Hursh et al. (2006), and Russell et al. (2006).

### 2.3. Experimental Paradigm and Sessions

In this sustained-attention experiment we adapted the *Lane Keeping Task* (LKT) as the driving paradigm (Huang et al., 2016), where subjects must maintain the cruising position on the central lane and compensate randomly induced vehicle deviations by turning the steering wheel (see **Figure 1**). The experiment was conducted at the Brain research Center at NCTU using a realistic driving simulator (Chuang et al., 2012). The ES of the subjects (reflecting the sleep quality of previous nights) were tracked and reported automatically. They were asked to come to the lab when a desirable score is detected, respecting a balance among the sleep-related fatigue levels NO, RR, and HR. Each LKT session lasted 30 min. Before it, they were instrumented

with the EEG and asked to sit and stay quiet for 2 min. The experimental paradigm simulated a night-view cruising and the lane departures were equally distributed between left and right deviations. Perturbations were presented at intervals of approximately 1 every 7 – 12 s jittered to prevent anticipatory reactions of the drivers (resulting in approximately 180 events per session). If there is no response to the deviation, the simulated vehicle hits the curb and keeps its movement with no feedback to the subject.

During a longitudinal study spanning a 5-month period of daily sampling, 12 subjects were able to complete 3 sessions within each of the three levels of the ES. The rest of the participants completed at least 2 sessions within two classification levels. The subjects attended the sessions within 1 – 3 week intervals and the total number of completed EEG sessions was 141.

### 2.4. EEG Data Acquisition and Preprocessing

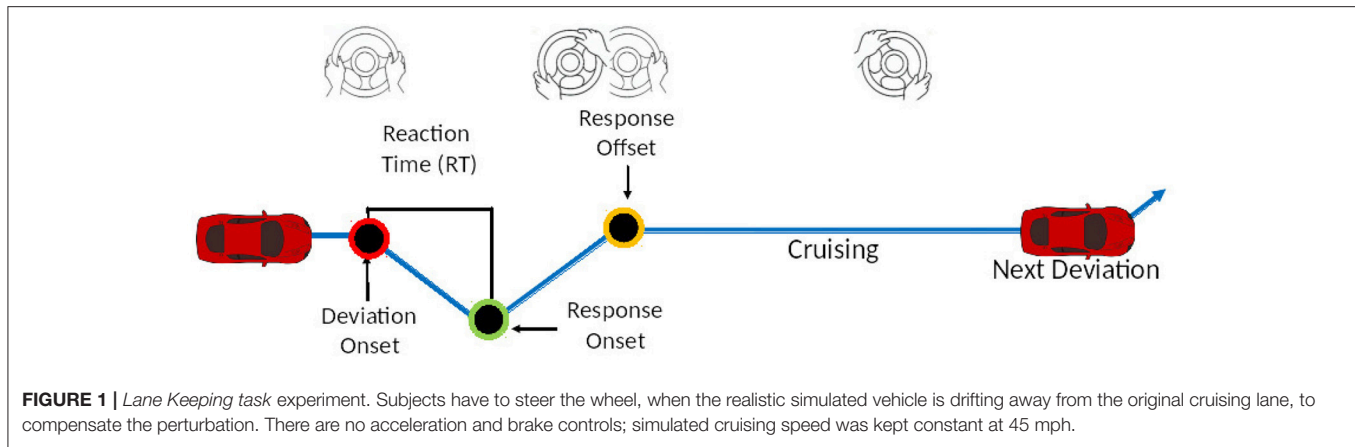
A 64-channel EEG system (Neuroscan Inc.) was used to collect EEG data during the driving task, with channel locations measured by a 3D digitizer following the international 10-20 system. The sampling rate was 1,000 Hz and the impedance was kept below  $5K\Omega$  for all electrodes. The ocular and muscular artifacts were identified in epochs with an amplitude exceeding  $\pm 70 \mu V$  (see **Figure S1** in Supplementary Material for an example) and removed by visual inspection (Tatum et al., 2007; Tandle et al., 2016). The signals were band-pass filtered between 0.5 and 50 Hz and then downsampled to 500 Hz. For our analysis, we selected brain areas and respective channels described in the **Table 1**, based on Lainscek's study (Lainscek et al., 2013). Our analyses focused on the EEG signals 1 s (or 500 points) before each lane-departure event. This choice aims to capture the tonic modulations of attention and engagement during a sustained performance in simulated driving tasks and it was based on the studies of Huang et al. (2007), Chuang et al. (2014), and Lin et al. (2016).

### 2.5. Hypotheses

We hypothesize that the lack of attention in drivers emerges from the interaction of neurobiological mechanisms associated with sleep- and task-related fatigue processes. More specifically, the performance decrements in fatigued drivers are accompanied by effective connectivity changes in several brain areas tied to different spectral behaviors associated with the real-world distractors, resulting in different patterns of the neural rhythms augmentation or suppression.

### 2.6. Reaction Time and Drive Performance

Defined as the elapsed time between the lane departure onset and the response onset, the *Reaction Time* (RT) has been used by several studies to detect subjects' fluctuations of performance in the simulated driving tasks (Huang et al., 2016; Lin et al., 2016). Short RTs are expected from alert drivers who respond quickly to cruising perturbations whereas drowsy drivers tend to react slower and produce longer RTs. To alleviate inter- and intra-subject variability, we define a *Normalized Reaction Time* (NRT)



**TABLE 1 |** Brain areas and the respective selected channels for the effective connectivity analysis.

| Areas            | Channels             |
|------------------|----------------------|
| Left anterior    | F1 F3 F5             |
| Right anterior   | F2 F4 F6             |
| Left motor       | FC1 FC3 FC5 C1 C3 C5 |
| Right motor      | FC2 FC4 FC6 C2 C4 C6 |
| Frontal midline  | FCz Cz               |
| Left parietal    | CP3 CP5 TP7 P3 P5 P7 |
| Right parietal   | CP6 CP4 TP8 P4 P6 P8 |
| Parietal midline | CPz Pz               |
| Left occipital   | PO3 PO7 O1           |
| Right occipital  | PO4 PO8 O2           |

For each session, the measures applied in this work were derived from single trials, normalized to the baseline information and then averaged over channels.

dividing the RTs by the average of the 10% shortest values within each session (sorted in ascending order). For our analyses, we consider a RT lower bound 1 s and upper bound 4 s to analyze transitions from alert to drowsy states. Subjects with NRT out of this interval are considered in very high or very low vigilance states. In the literature, significant changes in power spectra and in directed measures were empirically observed between 2 and 3 s (Chuang et al., 2012; Huang et al., 2015, 2016; Lin et al., 2016). We used a logistic transformation to rescale the NRT to those limits, defining a *Driving Performance* (DP) index (Huang et al., 2015):

$$DP(NRT) = \frac{2 + 2e^{-0.5}}{(1 + e^{-0.5NRT})(1 - e^{-0.5})} - \frac{1 + e^{-0.5}}{1 - e^{-0.5}}.$$

Notice that  $DP(1) = 1$ , DP tends to approximately 4.08 as NRT tends to infinity and it exhibits a close linear relation for NRT between 1 and 4. After the transformation, we set  $DP = 1$  for  $DP < 1$  and  $DP = 4$  for  $DP > 4$ . Therefore, DP maps the unbounded NRT to the interval [1, 4].

## 2.7. Phase-Space Reconstruction

Given an EEG signal,  $X = \{x_1, \dots, x_n\}$ , the spatial and time-delayed embedding coordinates are defined as  $X_{vec} = \{\vec{x}_i = (x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau}); i = 1, \dots, N\}$  where  $N = n - (m -$

$1)\tau$ . The embedding parameters  $m$  and  $\tau$  can be determined independently using the non-parametric Kozachenko-Leonenko estimator (Kozachenko and Leonenko, 1987), as done by Gautama, Mandic and Hulle (GMH) (Gautama et al., 2003). This procedure avoids oversampled trajectories and autocorrelated data effects (Kennel and Abarbanel, 2002). Using the GMH approach for the EEG signals from all subjects and sessions (more details and applications in Baggio and Fonseca, 2011; Fonseca et al., 2015), we obtained  $m = 4$  and  $\tau = 1$ , respectively the maximum embedding dimension and minimum time lag found (see section 3.4 in Supplementary Material for the reconstruction Matlab script).

## 2.8. CCM

Given two EEG signals  $X, Y$  with length  $n$ , we calculate the phase space reconstruction coordinates  $X_{vec}$  with embedding parameters  $m$  and  $\tau$ . For  $i = 1, \dots, N$  where  $N = n - (m - 1)\tau$ , we consider each vector  $\vec{x}_i$  (representing the system dynamical evolution) and obtain:

1 - the distances from  $\vec{x}_i$  to all other states in  $X_{vec}$ :  $D_i = \{d(\vec{x}_i, \vec{x}_j), i \neq j\}$ , where  $d$  represents the euclidean distance between vectors.

2 - the distance-related weights:  $u_i = e^{\frac{-d(\vec{x}_i, \vec{x}_j)}{\min}}$ , where  $\min$  is the minimum distance found in  $D_i$  calculations.

3 - the normalized weights:  $w_i = \frac{u_i}{\sum_{j=1}^{N-1} u_j}$ .

4 - the scalar  $y$ -value estimated by  $X_{vec}$ :  $\hat{y}_i = \sum_{j=1}^{N-1} w_j y_j$ .

We define the CCM from the source signal  $X$  to the target signal  $Y$ , as the correlation between  $\hat{Y} = \{\hat{y}_1, \dots, \hat{y}_N\}$  and  $Y = \{y_{n-N+1}, \dots, y_n\}$  where  $N = n - (m - 1)\tau$ .

Notice that steps 1 to 3 are about  $X$  information and, in step 4, we use the temporal correspondence between  $X_{vec}$  and  $Y_{vec}$  to predict  $Y$  information, where the weights defined in step 3



are the highest for the closest neighbors. By definition, CCM is asymmetric and lies in the interval  $[-1, 1]$  (see **section 3.5** in Supplementary Material for the CCM main Matlab script).

To test the efficiency of CCM, following the ideas reported in Ball et al. (2016), we designed a brain connectivity model with eight coupled damped oscillators sources (see **Figure 2**, left panel) defined by *Autoregressive Processes* (AR) of order 5. Sources 1 to 4 are coupled and located in the *Anterior Cingulate Cortex* (ACC) with respective rhythms 8, 10, 11, and 12 Hz, defining an alpha cluster. Sources 5 to 8 are coupled and lie in the *Posterior Cingulate Cortex* (PCC) with respective frequencies 20, 22, 25, and 30 Hz, configuring a beta cluster. The simulation was performed in three stages of 5 s each. ACC and PCC clusters are disconnected in stages 1 and 3 and coupled during stage 2. Intra- and inter-cluster couplings were defined by Gaussian mixture AR models.

Aiming the analysis of the changes at the causal relationship between the ACC and PCC clusters in the channel level, we used a *Boundary Element Method* (BEM) from the SIFT toolbox (Mullen, 2012) to generate 64-channel EEG signals. This realistic forward head model projects the source activations to the scalp using the “colin27” brain atlas as the reference (Holmes et al., 1998). Varying the white-noise variances in the AR processes from 0.1 to 1 s (step 0.1 s), we simulated ten 64-channel EEG signals with the sampling frequency of 200 Hz (see **sections 3.1** to **3.3** in Supplementary Material for the SIFT settings).

We decomposed the signals into the alpha and beta bands and calculated CCM in windows of 0.25 s (20 points per stage) from the channels in the Left and Right Anterior areas to the ones in the Left and Right Posterior areas (see **Table 1**). The averaged CCM over channels (see **Figure S2** in Supplementary Material for the flowchart of simulation study), is plotted for each noise level in **Figure 1**, right panel. The observed changes in the causal outflow from anterior to posterior channels are consistent with the brain connectivity model defined at the source level. CCM is robust to noise and insensitive to linear mixtures.

## 2.9. Statistical Analysis

Considering 141 sessions and an average number of 143 events per session (total of 20,182 events), we checked the statistical

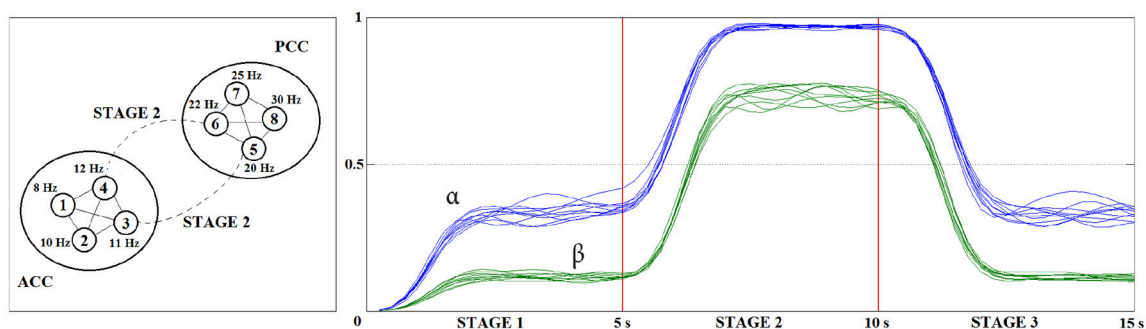
significance of CCM from source to target areas selected in this work. We performed a bootstrapping approach using surrogate data with the same power spectrum of the original signals (Baggio and Fonseca, 2011). A Wilcoxon rank sum test was used with 1% significance level to verify the null hypothesis that the original data, epochs of 1 s before the events for each subject and session, and its surrogates have the same distribution of the CCM values. The null hypothesis was rejected for all sessions indicating that the causal relations are a genuine non-linear feature of the data.

For each session and event, the signals were decomposed into the bands  $\theta$ : [4.5, 7.5] Hz;  $\alpha$ : [7.5, 12.5] Hz;  $\beta$ : [12.5, 20] Hz and  $\gamma$ : [25, 40] Hz, using a FFT procedure. For each band, CCM from source to target channels were calculated in the sessions. Considering in each session the baseline set as the CCM values corresponding to the 10% shortest DPs (in ascending order), CCM were normalized by subtracting the median and dividing by the quartile dispersion of the baseline set. Then, the normalized CCM values were averaged over channel pairs belonging to source and target areas (see **Table 1** for the channel sets definition), defining a baseline relative causal relation between areas. See **Figure 3** for the event signal processing pipeline. The same pipeline was applied to the spectral analysis calculations considering only the target channels  $Y$ . The spectral analysis was performed using the FFT procedure in Matlab (2012b). The results presented in this work will be always relative to the baseline set within sessions and averaged over channels.

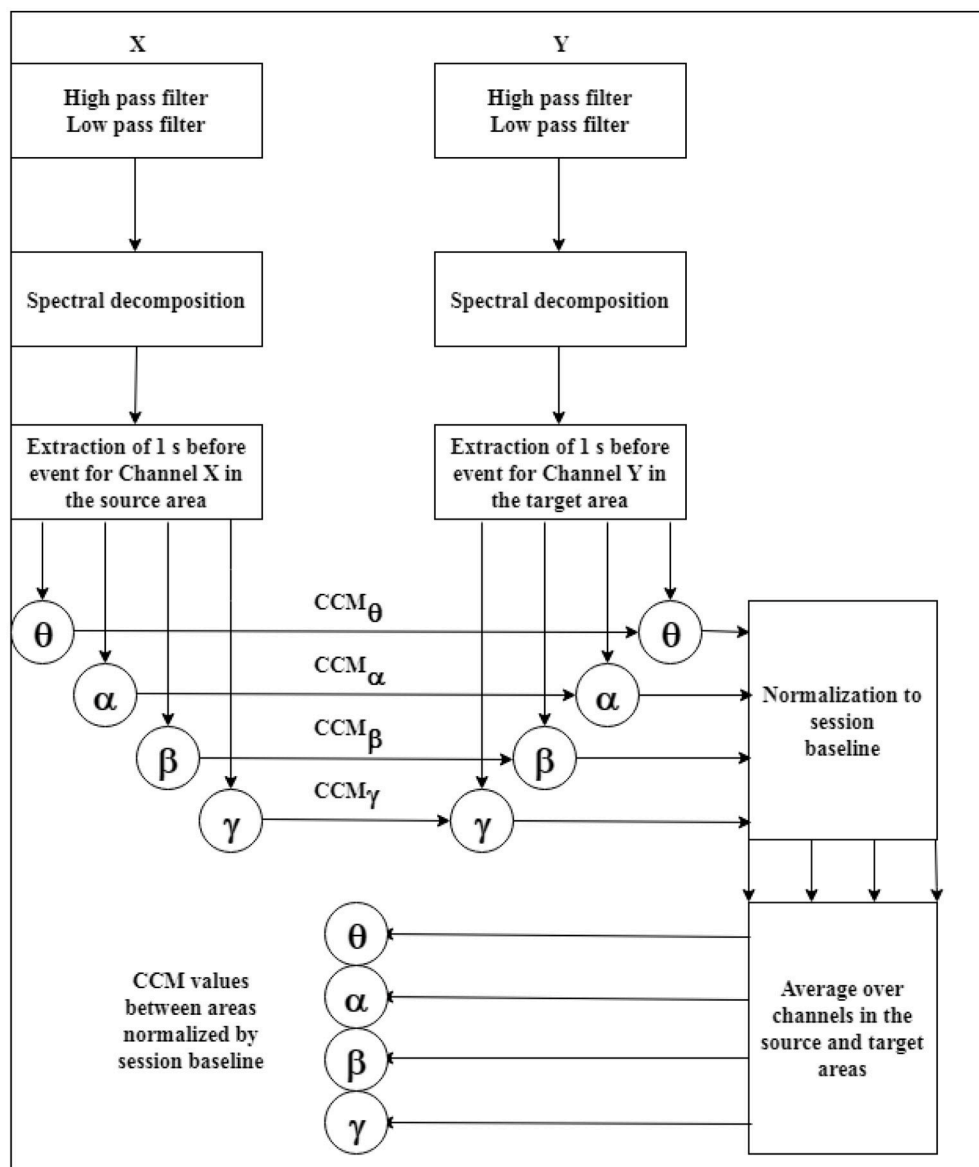
The normalized CCM and spectral values from all subjects and sessions were aggregated, sorted by DP, and then separated in the three levels of sleep-related fatigue NO, RR, and HR with respective sample sizes of 6811, 7136, and 6235.

The significant statistical difference for the normalized CCM values between categories was analyzed by two criteria: the distribution difference was validated by the Wilcoxon rank sum test with 1% significance level, and the slope difference was checked by the  $F$ -test with 1% significance level as well.

CCM-DP, power-DP, and CCM-power statistical relations were investigated by the Pearson's correlation (see **Figure S3** in Supplementary Material for a flowchart of the overall process).



**FIGURE 2 |** Illustration of the simulated eight dynamically coupled sources (**Left**) from the brain connectivity model performed in three stages of 5 s each. On stages 1 and 3 the alpha cluster in ACC and beta cluster in PCC are only intra-coupled. On stage 2, the clusters are intra- and inter-coupled. The source activations were projected to the scalp using a BEM forward head model. The 64-channel EEG signals were simulated for 10 different levels of noise and then decomposed into the alpha and beta bands. Averaged CCM values from anterior to posterior channels (**Right**) were consistent with the changes in the inter-cluster coupling during the stages.



**FIGURE 3 |** Event signal processing pipeline for the CCM values from source channels X to target channels Y. The power analysis for the target channels was conducted using the same steps.

### 3. RESULTS

#### 3.1. NRT and DP Distributions

The RTs were extracted from lane departure events for the 17 subjects and under three different sleep-related fatigue levels (defined by the quality of sleep). For each session, the NRTs were derived and then the DP indexes were obtained. **Table 2** shows the descriptive statistics across NO, RR, and HR conditions defined by the ES. The NRT and DP distributions are skewed to the right due to slow reactions of fatigued drivers and the experimental paradigm (no feedback for hitting the curb). Their distributions are super-Gaussians (Lee et al., 1999) with one and two peaks, respectively, as shown in **Figure 4**. The logistic transformation in the DP calculation was able to decrease

the normalized reaction time variance and keep the quartile dispersion in the same order of magnitude than NRTs, i.e., a non-linear transformation with close to linear effects. The conversion from NRT to DP is a useful procedure for correcting experimental distortions and rescaling an unbounded measure to a more practical behavioral performance index.

Also shown in **Figure 4**, the NRT-DP transformation keeps the ascending order among the sleep-related fatigue levels, for the NRTs- and DPs- distribution means and the peaks (lower values for HR, middle for NO and higher for RR). In the DP domain, it is clearly seen a higher probability of 4 (drowsy state) in the HR level of sleep-related fatigue, not noticed in the NRT domain. The DPs fit the interval [1, 4] (by definition) and reveal new features in the changes of alertness levels.

**TABLE 2 |** Descriptive Statistics of NRT and DP across the sleep-related fatigue levels NO, RR, and HR.

|                         | Normal  | Reduced risk | High risk |
|-------------------------|---------|--------------|-----------|
| <b>Number of events</b> | 6811    | 7136         | 6235      |
| <b>NRT</b>              |         |              |           |
| Mean                    | 1.9172  | 2.0589       | 2.8199    |
| Standard deviation      | 1.6897  | 3.2471       | 9.2638    |
| Quartile dispersion     | 0.2331  | 0.2581       | 0.2935    |
| <b>DP</b>               |         |              |           |
| Mean                    | 1.6873  | 1.7046       | 1.8858    |
| Standard deviation      | 0.5986  | 0.6274       | 0.7987    |
| Quartile dispersion     | 0.22097 | 0.2315       | 0.2579    |

The transformation NRT to DP provides distributions with lower variability, but with similar quartiles dispersion structure.

### 3.2. CCM Oscillations Indexed by DP

We first analyzed the relation between the normalized CCM and DP values. For different target areas and bands, the CCM values exhibit a strong oscillatory behavior in DP between 1 and 2. For DP between 2 and 4, a nearly monotonic behavior was noticed and the Pearson's correlation was evaluated. Considering the two categories of sleep-related fatigue NO and HR, for more than 90% of the 144 cases (2 source areas  $\times$  9 target areas  $\times$  4 frequency bands  $\times$  2 categories), the causal relations exhibited a strong positive or negative correlation (absolute value greater than 0.7) with the performance index DP. The strong correlation for the RR level of fatigue was not observed in this case. In the **Figure 5**, the top panel shows the normalized CCM values from the source, Frontal Midline, to the target, Parietal Midline areas, sorted by DPs, for the four frequency bands and three levels of sleep-related fatigue NO (blue), RR (green), and HR (red). In short DP's, between 1 and 2, related to alert states, it's possible to observe a mirror pattern between NO and HR levels. In the longer DP's, related to drowsy states, we observe different trends in the nearly monotonic behavior between the same two levels.

**Table 3** shows the significant changes in the normalized CCM values of sleep-related fatigue levels NO vs. HR, for DPs between 2 and 4, where a nearly monotonic behavior was observed. The source of the dynamical coupling was the Frontal Midline area and the targets were the other selected areas represented in different rows. For several targets and bands, the HR- and NO-normalized CCM values have different distributions and slopes. Gray background cells in **Table 3** indicate simultaneous significant statistical differences in distributions and slopes (considering the significance level of 1%) between the two levels of sleep-related fatigue, i.e., it points out the targets and bands where the causal relations have different trends (positive and negative slopes) with different probability of occurrences.

We also investigated the effective connectivity from the source, Parietal Midline, to the other selected targets. Although CCM is not symmetric by definition, the causal relations from the Parietal Midline to Frontal Midline are similar to its opposite direction values, indicating a bi-directional causation between those two areas. **Table 4** shows the statistical analyses of the

normalized CCM values from the Parietal Midline area between levels NO and HR, to different targets at different frequency bands, as shown in **Table 3**.

### 3.3. Spectral Power Indexed by DP

The relation between normalized EEG power and DP was studied for all ten areas defined in **Table 1**, for the bands  $\delta, \alpha, \beta, \gamma$ , and for the two categories of sleep-related fatigue NO and HR. In **Figure 5**, for instance, the bottom panel shows the fluctuations of the normalized power in different bands, for the Parietal Midline area.

As for the normalized CCM values, the normalized power in the targets exhibits an oscillatory behavior for DP less than 2. For the DP higher than 2, we observed a nearly linear behavior. In this domain, between 2 and 4, the Pearson's correlation between power and DP were calculated and the results exhibited a strong positive or negative correlation (absolute value greater than 0.75) between spectral activity and the performance index for more than 90% of the 80 cases analyzed.

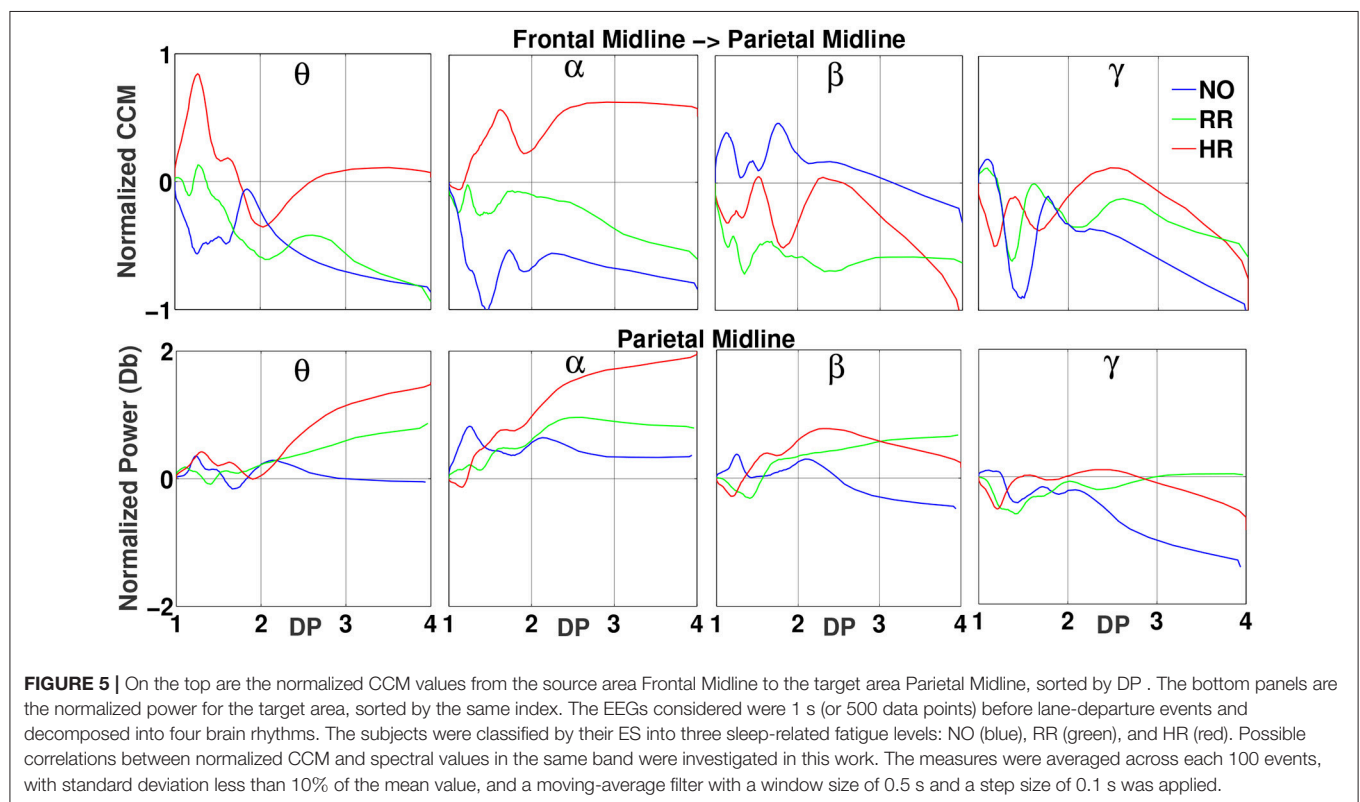
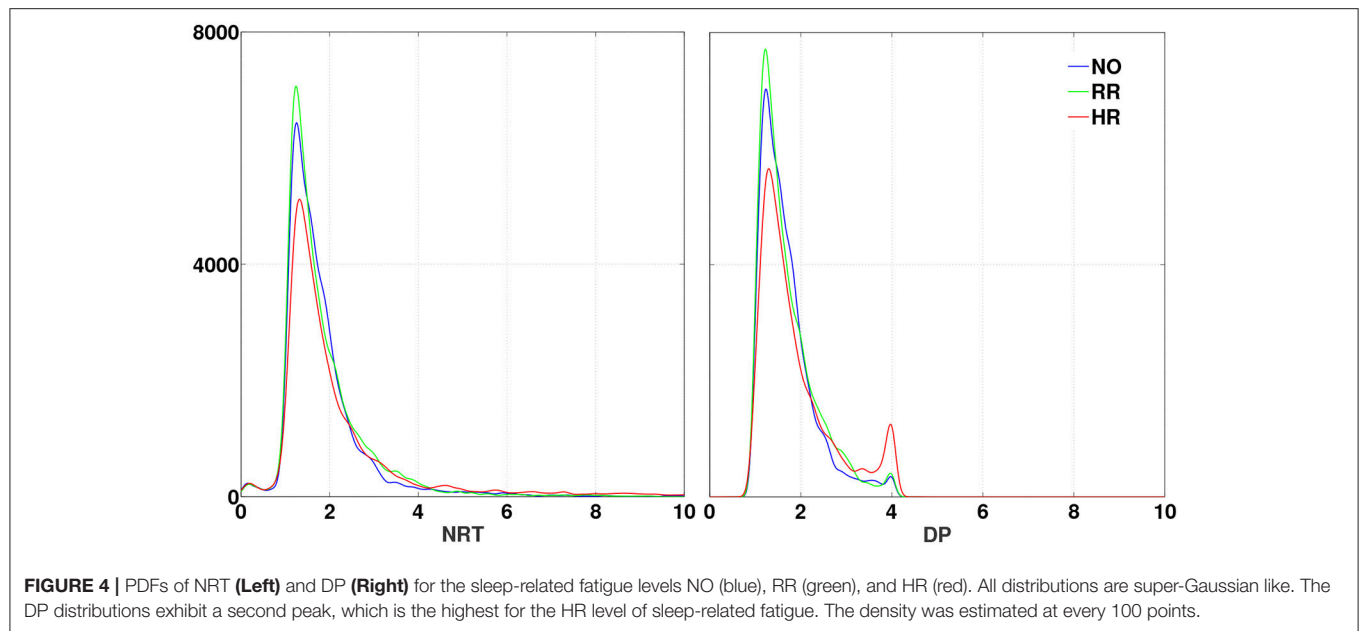
### 3.4. CCM-Power Correlation

After exploring the CCM-DP and power-DP relations, the next question is how CCM and spectral power interacted considering the same target area. The procedure of evaluating the causal relation to a specific target in different frequency bands allows a natural connection with spectral power in the same target. Considering the source of CCM as the Frontal Midline and Parietal Midline areas, we restricted our study to the cases where the distributions and slopes were significantly different between the levels of sleep-related fatigue NO and HR, marked as gray in **Table 3**. **Table 5** lists the correlations between normalized CCM values (considering both sources) and normalized spectral power sorted by DPs.

## 4. DISCUSSION

This study observed a group of young university students in their natural environment during a 20-week semester. We believe our subjects are a representative sample of healthy young adults in real-world environments, with expected high levels of stress and irregular sleep (Lund et al., 2010). With the sustained-attention experiment, we aim to understand the connections between those subjective parameters and the performance decrements in sleep-related fatigue, characterizing its variability and instability (Chua et al., 2014). To achieve this goal, the starting point was the signal-reconstruction process. The embedding coordinates revealed different recurrence structures linked to the three levels of sleep-related fatigue defined by the ES. We consider that this representation was sensible to the different quality and quantity of sleep across subjects, quantifying behavioral and physiologic information from the different fatigue states determined by the Readiband using the SAFTE model.

The choice of the performance index to sort the normalized CCM and power spectrum values was crucial. The NRTs exhibit high variance and positive-skew distributions, an expected



outcome since fatigued drivers can exhibit low performance, failures (Huang et al., 2009; Liu et al., 2010), and even fall asleep. As the subjects have no feedback from the driving simulator when the vehicle hits the curb and maintains a continuous cruising, the NRTs can deviate significantly from the baseline. The transformation from NRTs to the DPs,

considering the interval [1,4] to analyze the EEG correlates of alertness-drowsiness transitions, alleviates this issue. As a consequence of its nearly linear behavior we obtain lower standard deviations, but with no robust changes in the data structure observed in the quartile dispersion (see Table 2). The DPs-distributions have two peaks, the second peak can be



**TABLE 3 |** Statistical analysis of normalized CCM values in different bands (columns), considering the Frontal Midline area as the source and the other areas (rows) as targets.

| Targets          | Source: Frontal Midline          |    |   |                                 |    |   |                                   |    |   |                                  |    |   |
|------------------|----------------------------------|----|---|---------------------------------|----|---|-----------------------------------|----|---|----------------------------------|----|---|
|                  | $\theta$                         |    |   | $\alpha$                        |    |   | $\beta$                           |    |   | $\gamma$                         |    |   |
|                  | NO                               | HR | p | NO                              | HR | p | NO                                | HR | p | NO                               | HR | p |
| Left anterior    | -0.4455, 0.5050 ( $< 10^{-4}$ )  |    |   | 0.1723, 0.6949 ( $< 10^{-4}$ )  |    |   | -0.6168, 0.8107 ( $< 10^{-4}$ )   |    |   | -1.7217, 0.5473 ( $< 10^{-4}$ )  |    |   |
|                  | -0.0363, 0.7376 ( $< 10^{-4}$ )  |    |   | -0.1934, 0.3099 ( $< 10^{-4}$ ) |    |   | -0.1934, 0.3099 ( $< 10^{-4}$ )   |    |   | -1.863, 0.2306 ( $< 10^{-4}$ )   |    |   |
| Right anterior   | -0.6473, -0.2415 ( $< 10^{-4}$ ) |    |   | 0.2379, 0.3733, (0.0040)        |    |   | -0.0364, 0.5440, ( $< 10^{-4}$ )  |    |   | -0.6130, 0.8925, ( $< 10^{-4}$ ) |    |   |
|                  | -0.1636, -0.0306 (0.0139)        |    |   | -0.1444, -0.01206 (0.0190)      |    |   | -0.1172, 0.2219, ( $< 10^{-4}$ )  |    |   | -0.0956, 0.71532 ( $< 10^{-4}$ ) |    |   |
| Left motor       | -0.3352, -0.2282 (0.3771)        |    |   | 0.5908, 0.7794 (0.4944)         |    |   | -0.5266, 0.2940 ( $< 10^{-4}$ )   |    |   | -0.8709, 0.2390, ( $< 10^{-4}$ ) |    |   |
|                  | -0.0032, 0.4066 ( $< 10^{-4}$ )  |    |   | -0.1753, 0.6630 ( $< 10^{-4}$ ) |    |   | -0.2516, -0.2072 (0.2020)         |    |   | -0.5131, -0.2163 ( $< 10^{-4}$ ) |    |   |
| Right motor      | -0.7234, 0.2485 ( $< 10^{-4}$ )  |    |   | -0.1045, 0.37740 (0.0015)       |    |   | -0.1504, -0.0984 (0.0343)         |    |   | -1.2960, 0.7357 ( $< 10^{-4}$ )  |    |   |
|                  | -0.1144, 0.9210 ( $< 10^{-4}$ )  |    |   | 0.3059, 0.6328 ( $< 10^{-4}$ )  |    |   | -0.1367, -0.0882 (0.0717)         |    |   | -0.6241, 0.2950 ( $< 10^{-4}$ )  |    |   |
| Left parietal    | -0.4649, -0.5567 (0.582)         |    |   | 0.4727, 0.6611 (0.4533)         |    |   | -0.6212, 0.1600 ( $< 10^{-4}$ )   |    |   | -0.4561, 0.2505 ( $< 10^{-4}$ )  |    |   |
|                  | -0.3071, 0.1694 ( $< 10^{-4}$ )  |    |   | -0.2654, 0.7004 ( $< 10^{-4}$ ) |    |   | -0.2999, -0.0092 ( $< 10^{-4}$ )  |    |   | -0.1184, -0.1096 (0.8348)        |    |   |
| Parietal midline | -0.5704, -0.0533 ( $< 10^{-4}$ ) |    |   | -0.6464, 0.5359 ( $< 10^{-4}$ ) |    |   | 0.0687, -0.3857 (0.0001)          |    |   | -0.5312, -0.1902 (0.0026)        |    |   |
|                  | -0.2611, 0.1842 ( $< 10^{-4}$ )  |    |   | -0.1209, 0.0839 ( $< 10^{-4}$ ) |    |   | -0.2252, -0.6031 (0.0002)         |    |   | -0.3546, -0.4453 (0.2083)        |    |   |
| Right parietal   | -0.7648, -0.4996 (0.0029)        |    |   | -0.8135, 0.4037 ( $< 10^{-4}$ ) |    |   | -0.06380, -0.4208 ( $< 10^{-4}$ ) |    |   | -0.9670, 1.1025 ( $< 10^{-4}$ )  |    |   |
|                  | 0.1465, 0.3360 ( $< 10^{-4}$ )   |    |   | 0.2103, 0.2148 (0.9183)         |    |   | 0.2122, -0.2697 ( $< 10^{-4}$ )   |    |   | -0.4766, 0.3390 ( $< 10^{-4}$ )  |    |   |

For the sleep-related fatigue levels NO and HR, the distributions and slopes were analyzed for DP in the interval [2, 4]. In each cell, on the top, are the CCM means, respectively of NO and HR categories, and the p-value for the Wilcoxon rank test inside brackets, with the null hypothesis that the two levels of sleep-related fatigue have CCM values with the same distributions. On the bottom, are the slopes respectively of NO and HR values and the p-value for the F-test inside brackets, with the null hypothesis that those two levels have CCM values with identical slopes in their linear regressions. Simultaneous significant probabilities shift and trend changes between NO and HR levels are indicated by the gray background. See **Figure 5** (top) to visualize plots of CCM from Frontal Midline to Parietal Midline areas.

**TABLE 4 |** Statistical analysis for normalized CCM values between NO and HR levels of sleep-related fatigue in different bands (columns), considering the Parietal Midline area as the source.

| Targets         | Source: Parietal Midline         |    |   |                                  |    |   |                                  |    |   |                                   |    |   |
|-----------------|----------------------------------|----|---|----------------------------------|----|---|----------------------------------|----|---|-----------------------------------|----|---|
|                 | $\theta$                         |    |   | $\alpha$                         |    |   | $\beta$                          |    |   | $\gamma$                          |    |   |
|                 | NO                               | HR | p | NO                               | HR | p | NO                               | HR | p | NO                                | HR | p |
| Left motor      | -0.3324, -0.6485 ( $< 10^{-4}$ ) |    |   | -0.1870, 0.8239 ( $< 10^{-4}$ )  |    |   | -0.0292, -0.5834 (0.0003)        |    |   | -0.3783, 0.16174, ( $< 10^{-4}$ ) |    |   |
|                 | -0.1417, 0.1447 ( $< 10^{-4}$ )  |    |   | -0.5294, 0.0463 ( $< 10^{-4}$ )  |    |   | -0.2978, -0.7328 (0.0003)        |    |   | -0.5198, -0.2839 (0.0038)         |    |   |
| Right motor     | -0.2939, -0.5015, (0.0002)       |    |   | -0.5207, 0.6376, ( $< 10^{-4}$ ) |    |   | -0.2758, -0.3164, (0.6054)       |    |   | -0.8973, 0.6191 ( $< 10^{-4}$ )   |    |   |
|                 | 0.2246, 0.1532 (0.0126)          |    |   | -0.2958, -0.0573 ( $< 10^{-4}$ ) |    |   | -0.1883, -0.1997 (0.8385)        |    |   | -0.4069, 0.2053 ( $< 10^{-4}$ )   |    |   |
| Left parietal   | -0.2060, -0.6120 (0.0002)        |    |   | 0.0133, 0.8920 ( $< 10^{-4}$ )   |    |   | 0.1918, 0.9166 ( $< 10^{-4}$ )   |    |   | -0.8330, -0.0814 ( $< 10^{-4}$ )  |    |   |
|                 | -0.0516, 0.4103 ( $< 10^{-4}$ )  |    |   | -0.1737, 0.1325 (0.0072)         |    |   | -0.1124, 0.4233 ( $< 10^{-4}$ )  |    |   | -1.1663, -0.1527 ( $< 10^{-4}$ )  |    |   |
| Right parietal  | -0.9579, -0.2734 ( $< 10^{-4}$ ) |    |   | -0.3390, -0.0357 ( $< 10^{-4}$ ) |    |   | -0.3526, 0.0388 (0.0023)         |    |   | -0.4815, 0.9359 ( $< 10^{-4}$ )   |    |   |
|                 | 0.4247, 0.1958 ( $< 10^{-4}$ )   |    |   | -0.1048, -0.1430 (0.1833)        |    |   | 0.3601, -0.4709 ( $< 10^{-4}$ )  |    |   | -0.1518, 0.2458 ( $< 10^{-4}$ )   |    |   |
| Left occipital  | 0.9939, -0.7421 ( $< 10^{-4}$ )  |    |   | -0.8257, 0.5110 ( $< 10^{-4}$ )  |    |   | -0.3524, 0.3047 ( $< 10^{-4}$ )  |    |   | -0.4807, -0.4323 (0.4944)         |    |   |
|                 | 0.9123, -0.1092 ( $< 10^{-4}$ )  |    |   | -0.4302, 0.1268 ( $< 10^{-4}$ )  |    |   | -0.2363, -0.2011 (0.3262)        |    |   | -0.6942, -0.2405 ( $< 10^{-4}$ )  |    |   |
| Right occipital | -1.0210, -0.0349 ( $< 10^{-4}$ ) |    |   | -0.7530, -0.4193 ( $< 10^{-4}$ ) |    |   | -0.9199, -1.4260 (0.6054)        |    |   | -1.2437, -0.1614 ( $< 10^{-4}$ )  |    |   |
|                 | 0.0993, 0.0103 ( $< 10^{-4}$ )   |    |   | -0.1763, -0.1638 (0.7843)        |    |   | -0.2920, -2.0188 ( $< 10^{-4}$ ) |    |   | -0.9254, -0.4901 ( $< 10^{-4}$ )  |    |   |

The parameters and p-values are the same defined in **Table 3**. As done before, simultaneous significant probabilities shift and trend changes are indicated by the gray background.

attributed to the drowsiness state. For the HR level of sleep-related fatigue, the second peak is the highest, which is consistent with the putative fatigue level derived from the actigraphy data (ES).

Both normalized CCM and spectral values were strongly correlated (positively or negatively) with DPs between 2 and 4 in the levels NO and HR of sleep-related fatigue, as illustrated in **Figure 5**. For the shorter DPs (lying in the

interval [1,2]) when subjects were in the alert state under the sleep levels NO and HR, different oscillations in several sources and targets were observed, a mirror behavior, indicating opposite shifts in the effective connectivity. As for longer DPs (lying in the interval [2,4]), where subjects were drowsy, different trends and distributions for the CCM values were found between the NO and HR sleep levels, revealing again different shifts of the effective

**TABLE 5 |** Pearson's correlations for the normalized CCM-spectral values considering the same target areas in different frequency bands for the sleep-related fatigue levels NO and HR.

| Targets                         |            | NO      | HR      |
|---------------------------------|------------|---------|---------|
| <b>SOURCE: FRONTAL MIDLINE</b>  |            |         |         |
| Left anterior                   | $\theta$ : | 0.5622  | 0.9867  |
|                                 | $\alpha$ : | 0.9974  | 0.0684  |
|                                 | $\beta$ :  | 0.9773  | -0.9129 |
|                                 | $\gamma$ : | 0.9617  | -0.9387 |
| Right anterior                  | $\beta$ :  | 0.7320  | -0.8859 |
|                                 | $\gamma$ : | 0.0815  | -0.9877 |
| Right motor                     | $\theta$ : | 0.8766  | 0.9491  |
|                                 | $\gamma$ : | 0.9920  | -0.9716 |
| Parietal midline                | $\theta$ : | 0.7685  | 0.9666  |
|                                 | $\alpha$ : | 0.5330  | 0.9933  |
| Right parietal                  | $\beta$ :  | -0.9527 | 0.9676  |
|                                 | $\gamma$ : | 0.9938  | -0.7273 |
| <b>SOURCE: PARIETAL MIDLINE</b> |            |         |         |
| Left motor                      | $\theta$ : | 0.9319  | 0.9399  |
|                                 | $\alpha$ : | 0.9861  | 0.9038  |
| Right motor                     | $\gamma$ : | 0.9788  | -0.8707 |
| Left parietal                   | $\theta$ : | 0.6080  | 0.9916  |
|                                 | $\alpha$ : | 0.9180  | 0.1632  |
|                                 | $\beta$ :  | 0.6131  | -0.9460 |
| Right parietal                  | $\beta$ :  | -0.9363 | 0.9834  |
|                                 | $\gamma$ : | 0.9389  | -0.9494 |
| Left occipital                  | $\theta$ : | -0.8264 | -0.6177 |
|                                 | $\alpha$ : | 0.9674  | 0.6303  |

On the left, the source of CCM is the Frontal Midline Area. On the right, the source is the Parietal Midline Area. The first column of each Table specifies the target areas. The choice was based on the simultaneous significant differences in distributions and slopes between those two levels (marked as gray in **Tables 3, 4**). Both measures were sorted by DPs within the interval [2, 4]. See **Figure 5** for the plots of the case from the Frontal Midline source to the Parietal Midline target.

connectivity. Those results demonstrate that DP is an efficient index to understand alertness-drowsiness transitions (Huang et al., 2015).

The information transferred from the source areas Frontal Midline and Parietal Midline to their neighboring areas during the 1 s pre-stimulus period have different rates between subjects in the NO and HR levels of sleep-related fatigue. This difference can be attributed to specific patterns in the effective connectivity related to behavioral microsleeps, reported in Toppi et al. (2016). In both Frontal Midline and Parietal Midline sources of connectivity, for almost all analyzed targets (with the exception from the Parietal Midline area to the Right Occipital area) the normalized CCM values, in some frequency, have significantly different distributions, a negative slope in the NO condition and a positive slope for the HR of fatigue, indicating the ES classification (related to sleep quality) can distinguish new features in the fatigued drivers (with DPs between 2 and 4). In the HR fatigue level, for the bands indicated in the gray background cells in **Tables 3, 4**, the normalized CCM values increase with the increments

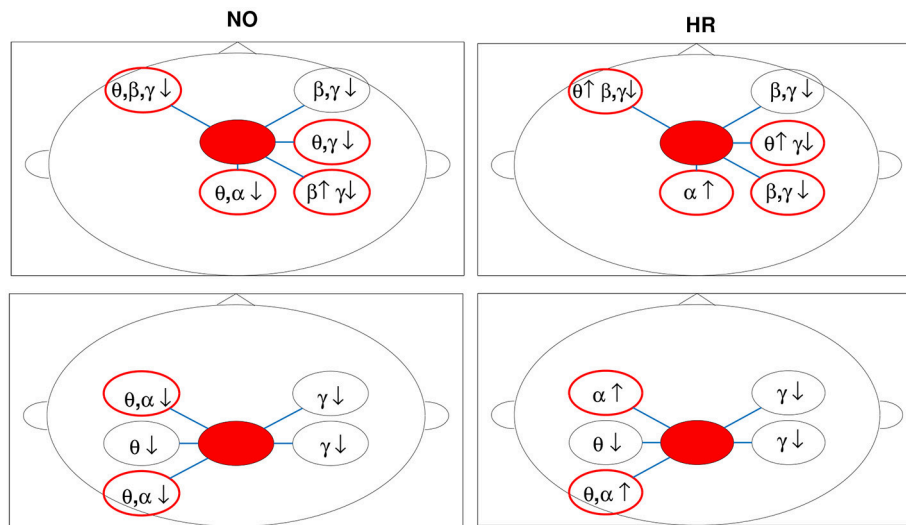
in DP (with 3 exceptions), suggesting enhanced coupling among the studied areas in the fatigued drivers with low sleep quality.

The correlations between the normalized CCM and spectral values are detailed in **Table 5** and represent a novel application to analyze the shifts in the effective connectivity in brain areas during the sustained-attention tasks, allowing us to explore its correlates with the subject fatigue level. We considered only the couplings and bands where study results showed significant differences in distributions and slopes of the causal relations between the sleep-related fatigue categories. Strong positive and negative values were derived.

The normalized CCM values sorted by DP with an increasing magnitude indicates tonic changes of brain dynamics associated with a decline in alertness (DP variation from 2 to 4 is associated with sub-optimal and poor performances; Huang et al., 2015). We focused our attention on those cases, where the CCM values either increased with DP (a positive slope) or decreased with DP (a negative slope). The effective connectivity measure applied in this work is based on the dynamical coupling of brain areas and can modulate the power spectra as reported in Soldatenko and Chichkine (2014) and Lacot et al. (2016), where new power peaks and the enhancement of the original harmonics are associated with the increasing of coupling strength. In brain networks, this modulation was noticed in the BOLD signal analysis, where fMRI-based connectivity and frequency-specific EEG power are related (Conner et al., 2011; Scheeringa et al., 2012). So, it is reasonable to claim that strong CCM-power correlations represent augmentation or suppression for a specific oscillatory activity in the target areas. Taking this into consideration, we combined the information from **Tables 3–5** and illustrated the brain network changes for the NO and HR levels of sleep-related fatigue in **Figure 6**. In the figure, the sources are indicated by the filled red circles and the augmentation or suppression are represented, respectively, by up and down arrows. The targets with significant differences between levels (augmentation to suppression or vice-versa) are indicated by red circles.

The  $\gamma$  band relates to the higher-order cognitive activities for internal modeling of motor control to form a representation shaping internal models to improve motor performance, the suppression of this oscillation observed in different areas for subjects in the NO and HR levels of sleep-related fatigue could indicate the weakening in such ability during fatigue. The  $\gamma$  rhythm suppression could also suggest a weakening in the complex cognitive functions related to attention and memory (Jensen et al., 2007) expressed, for instance, in a difficult of maintaining visual shapes in short-term memory (Tallon-Baudry et al., 1998), reasonable for fatigued subjects (DP is higher than 2 in both levels).

The  $\theta$  frequency is related to cognitive control. The increase of  $\theta$  power is to coordinate activities of various brain regions to update the motor plan in response to somatosensory inputs. There is a suppression of this oscillation for subjects in the NO level and augmentation for the HR level. This could show the increase of the drowsy drivers' efforts to maintain



**FIGURE 6 |** Brain network changes for fatigued drivers (grand averages for DPs between 2 and 4) in the sleep-related fatigue levels NO (good quality sleep) and HR (poor quality sleep). The Frontal Midline and Parietal Midline areas were considered sources (red filled circles) of the effective connectivity and the CCM-power correlations were analyzed. Augmentation and suppression in the neural rhythms are indicated respectively by up and down arrows. The red circles indicate target areas with different spectral activity (augmentation-suppression) between levels.

the similar driving performance. This significant increase in the  $\theta$  activity was also observed in drivers during the transitional phase from alertness to fatigue (Lal and Craig, 2002), in the frontal area was associated with mental fatigue (Wascher et al., 2014) and in the occipital-parietal areas was related to working-memory processing (Raghavachari et al., 2006).

We observed a suppression in the  $\theta$  and  $\alpha$  activities in the occipital area for subjects in the NO level of sleep-related fatigue. This finding suggests that the driver is more concentrated on the task than the ones in the HR level, for instance, processing some visual or auditory information from the realistic simulated vehicle, as observed in Lin et al. (2010). For subjects in the HR level,  $\theta$  and  $\alpha$  are activated in the occipital, motor and parietal areas (by the sources Frontal Midline and Parietal Midline). In this level of sleep-related fatigue representing a lack of sleep, the subjects tend more to mind-wandering under low perceptual demands (Lin et al., 2016). Similar findings were obtained during simulated driving in Huang et al. (2009).

The opposite trends in the change of  $\alpha$  and  $\beta$  activities in the parietal area between subjects in those two sleep-related fatigue levels can be associated with different mechanisms for movement processing. In this context, subjects in the HR level could be more sensitive to movement selection demands where an increasing  $\alpha$  and decreasing  $\beta$  were detected. Those findings are consistent with the actual and imagined movements reported in Brinkman et al. (2014).

The identification of distinct sleep-related fatigue levels was crucial for discriminating the effective connectivity patterns observed in the task-positive network of drivers. Their importance is based on the hypothesis that the sleep

loss may affect brain functions locally, in a bottom-up regulation of temporal changes in neurobehavioral performance (Van Dongen et al., 2011), suggesting a dependence on cumulative increase in activation of the neuronal groups. This summative activation requiring to gather cognitive resources can explain the neural network changes observed in different frequencies during the sustained-attention driving task. Our results from DP, normalized CCM and spectral values support this bottom-up theory where performance is readjusted by the circadian rhythm and time-on-task effects.

## 5. CONCLUSION

The combination of EEG, behavioral and physiological information (expressed respectively in the CCM, DP and ES measures) as well the information about the task and socio-environmental context in which the driving experiments were performed, can highlight the real-world fatigue phenomenon. The spectral changes observed in the alertness oscillations can be explained by effective connectivity measures. CCM analysis over specific brain areas brain areas can predict different patterns of augmentation and suppression in the neural rhythms. CCM results can improve the development of real time devices for monitoring driver vigilance.

## AUTHOR CONTRIBUTIONS

AF and T-PJ contributed conception and design of the study. C-TL and J-TK organized the database. AF and SK performed the statistical analysis. AF wrote the first draft of the manuscript.

SK wrote sections of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

## FUNDING

This work was partially supported by CAPES (Brazil), by the Cognition and Neuroergonomics Collaborative Technology Alliance Annual Program Plan through the Army Research Laboratory under Cooperative Agreement under

Grant W911NF-10-2-0022, and by the Australian Research Council (ARC) under discovery Grants DP180100670 and DP180100656.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2018.00418/full#supplementary-material>

## REFERENCES

- Baccalá, L. A., and Sameshima, K. (2001). Partial directed coherence: a new concept in neural structure determination. *Biol. Cybern.* 84, 463–474. doi: 10.1007/PL00007990
- Baechinger, M., Zerbi, V., Moisa, M., Polania, R., Liu, Q., Mantini, D., et al. (2017). Concurrent tacs-fmri reveals causal influence of power synchronized neural activity on resting state fmri connectivity. *J. Neurosci.* 37, 4766–4777. doi: 10.1523/JNEUROSCI.1756-16.2017
- Baggio, G., and Fonseca, A. (2011). Complex dynamics of semantic memory access in reading. *J. R. Soc. Interface* 9, 328–338. doi: 10.1098/rsif.2011.0289
- Balkin, T. J., and Wesensten, N. J. (2011). “Differentiation of sleepiness and mental fatigue effects,” in *Cognitive Fatigue: Multidisciplinary Perspectives on Current Research and Future Applications*, ed P. L. Ackerman (Washington, DC: American Psychological Association), 47–66.
- Ball, K., Bigdely-Shamlo, N., Mullen, T., and Robbins, K. (2016). PWC-ICA: a method for stationary ordered blind source separation with application to EEG. *Comput. Intell. Neurosci.* 2016:9754813. doi: 10.1155/2016/9754813
- Barnett, L., and Seth, A. K. (2014). The MVGC multivariate granger causality toolbox: a new approach to granger-causal inference. *J. Neurosci. Methods* 223, 50–68. doi: 10.1016/j.jneumeth.2013.10.018
- Belenky, G., Wesensten, N. J., Thorne, D. R., Thomas, M. L., Sing, H. C., Redmond, D. P., et al. (2003). Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study. *J. Sleep Res.* 12, 1–12. doi: 10.1046/j.1365-2869.2003.00337.x
- Bonnet, M. H., and Arand, D. L. (1995). We are chronically sleep deprived. *Sleep* 18, 908–911.
- Brinkman, L., Stolk, A., Dijkerman, H., de Lange F. P., and Toni, I. (2014). Distinct roles for alpha- and beta-band oscillations during mental simulation of goal-directed actions. *J. Neurosci.* 34, 14783–14792. doi: 10.1523/JNEUROSCI.2039-14.2014
- Brown, I. D. (1994). Driver fatigue. *Hum. Factors* 36, 298–314.
- Chua, E.-P., Yeo, S.-C., and Lee, I.-G. (2014). Sustained attention performance during sleep deprivation associates with instability in behavior and physiologic measures at baseline. *Sleep* 37, 27–39. doi: 10.5665/sleep.3302
- Chuang, C.-H., Ko, L.-W., Jung, T.-P., and Lin, C.-T. (2014). Kinesthesia in a sustained-attention driving task. *Neuroimage* 91, 187–202. doi: 10.1016/j.neuroimage.2014.01.015
- Chuang, S.-W., Ko, L.-W., Lin, Y.-P., Huang, R.-S., Jung, T.-P., and Lin, C.-T. (2012). Co-modulatory spectral changes in independent brain processes are correlated with task performance. *Neuroimage* 62, 1469–1477. doi: 10.1016/j.neuroimage.2012.05.035
- Conner, C. R., Ellmore, T. M., Pieters, T. A., DiSano, M. A., and Tandon, N. (2011). Variability of the relationship between electrophysiology and bold-fmri across cortical regions in humans. *J. Neurosci.* 31, 12855–12865. doi: 10.1523/JNEUROSCI.1457-11.2011
- Dawson, D., Noy, Y. I., Harma, M., Akerstedt, T., and Belenky, G. (2011). Modelling fatigue and the use of fatigue models in work settings. *Accid. Anal. Prev.* 43, 549–564. doi: 10.1016/j.aap.2009.12.030
- Fonseca, A., Boboeva, V., Brederoo, S., and Baggio, G. (2015). Disrupting morphosyntactic and lexical semantic processing has opposite effects on the sample entropy of neural signals. *Brain Res.* 1604, 1–14. doi: 10.1016/j.brainres.2015.01.030
- Friston, K. (2011). Functional and effective connectivity: a review. *Brain Connect.* 1, 13–36. doi: 10.1089/brain.2011.0008
- Friston, K., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. *Neuroimage* 19, 1273–1302. doi: 10.1016/S1053-8119(03)00202-7
- Friston, K., Moran, R., and Seth, A. K. (2013). Analysing connectivity with granger causality and dynamic causal modelling. *Curr. Opin. Neurobiol.* 23, 172–178. doi: 10.1016/j.conb.2012.11.010
- Gautama, T., Mandic, D. P., and Hulle, M. M. V. (2003). “A differential entropy based method for determining the optimal embedding parameters of a signal,” in *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on*, vol. 6 (Hong Kong), VI–29–32.
- Goldenberg, D., and Galvn, A. (2015). The use of functional and effective connectivity techniques to understand the developing brain. *Dev. Cogn. Neurosci.* 12, 155–164. doi: 10.1016/j.dcn.2015.01.011
- Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37, 424–438.
- Härmä, M., Kompier, M. A. J., and Vahtera, J. (2006). Work-related stress and health-risks, mechanisms and countermeasures. *Scand. J. Work Environ. Health* 32, 413–419. doi: 10.5271/sjweh.1047
- Harrison, Y., Jones, K., and Waterhouse, J. (2007). The influence of time awake and circadian rhythm upon performance on a frontal lobe task. *Neuropsychologia* 45, 1966–1972. doi: 10.1016/j.neuropsychologia.2006.12.012
- Hirata, Y., Amigó, J., Matsuzaka, Y., Yokota, R., Mushiaki, H., and Aihara, K. (2016). Detecting causality by combined use of multiple methods: Climate and brain examples. *PLoS ONE* 11:e0158572. doi: 10.1371/journal.pone.0158572
- Holmes, C. J., Hoge, R., Collins, L., Woods, R., Toga, A. W., and Evans, A. C. (1998). Enhancement of mr images using registration for signal averaging. *J. Comput. Assist. Tomogr.* 22, 324–333.
- Huang, C.-S., Pal, N. R., Chuang, C.-H., and Lin, C.-T. (2015). Identifying changes in EEG information transfer during drowsy driving by transfer entropy. *Front. Hum. Neurosci.* 9:570. doi: 10.3389/fnhum.2015.00570
- Huang, K.-C., Huang, T.-Y., Chuang, C.-H., King, J.-T., Wang, Y.-K., Lin, C.-T., et al. (2016). An EEG-based fatigue detection and mitigation system. *Int. J. Neural Syst.* 26:1650018. doi: 10.1142/S0129065716500180
- Huang, R.-S., Jung, T.-P., and Makeig, S. (2007). “Event-related brain dynamics in continuous sustained-attention tasks,” in *Foundations of Augmented Cognition*, eds D. D. Schmorrow and L. M. Reeves (Berlin; Heidelberg: Springer), 65–74.
- Huang, R.-S., Jung, T.-P., and Makeig, S. (2009). “Tonic changes in eeg power spectra during simulated driving,” in *HCI (16)*, volume 5638 of *Lecture Notes in Computer Science* (Hong Kong: Springer), 394–403.
- Hursh, S. R., Raslear, T. G., Kaye, A. S., and Franzone, J. F. (2006). *Validation and Calibration of a Fatigue Assessment Tool for Railroad Work Schedules, Summary Report*. U.S. Dept. of Transportation.
- Hursh, S. R., Redmond, D. P., Johnson, M. L., Thorne, D. R., Belenky, G., Balkin, T. J., et al. (2004). Fatigue models for applied research in war-fighting. *Aviat. Space Environ. Med.* 75, 44–53.
- Inoue, Y., and Komada, Y. (2014). Sleep loss, sleep disorders and driving accidents. *Sleep Biol. Rhythms* 12, 96–105. doi: 10.1111/sbr.12064
- Jensen, O., Kaiser, J., and Lachaux J. P. (2007). Human gamma-frequency oscillations associated with attention and memory. *Trends Neurosci.* 30, 317–324. doi: 10.1016/j.tins.2007.05.001
- Jiang, J.-J., Huang, Z.-G., Huang, L., Liu, H., and Lai, Y.-C. (2016). Directed dynamical influence is more detectable with noise. *Sci. Rep.* 6:24088. doi: 10.1038/srep24088



- Kaminski, M. J., and Blinowska, K. J. (1991). A new method of the description of the information flow in the brain structures. *Biol. Cybern.* 65, 203–210.
- Kennel, M. B., and Abarbanel, H. D. I. (2002). False neighbors and false strands: a reliable minimum embedding dimension algorithm. *Phys. Rev. E* 66, 1–18. doi: 10.1103/PhysRevE.66.026209
- Kozachenko, L. F., and Leonenko, N. N. (1987). Sample estimate of the entropy of a random vector. *Probl. Peredachi Inf.* 23, 9–16.
- Lacot, E., Houchmandzadeh, B., Girardeau, V., Hugon, O., and Jacquin, O. (2016). Nonlinear modification of the laser noise power spectrum induced by frequency-shifted optical feedback. *Phys. Rev. A* 94, 1–10. doi: 10.1103/PhysRevA.94.033843
- Lainscsek, C., Hernandez, M., Weyhenmeyer, J., Sejnowski, T., and Poizner, H. (2013). Non-linear dynamical analysis of eeg time series distinguishes patients with parkinsons disease from healthy individuals. *Front. Neurol.* 4:200. doi: 10.3389/fneur.2013.00200
- Lal, S., and Craig, A. (2002). Driver fatigue: electroencephalography and psychological assessment. *Psychophysiology* 39, 313–321. doi: 10.1017/S0048577201393095
- Lee, T.-W., Girolami, M., and Sejnowski, T. (1999). Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources. *Neural Comput.* 11, 417–441.
- Lin, C.-T., Chuang, C.-H., Kerick, S., Mullen, T., Jung, T.-P., Ko, L.-W., et al. (2016). Mind-wandering tends to occur under low perceptual demands during driving. *Sci. Rep.* 6:21353. doi: 10.1038/srep21353
- Lin, C.-T., Huang, K.-C., Chao, C.-F., Chen, J.-A., Chiu, T.-W., Ko, L.-W., et al. (2010). Tonic and phasic {EEG} and behavioral changes induced by arousing feedback. *Neuroimage* 52, 633–642. doi: 10.1016/j.neuroimage.2010.04.250
- Liu, J.-P., Zhang, C., and Zheng, C.-X. (2010). Estimation of the cortical functional connectivity by directed transfer function during mental fatigue. *Appl. Ergon.* 42, 114–121. doi: 10.1016/j.apergo.2010.05.008
- Lund, H. G., Reider, B. D., Whiting, A. B., and Prichard, J. R. (2010). Sleep patterns and predictors of disturbed sleep in a large population of college students. *J. Adolesc. Health* 46, 124–132. doi: 10.1016/j.jadohealth.2009.06.016
- May, J. F., and Baldwin, C. L. (2009). Driver fatigue: the importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. *Transport. Res. F Traffic Psychol. Behav.* 12, 218–224. doi: 10.1016/j.trf.2008.11.005
- McBride, J. C., Zhao, X., Munro, N. B., Jicha, G. A., Schmitt, F. A., Kryscio, R. J., et al. (2015). Sugihara causality analysis of scalp EEG for detection of early alzheimer's disease. *Neuroimage Clin.* 7, 258–265. doi: 10.1016/j.nicl.2014.12.005
- McCracken, J. M., and Weigel, R. S. (2014). Convergent cross-mapping and pairwise asymmetric inference. *Phys. Rev. E* 90:062903. doi: 10.1103/PhysRevE.90.062903
- Mullen, T. (2012). *The Source Information Flow Toolbox*. Available online at: <http://scn.ucsd.edu/wiki/SIFT> (Accessed May 24, 2018).
- Pack, A. I., Pack, A. M., Rodgman, E., Cucchiara, A., Dinges, D. F., and Schwab, C. W. (1995). Characteristics of crashes attributed to the driver having fallen asleep. *Acci. Anal. Prev.* 27, 769–775.
- Raghavachari, S., Lisman, J. E., Tully, M., Madsen, J. R., Bromfield, E. B., and Kahana, M. J. (2006). Theta oscillations in human cortex during a working-memory task: Evidence for local generators. *J. Neurophysiol.* 95, 1630–1638. doi: 10.1152/jn.00409.2005
- Russell, C. A., Arand, D., Myers, L. J., Wubbels, P., and Downs, H. (2006). *Controlling Fatigue risk With the Fatigue Science Readiband Actigraph and Associated Sleep/Wake Classification Algorithms: Use Case and Validation*. Retrieved from Fatigue Science: <https://www.fatiguescience.com/wp-content/uploads/2016/09/Readiband-Validation-Accuracy.pdf>
- Scheeringa, R., Petersson, K. M., Kleinschmidt, A., Jensen, O., and Bastiaansen, M. C. M. (2012). Eeg alpha power modulation of fmri resting-state connectivity. *Brain Connect.* 2, 254–264. doi: 10.1089/brain.2012.0088
- Schiecke, K., Pester, B., Piper, D., Feucht, M., Benninger, F., Witte, H., et al. (2017). Advanced nonlinear approach to quantify directed interactions within eeg activity of children with temporal lobe epilepsy in their time course. *EPJ Nonlinear Biomed. Phys.* 5:3. doi: 10.1051/epjnbp/2017002
- Schreiber, T. (2000). Measuring information transfer. *Phys. Rev. Lett.* 85, 461–464. doi: 10.1103/PhysRevLett.85.461
- Soldatenko, S., and Chichkine, D. (2014). Correlation and spectral properties of a coupled nonlinear dynamical system in the context of numerical weather prediction and climate modeling. *Discrete Dynam. Nat. Soc.* 2014, 1–16. doi: 10.1155/2014/498184
- Spurgeon, A., Harrington, J. M., and Cooper, C. L. (1997). Health and safety problems associated with long working hours: a review of the current position. *Occup. Environ. Med.* 54, 367–375.
- Sugihara, G., May, R., Ye, H., Hsieh, C.-H., Deyle, E., Fogarty, M., et al. (2012). Detecting causality in complex ecosystems. *Science* 338, 496–500. doi: 10.1126/science.1227079
- Tallon-Baudry, C., Bertrand, O., Peronnet, F., and Pernier, J. (1998). Induced gamma-band activity during the delay of a visual short-term memory task in humans. *J. Neurosci.* 18, 4244–4254.
- Tandle, A., Jog, N., D'cunha, P., and Chheta, M. (2016). Classification of artefacts in EEG signal recordings and eeg artefact removal using eeg subtraction. *Commun. Appl. Electron.* 4, 12–19. doi: 10.5120/cae2016651997
- Tatum, W., Husain, A. M., Benbadis, S., and Kaplan, P. (2007). *Handbook of EEG Interpretation*. New York, NY: Demos Medical Publishing.
- Toppi, J., Astolfi, L., Poudel, G. R., Innes, C. R., Babiloni, F., and Jones, R. D. (2016). Time-varying effective connectivity of the cortical neuroelectric activity associated with behavioural microsleeps. *Neuroimage* 124, 421–432. doi: 10.1016/j.neuroimage.2015.08.059
- Van Dongen, H. P., Belenky, G., and Krueger, J. M. (2011). A local, bottom-up perspective on sleep deprivation and neurobehavioral performance. *Curr. Top. Med. Chem.* 19, 2414–2422. doi: 10.2174/156802611797470286
- Wascher, E., Rasch, B., Sanger, J., Hoffmann, S., Schneider, D., Rinkenauer, G., et al. (2014). Frontal theta activity reflects distinct aspects of mental fatigue. *Biol. Psychol.* 96, 57–65. doi: 10.1016/j.biopsycho.2013.11.010

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer YA and handling Editor declared their shared affiliation.

Copyright   2018 Fonseca, Kerick, King, Lin and Jung. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# EEG-Based Mental Workload Neurometric to Evaluate the Impact of Different Traffic and Road Conditions in Real Driving Settings

Gianluca Di Flumeri<sup>1,2,3\*</sup>, Gianluca Borghini<sup>1,2,3</sup>, Pietro Aricò<sup>1,2,3</sup>, Nicolina Sciaraffa<sup>1,2,4</sup>, Paola Lanzi<sup>5</sup>, Simone Pozzi<sup>5</sup>, Valeria Vignali<sup>6</sup>, Claudio Lantieri<sup>6</sup>, Arianna Bichicchi<sup>6</sup>, Andrea Simone<sup>6</sup> and Fabio Babiloni<sup>1,3,7</sup>

<sup>1</sup> BrainSigns srl, Rome, Italy, <sup>2</sup> IRCCS Fondazione Santa Lucia, Neuroelectrical Imaging and BCI Lab, Rome, Italy,

<sup>3</sup> Department of Molecular Medicine, Sapienza University of Rome, Rome, Italy, <sup>4</sup> Department of Anatomical, Histological, Forensic and Orthopedic Sciences, Sapienza University of Rome, Rome, Italy, <sup>5</sup> Deep Blue srl, Rome, Italy, <sup>6</sup> Department of Civil, Chemical, Environmental and Materials Engineering (DICAM), School of Engineering and Architecture, University of Bologna, Bologna, Italy, <sup>7</sup> Department of Computer Science, Hangzhou Dianzi University, Hangzhou, China

## OPEN ACCESS

### Edited by:

Muthuraman Muthuraman,  
University Medical Center of the  
Johannes Gutenberg University  
Mainz, Germany

### Reviewed by:

Edmund Wascher,  
Leibniz-Institut für Arbeitsforschung  
an der TU Dortmund (IfADo),  
Germany  
Bahamn Nasserolelami,  
Trinity College Dublin, Ireland

### \*Correspondence:

Gianluca Di Flumeri  
gianluca.diflumeri@uniroma1.it;  
gluca.diflumeri@gmail.com

**Received:** 16 July 2018

**Accepted:** 05 December 2018

**Published:** 18 December 2018

### Citation:

Di Flumeri G, Borghini G, Aricò P, Sciaraffa N, Lanzi P, Pozzi S, Vignali V, Lantieri C, Bichicchi A, Simone A and Babiloni F (2018) EEG-Based Mental Workload Neurometric to Evaluate the Impact of Different Traffic and Road Conditions in Real Driving Settings.  
*Front. Hum. Neurosci.* 12:509.  
doi: 10.3389/fnhum.2018.00509

Car driving is considered a very complex activity, consisting of different concomitant tasks and subtasks, thus it is crucial to understand the impact of different factors, such as road complexity, traffic, dashboard devices, and external events on the driver's behavior and performance. For this reason, in particular situations the cognitive demand experienced by the driver could be very high, inducing an excessive experienced mental workload and consequently an increasing of error commission probability. In this regard, it has been demonstrated that human error is the main cause of the 57% of road accidents and a contributing factor in most of them. In this study, 20 young subjects have been involved in a real driving experiment, performed under different traffic conditions (rush hour and not) and along different road types (main and secondary streets). Moreover, during the driving tasks different specific events, in particular a pedestrian crossing the road and a car entering the traffic flow just ahead of the experimental subject, have been acted. A Workload Index based on the Electroencephalographic (EEG), i.e., brain activity, of the drivers has been employed to investigate the impact of the different factors on the driver's workload. Eye-Tracking (ET) technology and subjective measures have also been employed in order to have a comprehensive overview of the driver's perceived workload and to investigate the different insights obtainable from the employed methodologies. The employment of such EEG-based Workload index confirmed the significant impact of both traffic and road types on the drivers' behavior (increasing their workload), with the advantage of being under real settings. Also, it allowed to highlight the increased workload related to external events while driving, in particular with a significant effect during those situations when the traffic was low. Finally, the comparison between methodologies revealed the higher sensitivity of neurophysiological measures with respect to ET and subjective ones.

In conclusion, such an EEG-based Workload index would allow to assess objectively the mental workload experienced by the driver, standing out as a powerful tool for research aimed to investigate drivers' behavior and providing additional and complementary insights with respect to traditional methodologies employed within road safety research.

**Keywords:** electroencephalography, mental workload, human factor, machine-learning, asSWLDA, neuroergonomics, car driving, road safety

## INTRODUCTION

According to the reports of World Health Organization (WHO) (World Health Organization, 2015), every year traffic accidents cause the death of 1.3 million people around the world, and moreover about 50 million people suffer from a disability caused by accidents related to cars. By 2020, it is estimated that traffic accidents will be the fifth leading cause of death in the world, reaching 2.4 million deaths per year (World Health Organization, 2013). Among the principal causes of the car accidents and related mortality there is the human factor (Hansen, 2007; Subramanian, 2012). In particular, it has been demonstrated that human error is the main cause of the 57% of road accidents and a contributing factor in over 90% of them (Treat et al., 1979). Driver's common errors are largely correlated to overload, distractions, tiredness, or the simultaneous realization of other activities during driving (Allnutt, 1987; Horowitz and Dingus, 1992; Summala and Mikkola, 1994; Petridou and Moustaki, 2000). In fact, the human performance decrease, and consequently the errors commission, are directly attributable to aberrant mental states, in particular the mental workload while degrading in overload, which is considered one of the most important human factor constructs in influencing performance (Reason, 2000; Parasuraman et al., 2008; Paxion et al., 2014). The model theorized by De Waard (1996), widely used in automotive psychological research, establishes the relation between task demands and performance depending on the driver workload. This model describes the driving activity with a hierarchy of tasks on three levels, the *strategical*, the *tactical* and the *operational*, each of them divided into different subtasks, describing the driving as a very complex and often high-demanding activity. Therefore, the cognitive resources required in very complex situations can exceed the available resources, leading to an increase of workload and to performance impairments (Robert, 1997; Paxion et al., 2014).

The aforesaid statistics and findings justify the increasing attention received by the Human Factor within the road safety research during the last decades. As well as in other human-centered domains such as aviation and industry (Vicente, 2013; Toppi et al., 2016; Vecchiato et al., 2016; Borghini et al., 2017a), psychological disciplines have been taken on a considerable scientific importance receiving more and more attention. They have become a fundamental instrument for understanding and interpreting the behavior of the driver (Bucchi et al., 2012), trying to provide cognitive models in order to predict and avoid unsafe actions as well as to understand the relationship between such unsafe behaviors and different factors related to traffic, road complexity, car equipment and external events. The

most frequently adopted techniques in this research field are those based on questionnaires and interviews after large-scale experiments in naturalistic (i.e., real driving) and simulated (i.e., by using simulator) settings. They make it possible to acquire useful information for personality tests and profiles, they help to highlight and correct behavioral difficulties and, therefore, they shape the driver to have a safe relationship with driving in different conditions, and in particular in emergency situations, as well as to improve road and car design and adapt safety education with respect to the driver background (Cestac et al., 2014; Kaplan et al., 2015).

In order to increase the strength of such psychological research applied to road safety, this discipline could now benefit from recent advancements and outcomes coming from Neuroscience and Neuroergonomics. The field of the Neuroergonomics aims to study the relationship between the human behavior and the brain at work (Parasuraman and Rizzo, 2008). It provides a multidisciplinary translational approach that merges elements of neuroscience, cognitive psychology, human factors and ergonomics to study brain structure and function in everyday environments. Applied to the driving safety domain, a Neuroergonomic approach should allow to investigate the relationship between human mental behavior, performance and road safety, taking advantage from neurophysiological measures and providing a deeper understanding of human cognition and its role in decision making and possible error commission at the wheel (Lees et al., 2010). In fact, it is widely accepted in scientific literature the limit of using subjective measures alone, such as questionnaires and interview, because of their intrinsic subjective nature and the impossibility to catch the "unconscious" phenomena behind human behaviors (Gopher and Braune, 1984; Dienes, 2004; Wall et al., 2004; Aricò et al., 2017b). In this context, technological advancements enable the use of neurophysiological measures, for example the measure of brain activity, heart activity, eye movements, to obtain objective measures of specific mental states with low invasiveness (Aricò et al., 2017c). Among the several neuroimaging techniques, such as functional Magnetic Resonance and Magnetoencephalography, Electroencephalographic technique (EEG) has been demonstrated to be one of the best techniques to infer, even in real time, objective assessment of mental states and in particular the mental workload experienced by the user, since other than being a direct measure of brain activations, it is characterized by high temporal resolution, limited cost and invasiveness (Prinzel et al., 2000; Aricò et al., 2016b). EEG-based measures of drivers' mental states have been already investigated during the recent decades in order to determine brain cues of incoming risky psychophysical states, e.g., fatigue, drowsiness,

inattention, overload (Lin et al., 2005; Michail et al., 2008; Brookhuis and de Waard, 2010; Borghini et al., 2012, 2014; Maglione et al., 2014; Wang et al., 2015; Zhang et al., 2015; Kong et al., 2015, 2017), and to develop futuristic Human-Machine interaction solutions and automation (Kohlmorgen et al., 2007; Lin et al., 2009; Göhring et al., 2013; Aricò et al., 2015). Nevertheless, two important gaps are still present in this domain:

- (1) the majority of neurophysiological studies about drivers' behaviors have been conducted in simulated environments or in poor realistic settings, but it has been proven that same experimental tasks are perceived differently, in terms of mental workload, if performed in a simulator or in real environment (de Winter et al., 2014); also, not only the task perception but the driver behavior itself related to a specific condition could change if the same condition is reproduced in simulators or in a real scenario (Philip et al., 2005);
- (2) in scientific literature there is still the lack of a synthetic EEG-based workload index to adopt in a systematic way within the road safety research, in order to integrate results coming from traditional techniques, such as subjective measures and car parameters analysis, with additional insights arising directly from drivers' brain (Paxion et al., 2014). Several studies about EEG correlates of driver's mental workload have been carried on, however experimental examples in real settings of a multimodal approach integrating neurophysiological with traditional measures are still lacking (Xing et al., 2018).

In this study, it has been investigated the possibility to adopt the approach recently developed and patented by the authors of this work (Aricò et al., 2016b, 2017a), to evaluate the mental workload experienced by car drivers by means of their EEG activity. More specifically, such an approach is based on a machine-learning method able to assess, even online and in high-realistic environments, the user's mental workload through a synthetic index. The authors successfully employed and validated such approach in different aviation-related applications, such as adaptive automation (Aricò et al., 2016a), personnel training (Borghini et al., 2017c), personnel expertise evaluation (Borghini et al., 2017b), moreover highlighting the higher sensitivity of such measures compared with subjective ones (Di Flumeri et al., 2015; Aricò et al., 2016b). Furthermore, the feasibility of obtaining EEG-based measures of driver's workload has already been validated through a pilot study of the present work conducted with eight subjects while performing a simplified version of the real driving task employed within the present work (Di Flumeri et al., 2018).

For the present work, 20 young subjects have been involved in a real driving task along urban roads, performed under different traffic conditions (rush hour and not) and going through different road types (main and secondary streets). Also, during the driving tasks specific events, in particular a pedestrian crossing the road and a car entering the traffic flow just ahead of the experimental subject, have been acted. During the experiments the drivers' brain activity, through EEG technique, and eye movements, through Eye-Tracking (ET) devices, have been collected. In

addition, subjective measures, car parameters (e.g., position, speed, etc.) and videos around the car have been gathered. Thanks to this multimodal approach, the present study aimed at:

- Validating the machine-learning approach developed by the authors also in automotive domain, through an experiment in high-realistic settings, i.e., real driving;
- Employing the EEG-based Workload index obtained from the hence validated approach to evaluate the impact of different factors, specifically the road complexity, the traffic intensity (depending on the hour of the day), and two specific events (a pedestrian crossing the road and a car entering in the traffic flow), on the drivers' mental workload;
- Comparing the neurophysiological measures with eye movements and subjective ones, in order to provide evidence of the complementarity of the obtained insights.

In conclusion, the present work will explore the potential of integrating these new methodologies, i.e., neurophysiological measures, with traditional approaches in order to enhance and extent research on drivers' behaviors and road safety.

## MATERIALS AND METHODS

### The Experimental Protocol

Twenty male students ( $24.9 \pm 1.8$  years old, licensed from  $5.9 \pm 1$  years, with a mean annual mileage of 10350 km/year) from the University of Bologna (Italy) have been recruited and involved on a voluntary basis in this study. They were selected in order to have a homogeneous experimental group in terms of age, sex, and driving expertise. The experiment was conducted following the principles outlined in the Declaration of Helsinki of 1975, as revised in 2000. Informed consent and authorization to use the video graphical material were obtained from each subject on paper, after the explanation of the study.

Two equal cars have been used for the experiments, i.e., Fiat 500L 1.3 Mjt, with diesel engine and manual transmission. The subjects had to drive the car along a route going through urban roads at the periphery of Bologna (Italy). In particular, the route consisted in three laps of a "circuit" about 2500 m long to be covered with the daylight (**Figure 1**).

The circuit was designed with the aim to include two segments of interest, both about 1000 m long but different in term of road complexity and so supposed different also in terms of cognitive demand, thus named hereafter "*Easy*" and "*Hard*": (i) *Easy* was a secondary road, mainly straight, with an intersection halfway with the right-of-way, one lane and low traffic capacity, serving a residential area; (ii) *Hard* was a main road, mainly straight, with two roundabouts halfway, three lanes and high traffic capacity, serving a commercial area. This factor will be hereafter named "*ROAD*." This assumption has been made on the basis of several evidences coming from scientific literature about road safety and behavior (Harms, 1991; Verwey, 2000; Paxion et al., 2014).

Furthermore, each subject had to repeat the task two times within the same day, one time during rush and one during normal hour: this factor will be hereafter named "*HOURL*," while the





**TABLE 1 |** Data extracted from the General Plan of Urban Traffic of Bologna (Italy) referred to the traffic flow intensities in the experimental area during the day.

| Transits                     | Total<br>14 h (6 ÷ 20) | RUSH hour                |                            | NORMAL<br>12 h |
|------------------------------|------------------------|--------------------------|----------------------------|----------------|
|                              |                        | Morning<br>(12:30–13:30) | Afternoon<br>(16:30–17:30) |                |
| Total                        | 19385                  | 2024                     | 2066                       | 15295          |
| Frequency<br>(Transits/hour) | –                      | 2024                     | 2066                       | 1274,6         |

*These data have been used to design two experimental conditions different in terms of traffic: the RUSH hours are characterized by traffic higher than during NORMAL hours.*

two conditions “*Rush*” and “*Normal*.” The rush hours of that specific area have been determined according to the General Plan of Urban Traffic of Bologna (PGTU, please see **Table 1**): the two “Rush hour” time-windows were from 12:30 to 13:30 (lunchtime) and from 16:30 to 17:30 (work closing time), with the experiments performed from 9.30 to 17.30, in order to ensure a homogeneous daylight condition.

Finally, during the last lap (i.e., the 3rd one) of each task repetition (i.e., Rush and Normal hour) two different events have been simulated, by involving actors, twice (i.e., along the Hard and the Easy circuit segment) along the route: a pedestrian crossing the road, and a car entering the traffic flow just ahead of the experimental subject, hereafter labeled respectively

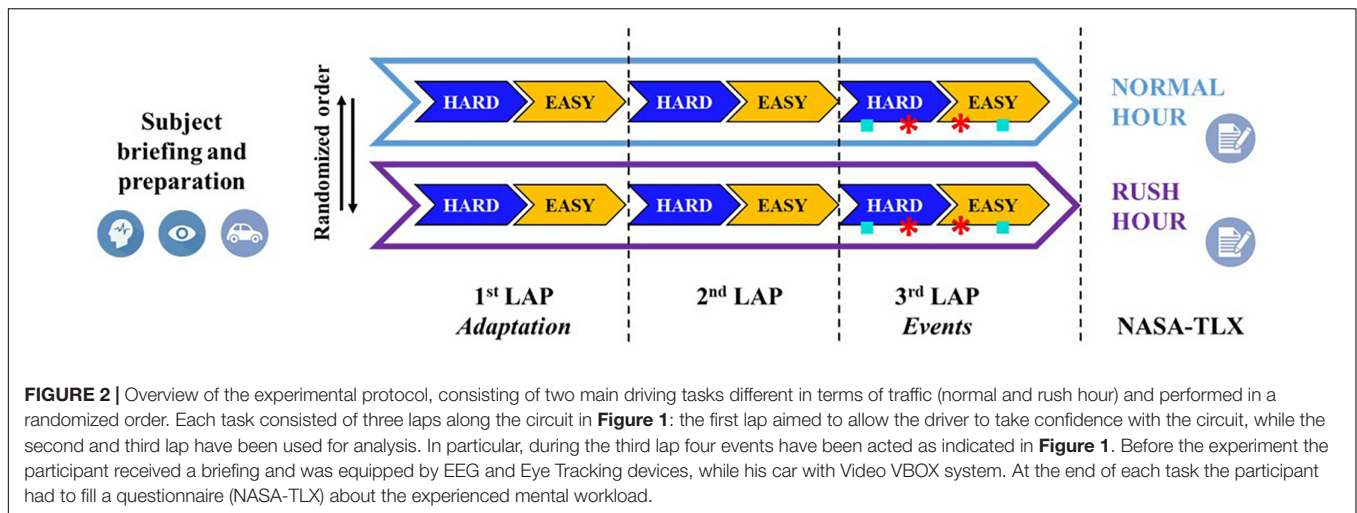
“*Pedestrian*” and “*Car*.” The event types have been selected as the most probable events coherently with the urban context, as well as the safest to act, i.e., without introducing any risk for the actors, for the experimental subjects and for the traffic in general.

The **Figure 1** shows the experimental circuit along Bologna roads, highlighting the “ROAD complexity” distribution as well as the occurred events.

To summarize, each subject, after a proper experimental briefing, performed a driving task of three laps along a circuit through urban roads two times, during Rush and Normal hours. The order of Rush and Normal conditions has been randomized among the subjects, in order to avoid any order effect (Kirk, 2015). Each lap consisted in a Hard and an Easy segment, where hard and easy are referred to the road complexity and thus task difficulty. Also, despite the initial briefing, the first lap of both the tasks has been considered an “adaptation lap,” while the data recorded during the second and third laps have been taken into account for the analysis. Finally, during the third lap two equal events have been simulated both along the Easy and the Hard segment (i.e., four events in total for each subject for each task, Rush and Normal).

The **Figure 2** shows a graphical representation of the experimental protocol.

During the whole protocol physiological data, in terms of brain activity through Electroencephalographic (EEG) technique and eye gazes through ET devices, and data about driving



behavior, through a professional device mounted on the car (i.e., a VBOX Pro), have been recorded. In addition, subjective measures of perceived Mental Workload have been collected from the subjects after both the tasks through the NASA Task Load Index (NASA-TLX) questionnaire (Hart and Staveland, 1988). It was possible to use Eye Tracker just with half of the subjects' sample (i.e., 10 subjects) because of device availability, so eye tracker-related data have been analyzed for 10 subjects. The following paragraphs will describe in detail the collection and processing of the aforementioned data, while the **Figure 3** shows the subject preparation and the recording setup within the car.

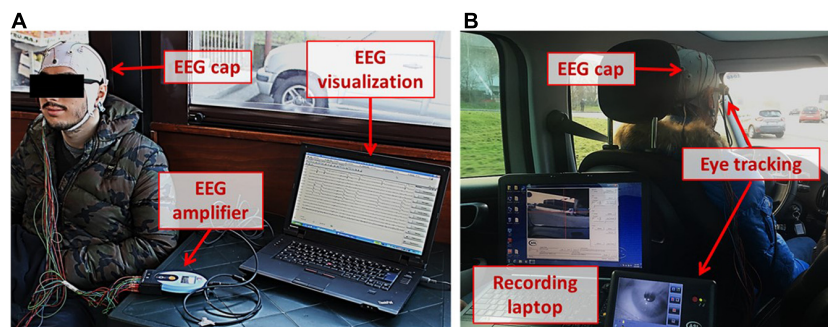
## The Data Collection

### Electroencephalographic Signal Recording and Processing

The EEG signals have been recorded using the digital monitoring BEmicro system (EBNeuro, Italy). Twelve EEG channels (FPz, AF3, AF4, F3, Fz, F4, P3, P7, Pz, P4, P8, and POz), placed according to the 10–20 International System, were collected with a sampling frequency of 256 Hz, all referenced to both the

earlobes, grounded to the Cz site, and with the impedances kept below 20 k $\Omega$ . During the experiments the EEG data have been recorded without any signal conditioning, the whole processing chain has been applied offline. In particular, EEG signal has been firstly band-pass filtered with a fourth-order Butterworth filter (high-pass filter cut-off frequency: 1 Hz, low-pass filter cut-off frequency: 30 Hz). The Fpz channel has been used to remove eyes-blink contributions from each channel of the EEG signal by using the REBLINCA algorithm (Di Flumeri et al., 2016). This step is necessary because the eyes-blink contribution could affect the frequency bands correlated to the mental workload, in particular the theta EEG band. This method allows to correct EEG signal without losing data.

For other sources of artifacts (i.e., environmental noise, drivers' movements, etc.), specific procedures of the EEGLAB toolbox (Delorme and Makeig, 2004) have been employed. Firstly, the EEG signal is segmented into epochs of 2 s (Epoch length), through moving windows shifted of 0.125 s (Shift), thus with an overlap of 0.875 s between two contiguous epochs. This windowing has been chosen with the compromise to have both a high number of observations, in comparison with the number



**FIGURE 3 |** On the left (**A**), the participant preparation phase. In particular, the EEG signal has been acquired through the EEG amplifier in *holter* modality: the EEG signal and the electrodes impedances were checked on a computer before starting the experiments. On the right (**B**), a picture representing the experimental setup within the car: in particular, other than the EEG cap, also the Eye Tracking device and its recording laptop are shown. The subject in picture, as all the participants, gave their signed authorization to use the video graphical material for dissemination purposes.

of variables, and to respect the condition of stationarity of the EEG signal (Elul, 1969). In fact, this is a necessary assumption in order to proceed with the spectral analysis of the signal. The EEG epochs with the signal amplitude exceeding  $\pm 100 \mu\text{V}$  (*Threshold criterion*) are marked as “artifact.” Then, each EEG epoch has been interpolated in order to check the slope of the trend within the considered epoch (*Trend estimation*). If such a slope is higher than  $10 \mu\text{V/s}$ , the considered epoch is marked as “artifact.” Finally, the signal sample-to-sample difference (*Sample-to-sample criterion*) has been analyzed: if such a difference, in terms of absolute amplitude, is higher than  $25 \mu\text{V}$ , i.e., an abrupt variation (no-physiological) happened, the EEG epoch is marked as “artifact.” At the end, the EEG epochs marked as “artifact” have been removed from the EEG dataset with the aim to have a clean EEG signal to perform the analyses.

From the clean EEG dataset, the Power Spectral Density (PSD) has been calculated for each EEG channel for each epoch using a Hanning window of the same length of the considered epoch (2 s length, that means 0.5 Hz of frequency resolution). Then, the EEG frequency bands of interest has been defined for each subject by the estimation of the *Individual Alpha Frequency* (IAF) value (Klimesch, 1999). In order to have a precise estimation of the alpha peak and, hence of the IAF, the subjects were been asked to keep the eyes closed for a minute before starting the experimental tasks. Finally, a spectral features matrix (EEG channels  $\times$  Frequency bins) has been obtained in the frequency bands directly correlated to the mental workload. In particular, only the theta band [ $\text{IAF} - 6 \div \text{IAF} - 2$ ], over the EEG frontal channels, and the alpha band [ $\text{IAF} - 2 \div \text{IAF} + 2$ ], over the EEG parietal channels, were considered as variables for the mental workload evaluation (Gevins and Smith, 2003; Aricò et al., 2016b; Borghini et al., 2017a).

At this point the automatic-stop-StepWise Linear Discriminant Analysis (asSWLDA), a specific Machine-Learning algorithm (basically an upgrade version of the well-known StepWise Linear Discriminant Analysis) previously developed (Aricò et al., 2016b), patented (Aricò et al., 2017a) and applied in different applications (Aricò et al., 2016a; Borghini et al., 2017b,c) by the authors has been employed. On the basis of the calibration dataset, the asSWLDA is able to find the most relevant spectral features to discriminate the Mental Workload of the subjects during the different experimental conditions (i.e., EASY = 0 and HARD = 1). Once identified such spectral features, the asSWLDA assigns to each feature specific weights ( $w_{i \text{ train}}$ ), plus a bias ( $b_{\text{train}}$ ), such that an eventual discriminant function computed on the training dataset [ $y_{\text{train}}(t)$ ] would take the value 1 in the hardest condition and 0 in the easiest one. This step represents the calibration, or “*Training phase*” of the classifier. Later on, the weights and the bias determined during the training phase are used to calculate the Linear Discriminant function [ $y_{\text{test}}(t)$ ] over the testing dataset (*Testing phase*), that should be comprised between 0 (if the condition is Easy) and 1 (if the condition is Hard). Finally, a moving average of 8 s (8MA) is applied to the  $y_{\text{test}}(t)$  function in order to smooth it out by reducing the variance of the measure: its output is defined as the *EEG-based Workload index* ( $WL_{\text{SCORE}}$ ). For the present work, the training data consisted in the *Easy* segment of the

2nd lap during the *Normal* condition and the *Hard* segment of the 2nd lap during the *Rush* condition (they have been hypothesized the two conditions characterized by respectively the lowest and highest mental workload demand), while the testing data consisted of the data of the 3rd lap of both the conditions.

Here below the training asSWLDA discriminant function (Equation 1, where  $f_{i \text{ train}}(t)$  represents the PSD matrix of the training dataset for the data window of the time sample  $t$ , and of the  $i^{\text{th}}$  feature), the testing one (Equation 2, where  $f_{i \text{ test}}(t)$  is as  $f_{i \text{ train}}(t)$  but related to the testing dataset) and the equation of the *EEG-based workload index* computed with a time-resolution of 8 s ( $WL_{\text{SCORE}}$ , Equation 3), are reported.

$$y_{\text{train}}(t) = \sum_i w_{i \text{ train}} \cdot f_{i \text{ train}}(t) + b_{\text{train}} \quad (1)$$

$$y_{\text{test}}(t) = \sum_i w_{i \text{ train}} \cdot f_{i \text{ test}}(t) + b_{\text{train}} \quad (2)$$

$$WL_{\text{SCORE}} = 8MA(y_{\text{test}}(t)) \quad (3)$$

## Eye-Tracking Data and Its Processing

Eye movements of the participants have been recorded through an ASL Mobile Eye-XG device (EST GmbH, Germany), a system based on lightweight eyeglasses equipped with two digital high-resolution cameras. One camera recorded the scene image and the other the participant's eye, that is monitored through infrared rays. The data were recorded with a sampling rate of 30 Hz (i.e., 33 ms time resolution), and a spatial resolution of  $0.5 \div 1^\circ$ . ASL software was used to analyze the data, obtaining information about the drivers' fixation points frame by frame (33 ms). A preliminary calibration procedure was carried out for each subject inside the car before starting driving, asking them to fix their gaze on thirty fixed visual points spread across the whole scene, in order to get a good accuracy of the eye-movement recorder. The gazes recorded during the driving task were manually analyzed, in order to group them into three different categories: road infrastructure, traffic vehicles, and external environment. For each subject, each lap (second and third), and each condition (Easy and Hard ROAD, Rush and Normal HOUR) the distribution of eye fixations between the three categories was calculated in terms of percentage of the total.

## Additional Measures

Each car has been equipped with a Video VBOX Pro (Racelogic Ltd., United Kingdom), a system able to continuously monitor the cinematic parameters of the car, integrated with GPS data and videos coming from up to four high-resolution cameras. The system has been fixed within the car, at the center of the floor of the back seats, in order to put it as close as possible to the car barycenter, while two cameras have been fixed over the top of the car. The system recorded car parameters (e.g., speed, acceleration, position, etc.) with a sampling rate of 10 Hz. For the purpose of the present study, the average speed for each task has been computed. Also, the cameras' videos have been used to count the number of vehicles encountered by the driver during each task.



Also, at the end of each task (thus only the HOUR condition, i.e., Rush vs. Normal, can be compared) the subjects had to evaluate the experienced workload by filling the NASA-TLX questionnaire (Hart and Staveland, 1988). In particular, the subject had (i) to assess, on a scale from 0 to 100, the impact of six different factors (i.e., Mental demand, Physical demand, Temporal demand, Performance, Effort, Frustration), and (ii) to assess the more impacting factor through 15 comparisons between couple of the previously evaluated factors. The result of this questionnaire is a score from 0 to 100 corresponding to the driver's mental workload perception.

## Performed Analyses

### Validation of Experimental Design Assumptions

The first analysis aimed to validate the assumptions in terms of experimental design, that is:

- (i) The subjects drove during two conditions different in terms of traffic, i.e., Rush and Normal hour;
- (ii) The circuit was constituted by two segments different in terms of road complexity, thus in terms of difficulty, i.e., Hard and Easy.

In order to validate the first assumption, the number of vehicles encountered by the experimental subjects and the average driving speed during the two conditions have been computed and statistically compared. It is expected that the number of vehicles is significantly higher and the average speed significantly lower during rush hours (Bucchi et al., 2012).

The second assumption has been validated by investigating the percentage of fixations over the external environment, since such indicator has been proven to be inversely correlated with mental workload: the more the experienced workload is, the less the number of fixations over the external environment is, since the driver gaze will mostly focus on infrastructure and vehicles (Costa et al., 2014; de Winter et al., 2014; Lantieri et al., 2015). Also, we verified the difference in terms of mental workload from a neurophysiological point of view: we computed the ratio between Theta rhythms over frontal sites ("*ThetaF*") and Alpha rhythms over parietal sites ("*AlphaP*"), since it is considered a well-established metric of mental workload (Borghini et al., 2014). In particular, The *ThetaF/AlphaP* has been proven to increase if the mental workload experienced by the user is increasing as well (Gevins and Smith, 2003; Holm et al., 2009; Borghini et al., 2015). The metric has been computed as the ratio between the averaged PSD values in theta band over the frontal electrodes (AF3, AF4, F3, Fz, F4) and the averaged PSD values in alpha band over the parietal electrodes (P3, P7, Pz, P4, P8, POz). Both the analysis have been performed comparing the two conditions employed to train the classifier (please see Electroencephalographic Signal Recording and Processing), i.e., the *Easy* segment of the 2nd lap during the *Normal* condition and the *Hard* segment of the 2nd lap during the *Rush* condition, assumed as the two conditions characterized by respectively the lowest and highest mental workload demand.

All the statistical comparisons have been performed through two-sided Wilcoxon signed rank tests. In fact, data come from multiple observations on the same subjects, but it is not possible to assume or robustly assess (the number of observations is always equal or less than 16) that the observations distribution is Gaussian, therefore paired non-parametric tests have been used (Siegel, 1956).

### Classification Performance

Firstly, a synthetic analysis of the brain features selected by the algorithm has been performed in order to evaluate any eventual recurrence of a specific feature. The initial features domain for each subject consisted in a matrix of 187 features (11 EEG channels \* 17 bins of frequency – from IAF-6 Hz to IAF+2 Hz with a resolution of 0.5 Hz –). Actually, only 99 of these features can be selected by the algorithm because of the Regions of Interest defined *a priori*: 45 features related to frontal Theta and 54 related to parietal Alpha.

Then, in order to investigate the algorithm (i.e., the asSWLDA) classification accuracy, the analysis of the Area Under Curve (AUC) of the Receiver Operator Characteristic (ROC) curve of the classifier has been performed (Bamber, 1975). In particular, AUC represents a widely used methodology to test the performance of a binary classifier: the classification performance can be considered good with an AUC higher than at least 0.7 (Fawcett, 2006). In this case there are actually two classes in terms of mental workload, i.e., Easy and Hard, related to the two different difficulty levels characterizing the circuit. As previously described, for each subject the training dataset consisted in the Easy segment of the 2nd lap during the Normal condition and the Hard segment of the 2nd lap during the Rush condition (they have been hypothesized the two conditions characterized by respectively the lowest and highest cognitive demand), while the testing dataset consisted of the data of the 3rd lap of both the conditions (*Real data*). Therefore, the classifier has been tested shuffling the testing dataset related labels (*Random*), in order to verify that classifier performance on measured data (*Real data*) was significantly higher than that one obtained on random data (*Random*), independently from the traffic intensity (i.e., both in Rush and Normal hour conditions). In both the cases (Real and Random), the time resolution of  $WL_{scores}$  is equal to 8 s, obtained as the best compromise between a high time resolution and good classification performance. Three two-sided Wilcoxon signed rank tests have been performed between *Real* and *Random data*, one for each HOUR condition (i.e., comparison Real vs. Random in *Normal* and *Rush* hour) and one comparing the *Normal* and *Rush* conditions only in terms of real data. The results of these multiple comparisons have been validated by applying the False Discovery Rate (FDR) correction (Benjamini and Hochberg, 1995).

### Workload Assessment

Once demonstrated the reliability of the classification algorithm to obtain the EEG-based index of mental workload in the specific driving scenarios, the workload scores (*WL score*) have been used to evaluate the impact of different factors, that is the road



complexity and the traffic as well as specific events along the driving experience. Depending on the analysis, the EEG-based WL scores have been analyzed in relation to ET and subjective data.

#### Evaluation of traffic and road complexity impact

The WL indexes obtained with a time resolution of 8 s from the testing dataset (i.e., the third lap) were averaged for each subject and for each condition (i.e., HOUR and ROAD). A Friedman test, the non-parametric version of the repeated measures ANOVA (Analysis of Variance), has been performed in order to investigate any possible effect due to traffic and road complexity on the workload perceived by the subject. Furthermore, since *post hoc* tests specifically designed for Friedman test do not exist but both the factors have been measured on the same subjects, two Wilcoxon signed rank tests have been performed in order to investigate potential within effects among the two factors, i.e., HOUR and ROAD.

Also, the results in terms of workload indexes have been compared with those obtained from ET in order to evaluate the different sensitivity to the phenomenon (i.e., mental workload variations) of the two technologies. In terms of ET measures, it has been investigated the percentage of fixations on the road infrastructure and vehicles, since such indicator has been proven to be directly correlated with mental workload while driving: the more the experienced workload is, the more the number of fixations over the road will be, since the driver gaze will mostly focus on infrastructure and vehicles (Costa et al., 2014; de Winter et al., 2014). Multiple two-sided Wilcoxon signed rank tests have been performed in order to reveal any difference with respect to the two investigated factors.

Furthermore, a two-sided Wilcoxon signed rank test has been performed on the NASA-TLX measures. Please note that for the continuity of the experiment the questionnaires were filled by the subjects only after the tasks end, therefore only the comparison between Normal and Rush hour has been possible (please refer to Section "Additional Measures").

#### Evaluation of single events impact

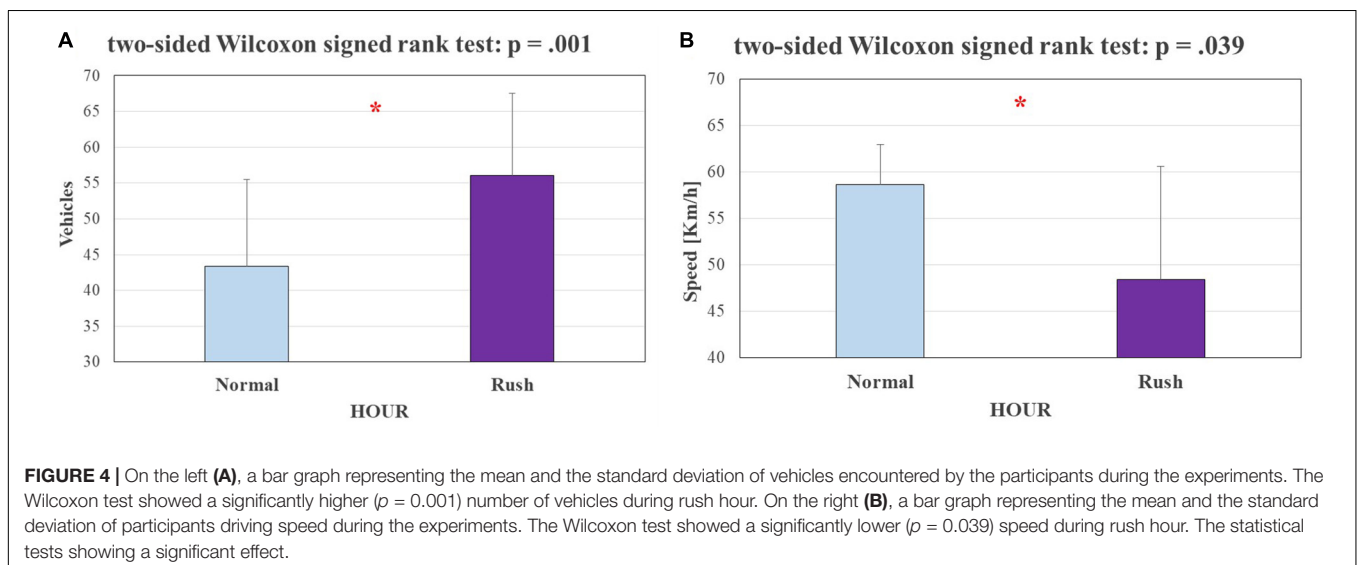
On the basis of the average duration of the events among the subjects during the driving experience, and to homogenize the measures with respect of this parameter (i.e., event duration), a fixed window of 20 s for the car event (from the first fixation of the car to its overtaking) and of 10 s for the pedestrian event (from the first fixation of the pedestrian to the acceleration after its road crossing) has been defined, independently from the traffic and the road complexity. Remembering that the events were acted only during the third lap of each task repetition, similar windows corresponding to the same circuit position were defined during the second lap in order to compare the event's happening vs. no-happening. The WL indexes were averaged for each subject, for each condition (i.e., HOUR and ROAD) and for each event. Multiple two-sided Wilcoxon signed rank tests have been performed in order to reveal any difference (i) with respect to the events' happening, and (ii) among the events types.

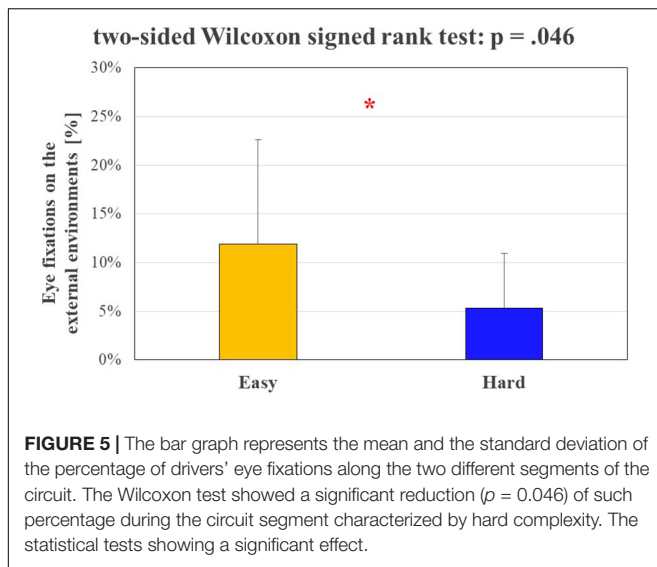
## RESULTS

The following results are referred to a sample of 16 subjects (8 with Eye Tracking), since one subject has been discarded because of technical issues on the EEG data, while three subjects have been discarded because of no objective difference in terms of encountered vehicles (measured through the VBOX cameras) between the two tasks, i.e., during Rush and Normal hours.

#### Experimental Design Validation

Figure 4 shows the results of the comparisons between (a) the number of vehicles encountered by the experimental subjects and (b) the average driving speed during the two different traffic conditions, i.e., during *Normal* and *Rush* hours. The performed statistical analysis revealed a significant increasing ( $p = 0.001$ ) of vehicles encountered by the experimental subjects and a significant decreasing ( $p = 0.039$ ) of driving average speed from

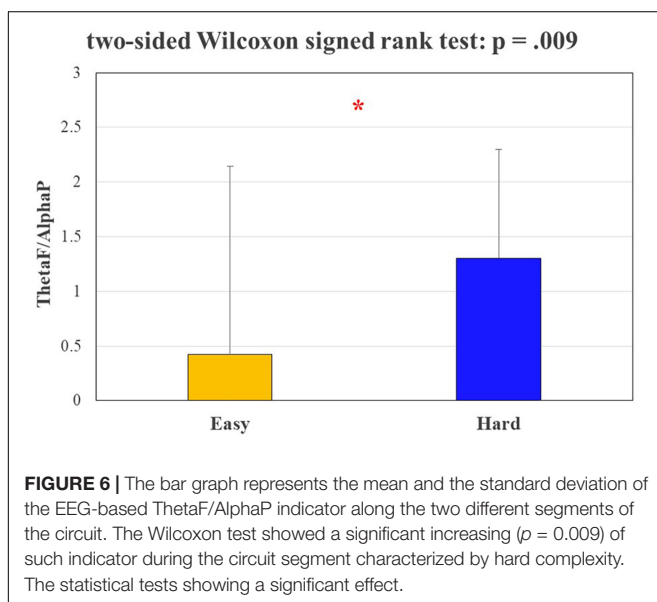




*Normal* to *Rush* hours, validating the experimental hypothesis about the two different conditions of traffic made *a priori* on the basis of the General Plan of Urban Traffic of Bologna (see The Experimental Protocol).

**Figure 5** shows the results in terms of percentage of fixations over the external environment between the *Easy* and *Hard* segments of the circuit, since such indicator has been proven to be inversely correlated with mental workload. The performed statistical analysis revealed a significant decreasing ( $p = 0.046$ ) of driver gazes over the external environment, validating the experimental hypothesis about the two different conditions of difficulty made *a priori* on the basis of scientific literature (see The Experimental Protocol).

**Figure 6** shows the results in terms of *ThetaF/AlphaP* value between the *Easy* and *Hard* segments of the circuit, since



such ratio has been proven to be a physiological indicator directly correlated to mental workload. The performed statistical analysis revealed a significant increasing ( $p = 0.009$ ) of the proposed index, validating the assumption about the different cognitive demand related to the two conditions, made *a priori* on the basis of scientific literature (see The experimental Protocol).

## Classification Performance

**Figure 7** shows the distribution of the features, and the relative frequency of selection, chosen by the asSWLDA during the training phase. The analysis of features selected by the algorithm revealed that the asSWLDA selected on average 4 features per subject, coming from 3 of the 11 channels available. The frequency bins, actually equal to 17 because included between IAF-6 Hz and IAF+2 Hz with a resolution of 0.5 Hz, have been grouped into four areas of interest: Lower Theta [IAF - 6 ÷ IAF - 4], Upper Theta [IAF - 4 ÷ IAF - 2], Lower Alpha [IAF - 2 ÷ IAF] and Upper Alpha [IAF ÷ IAF + 2]. The results show that Lower Theta over F4 and Upper Alpha over POz have been used for more than the 50% of subjects.

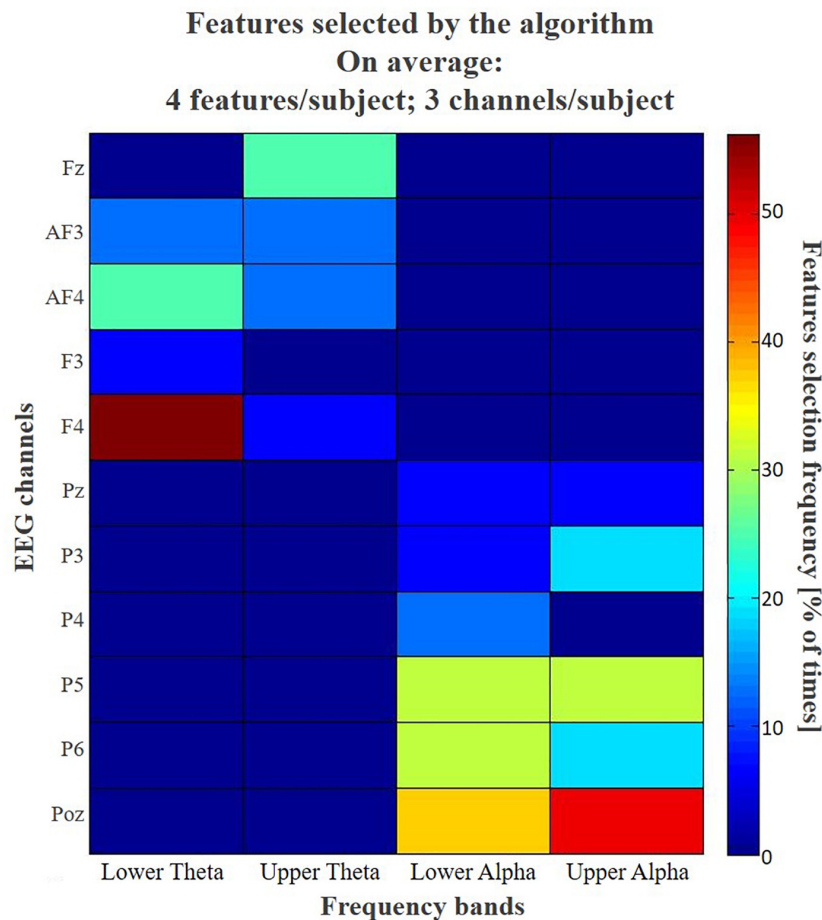
The AUC analysis (**Figure 8**) revealed that, by using such approach, it has been possible to achieve mean AUC values of  $0.744 \pm 0.13$  for the *Normal* hour and of  $0.727 \pm 0.06$  for the *Rush* hour. In particular, the two Wilcoxon tests demonstrated that the classifier performance on the Real data was significantly higher than on Random data in both the conditions (respectively  $p = 0.01$  and  $p = 0.0005$ ). Also, there were no significant differences ( $p = 0.64$ ) in terms of AUC values on Real data between *Normal* and *Rush* hours, in other words the classification performance was not dependent on the traffic condition. Because of the three repeated tests, the False Discovery Rate correction has been performed: with respect to the  $p$ -values obtained and ordered (0.0005, 0.01, and 0.64), the three corrected  $q$ -values are respectively 0.0015, 0.015, and 0.64, thus the first two results are still significant.

## Workload Assessment

### Evaluation of Traffic and Road Complexity Impact

**Figure 9** shows the results of the non-parametrical statistical analysis in terms of effects of the two investigated factors, i.e., the traffic (HOUR) and the road complexity (ROAD), on the mental workload experienced by the drivers. In particular, the Friedman test at the top of **Figure 9A** highlights a significant main effect ( $p = 0.00001$ ) among the different factors: the mental workload significantly increased because of the higher road complexity (i.e., from Easy to Hard), and even more because of the higher traffic intensity (i.e., from Normal to Rush hours). The Wilcoxon tests performed in order to investigate any *within* effect showed two significant main effects in term of workload increasing if both complexity [bottom left (**Figure 9B**), ROAD,  $p = 0.0038$ ] and traffic [bottom right (**Figure 9C**), HOUR,  $p = 0.0032$ ] increase.

**Figures 10, 11** show the results of the Wilcoxon tests comparing the sensitivity of ET measures with respect to EEG-based ones. For these analyses the EEG-based WL scores of only the subjects wearing also the Eye Tracker (eight of sixteen) have been considered, in order to make the results comparable (i.e.,



**FIGURE 7 |** The colormap summarizes the frequency of the selection of each feature for the whole subjects' sample. The initial features domain for each subject consisted of a matrix of 187 features (11 EEG channels \* 17 bins of frequency – from IAF – 6 Hz to IAF+2 Hz with a resolution of 0.5 Hz –). Actually, only 99 of these features can be selected by the algorithm, because of the Regions of Interest defined *a priori*: 45 features related to frontal Theta and 54 related to parietal Alpha. For a synthetic and effective representation, the frequency bins, actually equal to 17 because included between IAF-6 Hz and IAF+2 Hz with a resolution of 0.5 Hz, have been grouped into four areas of interest: Lower Theta [IAF – 6 ÷ IAF – 4], Upper Theta [IAF – 4 ÷ IAF – 2], Lower Alpha [IAF – 2 ÷ IAF] and Upper Alpha [IAF ÷ IAF + 2]. The results show that Lower Theta over F4 and Upper Alpha over POz have been used for more than the 50% of subjects.

both the measures have been collected during same experience). In particular:

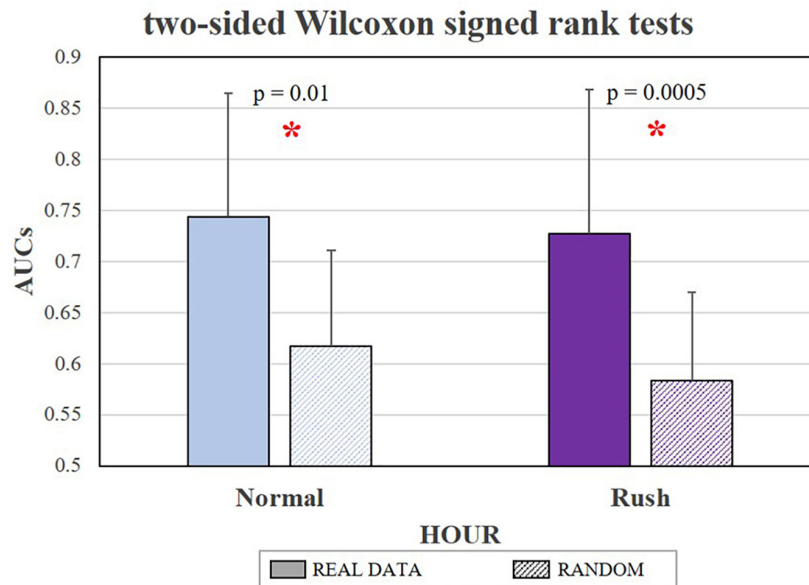
- **Figure 10:** in terms of ROAD complexity, while the EEG-based measures have been able to significantly discriminate ( $p = 0.008$ ) the two conditions at least during Normal hour, the ET-based ones have not been able to show any significant difference both during Normal and Rush hours;
- **Figure 11:** in terms of traffic HOUR, while the EEG-based measures have been able to significantly discriminate the two conditions both along Easy ( $p = 0.019$ ) and Hard ( $p = 0.039$ ) segments, the ET-based ones have been able to significantly discriminate Normal and Rush hours only along the Hard segment ( $p = 0.0192$ ).

Finally, **Figure 12** shows the results in terms of NASA-TLX scores, revealing that there is not any significant difference in terms of workload subjectively assessed between the *Normal* and *Rush* hour conditions.

### Evaluation of Single Events Impact

**Figure 13** shows the results in terms of EEG-based WL scores about how the presence of a specific event impacts the mental workload of the driver, with respect to the different experimental conditions. In terms of external events (the condition *EVENT* is referred to the event actually happened during the 3rd lap, the condition *NO EVENT* is referred to the same circuit portion during the 2nd lap when no events were acted), the pedestrian crossing the road induced a significantly higher workload only during the *Normal* hour along the *Hard* circuit segment (Wilcoxon test's  $p = 0.037$ ), while the car induced a significantly higher workload along both the *Easy* and *Hard* circuit segments but only during *Normal* hour (respectively Wilcoxon test's  $p = 0.007$  and  $p = 0.008$ ).

Considering only the condition “EVENT,” despite a decreasing trend from *Easy* to *Hard* segments, no significant differences ( $p > 0.05$ ) have been found for each event during the same traffic



**FIGURE 8 |** The bar graph represents the mean and the standard deviation of AUC values obtained in discriminating the Easy and Hard circuit segments. Both in Normal and Rush hour conditions, the classification performance obtained by training the classifier with real data (solid color) have been significantly higher (respectively  $p = 0.01$  and  $p = 0.0005$ ) than by using random data (lines pattern), achieving mean AUC values of respectively 0.74 and 0.73. The statistical tests showing a significant effect.

condition (HOUR). However, if considering the same difficulty level (ROAD), all the events induced a significant workload increasing during *Rush* hours, except the pedestrian along the Easy segment (Pedestrian *Hard*:  $p = 0.009$ ; Car *Easy*:  $p = 0.023$ ; Car *Hard*:  $p = 0.002$ ).

## DISCUSSION

Since the impact of drivers' errors in terms of human lives and costs is very high and the next future previsions are even worse (World Health Organization, 2015), the relationship between human errors and driving performance impairment due to a high mental workload has been deeply investigated in the automotive domain. Recent technological advancements as well as the growth of disciplines such as Neuroscience and Neuroergonomics now allow to record human neurophysiological signals, such as in this study brain activity through Electroencephalographic technique, in a robust way also outside the laboratory, and to obtain from them objective neurometrics of human mental states (i.e., workload) (Aricò et al., 2017c, 2018). The present work aimed to validate a machine-learning approach, i.e., the asSWLDA (Aricò et al., 2016a), for the objective assessment of human mental workload while driving in real settings, as well as its integration with traditional tools (e.g., questionnaires, car parameters, eye tracking) in order to evaluate the impact of different factors (road complexity, traffic intensity, external events), thus suggesting new innovative tools for enhancing research in road safety. In order to achieve these objectives, 20 young subjects have been involved in a real driving task along urban roads, performed under different traffic conditions (rush hour and not), driving through different

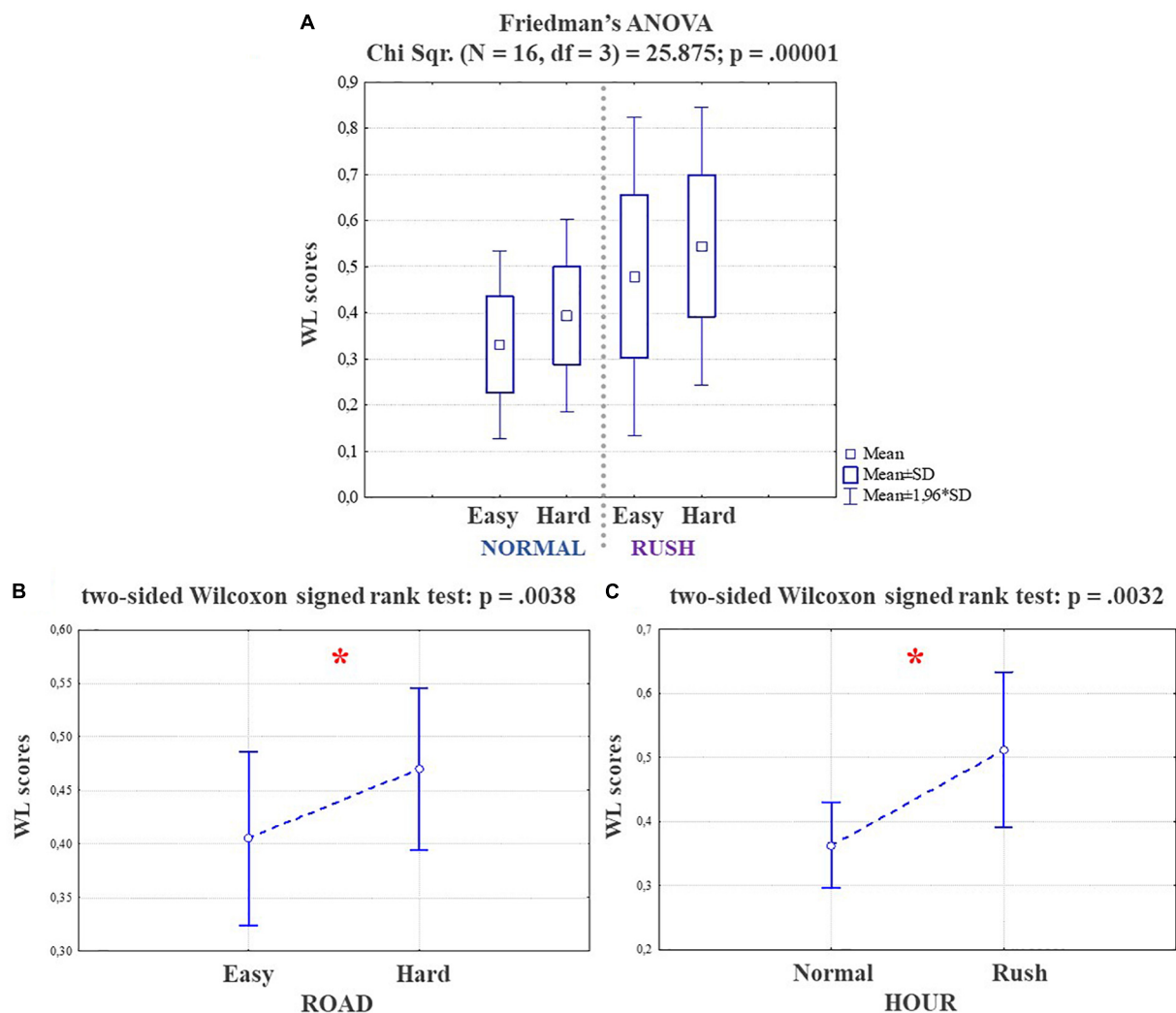
road types (main and secondary streets) and facing to external events.

Firstly, the experiments have been designed making two *a priori* assumptions:

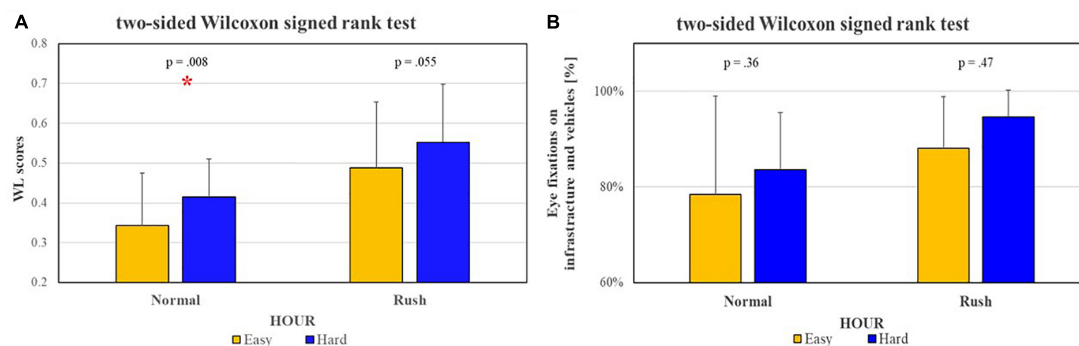
- (1) the experiments have been conducted in two different conditions of traffic intensity, depending on the hours (i.e., normal and rush hour) of the day; the experimental design initially referred to the General Plan of Urban Traffic of Bologna;
- (2) the circuit consisted of two segments of different difficulty, i.e., Easy and Hard, because of the related road complexity (Harms, 1991; Verwey, 2000; Paxion et al., 2014).

The statistical analysis performed on the average speed of the experimental subjects and the number of vehicles encountered during the experiments (Figure 4) validated the Assumption 1: in fact, the subjects encountered a significantly higher ( $p = 0.001$ ) number of vehicles and they drove at a significantly lower ( $p = 0.039$ ) speed during the rush hours, as expected from scientific literature (Bucchi et al., 2012). Statistical analysis of driver's eye fixations over the external environment (Figure 5) and physiological brain patterns (Figure 6) validated the Assumption 2: in fact the drivers' gazes over the external environment (such index inversely correlates with mental workload; de Winter et al., 2014; Lantieri et al., 2015) have been significantly lower ( $p = 0.046$ ) along the circuit segment that was hypothesized as *Hard*, while the ratio between frontal theta and parietal alpha rhythms significantly increased ( $p = 0.009$ ). These results confirmed the properness of the experimental design. Nevertheless, the analysis of encountered vehicles, determined

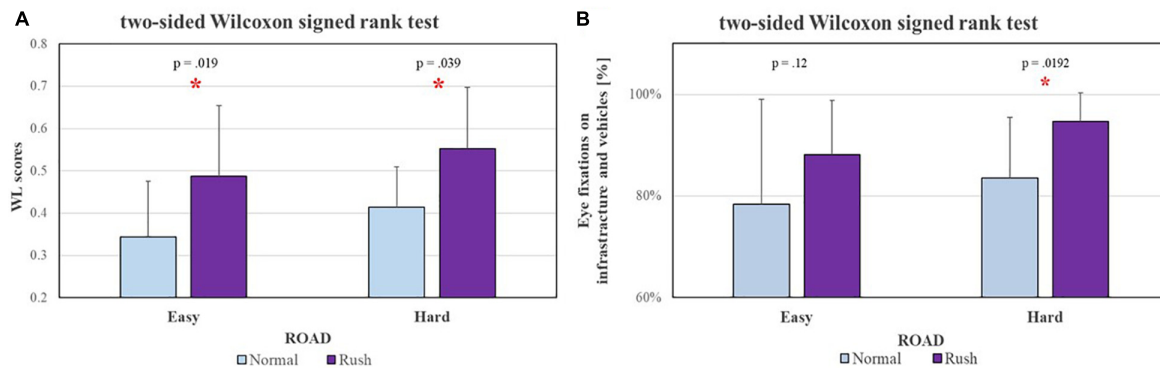




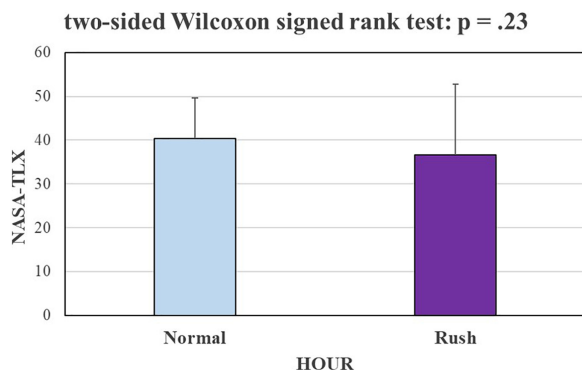
**FIGURE 9 |** At the top (A), the Friedman test highlighting a significant main effect ( $p = 0.00001$ ), in terms of mental workload increasing among the different factors. At the bottom, on the left (B) the Wilcoxon test on the factor ROAD and on the right (C) the same test on the factor HOUR, showing how both the factors produced a significant mental workload increasing (respectively  $p = 0.004$  and  $p = 0.003$ ). The statistical tests showing a significant effect.



**FIGURE 10 |** The Wilcoxon tests performed to investigate eventual sensitivity differences between Eye-Tracking [left (A)] and EEG [right (B)] measures, considered on the same subjects, in relation to ROAD complexity showed that EEG-based measures have been able to significantly discriminate ( $p = 0.008$ ) the two conditions at least during Normal hour, while the ET-based ones have not been able to show any significant difference both during Normal and Rush hours. The statistical tests showing a significant effect.



**FIGURE 11 |** The Wilcoxon tests performed to investigate eventual sensitivity differences between Eye-Tracking [left (A)] and EEG [right, (B)] measures, considered on the same subjects, in relation to traffic intensity (i.e., HOUR) showed that, while the EEG-based measures have been able to significantly discriminate the two conditions both along Easy ( $p = 0.019$ ) and Hard ( $p = 0.04$ ) segments, the ET-based ones have been able to significantly discriminate Normal and Rush hours only along the Hard segment ( $p = 0.02$ ). The statistical tests showing a significant effect.



**FIGURE 12 |** The bar graph represents the mean and the standard deviation of NASA-TLX scores, i.e., the subjective assessment of the mental workload experienced by the participants of the circuit. The Wilcoxon test does not reveal any significant difference in terms of workload subjectively assessed by the subjects between the Normal and Rush hour conditions.

through videos from the VBOX videos, led to discard three subjects because of no differences between rush and normal hours (see Results). Therefore, this validation approach should be taken into account for future works in real driving conditions, where external conditions and events are less controllable, even unpredictable, if compared with laboratory experiments.

Once validated the experiment in terms of differences between the road and traffic conditions, the EEG-based Workload measures have been validated. In particular, the analysis of AUC related to the asSWLDA-based classifier demonstrated that the adopted approach achieves considerable performance, i.e., AUCs  $> 0.7$  (Fawcett, 2006). More in detail, the AUC analysis (Figure 8) revealed that it has been possible to achieve mean AUC values of 0.74 for the Normal hour and of 0.73 for the Rush hour, significantly higher than a random classification in both the conditions (respectively  $p = 0.01$  and  $p = 0.0005$ ). Also, there were no significant differences ( $p = 0.64$ ) in terms of AUC values on Real data between Normal and Rush hours. All the

previous results have been also confirmed by the correction for multiple comparisons, in this case the False Discovery Rate. It is also true that, within the machine-learning theory, AUCs greater than 0.7 are considered remarkable if compared with a random distribution that is assumed to produce AUCs equal to 0.5. In the present study, the performance of the classifier on randomized data achieved AUCs values of about 0.6. A possible explanation could be that the random value would be closer and closer to 0.5 only if the number of repetitions tends to infinite, however, this result undoubtedly encourages research about improving the proposed method. Of course, classification performance of about 0.75 are anyway remarkable, in particular because of the novelty of such application (the EEG-based Workload index is provided with a time resolution equal to 8 s) and the real settings, where mental states assessment is more prone to misclassification: in fact, it is plausible to assume that outside the high controlled laboratory settings, the user experiences more complex mental states that consist of multiple different components having the potential to influence neurophysiological signals used to infer a specific state.

The analysis of the patterns of features selected by the algorithm during its training phase (Figure 7) provided interesting insights about its usability: in fact, the asSWLDA selected on average 4 discriminant features for each subject, and even more interesting, by involving 3 of the 11 available channels. It means that, once calibrated the system on a specific user, it would be able to work online during the driving experience involving only three EEG channels, in other words reducing significantly its invasiveness and increasing wearability, two critical aspects for applications outside the laboratory.

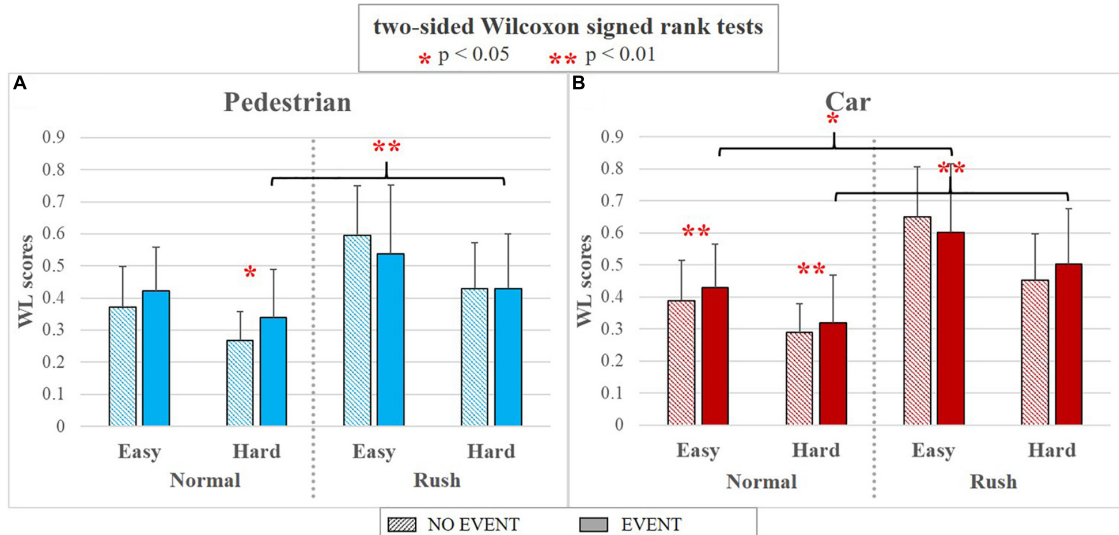
At this point, the asSWLDA output, in terms of EEG-based Workload index, has been used to evaluate the effects of road complexity, traffic intensity and external events on drivers' workload (Figure 9).

The Friedman ANOVA test (please see Figure) shows the results in terms of effects of the two investigated factors, i.e., the traffic (HOUR) and the road complexity (ROAD), on the mental workload experienced by the drivers: both the

traffic and road complexity contributed to significantly increase (main effect:  $p = 0.00001$ ; Wilcoxon tests respectively: HOUR,  $p = 0.0032$ ; ROAD:  $p = 0.0038$ ) the mental workload. In other words, the drivers' workload increased if traffic increased as well independently from the road complexity. At the same time, the drivers' workload increased while driving along more complex roads independently from the traffic intensity. These results have to be considered with respect to the experimental task: actually, the *Hard* segment was a three-lanes main street, that with respect to a one-lane main street (*Easy* segment) implies several additional decisions and actions, such as eventual car overtaking as well as looking at rear-view mirrors because of possible cars coming on lateral lanes. Of course, these actions increase with traffic increasing, because of the higher number of vehicles along the circuit (as demonstrated by Video analysis, please see **Figure 4**). Apparently, the *Easy* segment should not suffer traffic increasing, since being a one-lane segment the overtaking are very limited and drivers have not to frequently check rear-view mirrors since they cannot change lane. Nevertheless, because of the higher number of vehicles along the circuit during rush hours, the drivers had to continuously monitor eventual preceding cars, adapting safety distance and speed (in fact average speed during rush hour has been lower and drivers' gazes on infrastructure and vehicles higher also along *Easy* segment). These actions also induced a no-negligible workload increasing, giving a possible justification of the high accident rate along rural roads (Shankar et al., 1995), that are generally considered "*Easy to drive*" if compared with urban main roads (Harms, 1991; Paxion et al., 2014), thus mismatching the driver's expectations.

Very interestingly but not surprisingly, the neurophysiological measures showed a significantly higher sensitivity with respect to the ET ones (**Figures 10, 11**) in discriminating the different impact of road complexity and traffic intensity on mental workload. It is important to consider that ET measures were available only for a reduced group of the experimental sample (8 of 16 subjects), therefore it could have affected the performance of such measures in discriminating the mental workload related to different factors. However, the paired statistical analysis highlighted that on the same subjects, EEG-based measures were more sensitive to workload fluctuations. Their high sensitivity has been pointed out also with respect to subjective measures (i.e., NASA-TLX questionnaires, **Figure 12**), that on the contrary were not able to discriminate ( $p = 0.23$ ) normal from rush hours.

Finally, EEG-based workload measures revealed a significant workload increasing ( $p < 0.05$ ) related to both the investigated events, that is the car and the pedestrian crossing the road, especially in normal hours independently from the road complexity. Instead, no significant workload increasing were associated to the event during rush hours, despite a significantly higher workload in comparison with the same events during normal hours (**Figure 13**). Although for this analysis neurophysiological measures are not integrated with additional ones (it was impossible to collect subjective data related to specific events, while from the ET point of view it was possible to assess only if the event was been perceived or not), it is possible to deduce that external events could lead to eventually risky situations especially with low traffic (normal hours). In fact, although a lower absolute workload if compared with high traffic



**FIGURE 13 |** The bar graphs show the mean values and the standard deviation of the EEG-based WL scores related to the different events along the various experimental conditions. In particular, the results are divided per events category, i.e., Pedestrian on the left (**A**) and Car on the right (**B**). In both the cases, the condition EVENT (solid color) is referred to the event actually happened during the 3rd lap, the condition NO EVENT (lines pattern) is referred to the same circuit portion during the 2nd lap when no events were acted. The Wilcoxon tests revealed a significant workload increasing (one red asterisk stands for  $p < 0.05$ ; two red asterisks stand for  $p < 0.01$ ) related to both the investigated events, that is the car and the pedestrian crossing the road, especially in normal hours independently from the road complexity. Instead, no significant workload increasing were associated to the event during rush hours, despite a significantly higher workload in comparison with the same events during normal hours.

condition, they are characterized by an immediate cognitive demand increase, that could become dangerous if not expected by the driver.

Nevertheless, the main limit that affects the present study is the algorithm calibration with data coming from the task itself and recorded in very similar conditions. From one side, it could be argued that in everyday life context such a calibration would be unfeasible; from the other side it could be argued that the proposed algorithm is not classifying the targeted mental state, i.e., mental workload, but only two conditions that are very similar. Regarding the calibration, actually it is one of the main still open issues in transferring machine learning approaches from research to applied field: several solutions have been explored, such as cross-task calibration or employment of unsupervised algorithms, but the problem is still open and needs further investigation (Aricò et al., 2018). However, the present work did not aim at addressing such issue, but at investigating the possibility of applying a machine-learning algorithm for the mental workload evaluation, already validated in other domains, also in automotive applications. The highly challenging conditions of a “real driving experiment” with twenty subjects, jointly with the employment of high-quality instrumentation, already make the present work very innovative and of interest. Secondly, it is true that the algorithm has been calibrated on two conditions and employed in classifying two similar conditions, but it is also important to consider that calibration data for each subject came from two different repetitions (please refer to Section “Electroencephalographic Signal Recording and Processing” for more information): in fact data recorded during the *Easy* segment of *Normal* hour (2nd lap) have been used as EASY CLASS, while data recorded during the *Hard* segment of *Rush* hour (2nd lap) have been used as HARD CLASS. Even if assuming that *Easy* segment of *Normal* hour and *Hard* segment of *Rush* hour of 2nd and 3rd lap were intrinsically similar, no data from *Hard* segment of *Normal* hour and *Easy* segment of *Rush* hour have been used to train the classifier, therefore their coherent classification (e.g., *Hard* segment of *Normal* hour is not easier than the *Easy* segment during the same hour) is a mere and appreciable result of the proposed algorithm. Undoubtedly, mental workload is a Human Factor concept hard to define and even worse to measure (Moray, 2013), and confounds arising from different mental states are probably present, however, the results of the present study are already remarkable, especially if considering previous results obtained by the employment of the same algorithm in different applications (Aricò et al., 2016a,b, Borghini et al., 2017b,c).

It is important to remark how it is possible to achieve this kind of results only thanks to the proposed methodology: in fact, subjective measures cannot be gathered with high time resolution and without interfering with the main task, briefing and debriefing sessions can be performed only before and after the experience, while eye-tracker as well as other neurophysiological metrics (for example the *ThetaF/AlphaP* showed in Figure 6) are able to provide only an overall evaluation about a “long” condition. On the contrary, the proposed methodology is able to overcome these limitations, providing workload assessment with high time resolution (i.e.,

in this case 8 s) and thus allowing to evaluate also specific events.

In conclusion, the obtained results appear very interesting in terms of understanding driver's behaviors and its relationship with road environment, highlighting the added value of neurophysiological measures in providing insights about human mind that are not obtainable, or at least difficult to obtain, with traditional approaches. Certainly, further analyses are necessary in order to validate this multimodal approach with a larger sample of subjects, exploring the impact of other factors, such as different events, road signage and so on, and involving additional tools typical of road safety research, as well as exploring the possibility of calibrating the proposed algorithm without any task-related data.

## CONCLUSION

The present study, through a real driving experiment, aimed to validate a methodology able to infer driver's mental workload on the basis of his/her brain activity through Electroencephalographic technique. Once validated, such methodology has been successfully employed to evaluate the impact of different factors, specifically the road complexity, the traffic intensity (depending on the hour of the day), and two specific events (a pedestrian crossing the road and a car entering in the traffic flow), on the drivers' experienced mental workload. The analyses have been supported by information coming from subjective measures, drivers' eye movements tracking and car parameters. The results demonstrated (i) the reliability and effectiveness of the proposed methodology based on human EEG signals to objectively measure driver's mental workload with respect to different road factors, and (ii) the added value of neurophysiological measures in providing insights about human mind while dealing with tasks that are difficult or even impossible to obtain by using traditional approaches. In conclusion, other than the specific obtained results, the present work breaks new ground for the integration of these new methodologies, i.e., neurophysiological measures, with traditional approaches in order to enhance and extend research on drivers' behaviors and road safety.

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Good Clinical Practice (International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use) with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the ‘University of Bologna.’

## AUTHOR CONTRIBUTIONS

GDF is the main author of the paper. Also, he was actively involved within the experiments as well as in the EEG data



analysis and interpretation. PA, GB, and NS supported the experimental design, the data recording, the EEG data analysis, and the manuscript writing. PL and SP supported the experimental design and the results interpretation, providing their contribute in particular about the Human Factor concepts. VV, CL, AB, and AS were in charge of experiments planning, they contributed actively to the experiments execution, they analyzed Eye Tracking and subjective measures analysis, and supported the results interpretation. FB coordinated the research group, from the experimental design to the manuscript editing.

## REFERENCES

- Allnutt, M. F. (1987). Human factors in accidents. *Br. J. Anaesth.* 59, 856–864. doi: 10.1093/bja/59.7.856
- Aricò, P., Borghini, G., Di Flumeri, G., and Babiloni, F. (2017a). *Method for Estimating a Mental State, in Particular a Workload, and Related Apparatus*. EP3143933 A1, filed September 14, 2016, and issued March 22, 2017. Available at: <http://www.google.it/patents/EP3143933A1>
- Aricò, P., Borghini, G., Di Flumeri, G., Bonelli, S., Golfetti, A., Graziani, I., et al. (2017b). Human factors and neurophysiological metrics in air traffic control: a critical review. *IEEE Rev. Biomed. Eng.* 10, 250–263. doi: 10.1109/RBME.2017.2694142
- Aricò, P., Borghini, G., Di Flumeri, G., Sciaraffa, N., Colosimo, A., and Babiloni, F. (2017c). Passive BCI in operational environments: insights, recent advances, and future trends. *IEEE Trans. Biomed. Eng.* 64, 1431–1436. doi: 10.1109/TBME.2017.2694856
- Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Bonelli, S., Golfetti, A., et al. (2016a). Adaptive automation triggered by eeg-based mental workload index: a passive brain-computer interface application in realistic air traffic control environment. *Front. Hum. Neurosci.* 10:539. doi: 10.3389/fnhum.2016.00539
- Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Pozzi, S., and Babiloni, F. (2016b). A passive brain-computer interface application for the mental workload assessment on professional air traffic controllers during realistic air traffic control tasks. *Progr. Brain Res.* 228, 295–328. doi: 10.1016/bs.pbr.2016.04.021
- Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Graziani, I., Imbert, J. P., et al. (2015). "Reliability over time of EEG-based mental workload evaluation during Air Traffic Management (ATM) tasks," in *Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (New York, NY: IEEE), 7242–7245. doi: 10.1109/EMBC.2015.7320063
- Aricò, P., Borghini, G., Di Flumeri, G., Sciaraffa, N., and Babiloni, F. (2018). Passive BCI beyond the lab: current trends and future directions. *Physiol. Meas.* 39:08TR02. doi: 10.1088/1361-6579/aad57e
- Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J. Mathemat. Psychol.* 12, 387–415. doi: 10.1016/0022-2496(75)90001-2
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodological)* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Borghini, G., Aricò, P., Di Flumeri, G., and Babiloni, F. (2017a). *Industrial Neuroscience in Aviation: Evaluation of Mental States in Aviation Personnel*. Cham: Springer International Publishing Springer. doi: 10.1007/978-3-319-58598-7
- Borghini, G., Aricò, P., Di Flumeri, G., Cartocci, G., Colosimo, A., Bonelli, S., et al. (2017b). EEG-based cognitive control behaviour assessment: an ecological study with professional air traffic controllers. *Sci. Rep.* 7:547. doi: 10.1038/s41598-017-00633-7
- Borghini, G., Aricò, P., Di Flumeri, G., Sciaraffa, N., Colosimo, A., Herrero, M., et al. (2017c). A new perspective for the training assessment: machine learning-based neurometric for augmented user's evaluation. *Front. Neurosci.* 11:325. doi: 10.3389/fnins.2017.00325
- Borghini, G., Aricò, P., Di Flumeri, G., Salinari, S., Colosimo, A., Bonelli, S., and Babiloni, F. (2015). "Avionic technology testing by using a cognitive neurometric index: a study with professional helicopter pilots," in *Proceedings of the Engineering in Medicine et al. (EMBC), 2015 37th Annual International Conference of the IEEE* (New York, NY: IEEE), 6182–6185. doi: 10.1109/EMBC.2015.7319804
- Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., and Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev.* 44, 58–75. doi: 10.1016/j.neubiorev.2012.10.003
- Borghini, G., Vecchiato, G., Toppi, J., Astolfi, L., Maglione, A., Isabella, R., et al. (2012). "Assessment of mental fatigue during car driving by using high resolution EEG activity and neurophysiologic indices," in *Proceedings of the Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Diego, CA. 6442–6445. doi: 10.1109/EMBC.2012.6347469
- Brookhuis, K. A., and de Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. *Accid. Anal. Prevent.* 42, 898–903. doi: 10.1016/j.aap.2009.06.001
- Bucchi, A., Sangiorgi, C., and Vignali, V. (2012). Traffic psychology and driver behavior. *Proc. Soc. Behav. Sci.* 53, 972–979. doi: 10.1016/j.sbspro.2012.09.946
- Cestac, J., Paran, F., and Delhomme, P. (2014). Drive as i say, not as i drive: influence of injunctive and descriptive norms on speeding intentions among young drivers. *Trans. Res. Part F Traffic Psychol. Behav.* 23, 44–56. doi: 10.1016/j.trf.2013.12.006
- Costa, M., Simone, A., Vignali, V., Lantieri, C., Bucchi, A., and Dondi, G. (2014). Looking behavior for vertical road signs. *Transport. Res. Part F Traffic Psychol. Behav.* 23, 147–155. doi: 10.1016/j.trf.2014.01.003
- De Waard, D. (1996). *The Measurement of Drivers' Mental Workload*. Groningen: University of Groningen, Traffic Research Centre.
- de Winter, J. C. F., Happee, R., Martens, M. H., and Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: a review of the empirical evidence. *Transport. Res. Part F Traffic Psychol. Behav.* 27, 196–217. doi: 10.1016/j.trf.2014.06.016
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009
- Di Flumeri, G., Aricò, P., Borghini, G., Colosimo, A., and Babiloni, F. (2016). "A new regression-based method for the eye blinks artifacts correction in the eeg signal, without using any EOG channel," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Orlando, FL.
- Di Flumeri, G., Aricò, P., Borghini, G., Sciaraffa, N., Vignali, V., Lantieri, C., et al. (2018). "Mental workload assessment as taxonomic tool for neuroergonomics," in *Neuroergonomics. The Brain at Work and in Everyday Life*, 1st Edn, eds H. Ayaz and F. Dehais (Amsterdam: Elsevier).

## ACKNOWLEDGMENTS

The authors are grateful to the Unipol Group Spa and, in particular, ALFAEVOLUTION TECHNOLOGY, for the considerable help given in the research study. This work has been also co-financed by the European Commission through the Horizon2020 project H2020-MG-2016 *Simulator of behavioral aspects for safer transport*, "SimuSafe," (GA n. 723386), and the project "BrainSafeDrive: A Technology to Detect Mental States During Drive for Improving the Safety of the Road" (Italy-Sweden collaboration) with a grant of Ministero dell'Istruzione dell'Università e della Ricerca della Repubblica Italiana.

- Di Flumeri, G., Borghini, G., Aricò, P., Colosimo, A., Pozzi, S., Bonelli, S., et al. (2015). "On the use of cognitive neurometric indexes in aeronautic and air traffic management environments," in *Proceedings of the Symbiotic Interaction: 4th International Workshop, Symbiotic 2015*, Berlin, Germany. doi: 10.1007/978-3-319-24917-9\_5
- Dienes, Z. (2004). Assumptions of subjective measures of unconscious mental states: higher order thoughts and bias. *J. Consciousn. Stud.* 11, 25–45.
- Elul, R. (1969). Gaussian behavior of the electroencephalogram: changes during performance of mental task. *Science* 164, 328–331. doi: 10.1126/science.164.3877.328
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recogn. Lett.* 27, 861–874. doi: 10.1016/j.patrec.2005.10.010
- Gevens, A., and Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theor. Issues Ergon. Sci.* 4, 113–131. doi: 10.1080/14639220210159717
- Göhring, D., Latotzky, D., Wang, M., and Rojas, R. (2013). "Semi-autonomous car control using brain computer interfaces," in *Intelligent Autonomous Systems 12*, vol. 194, eds S. Lee, H. Cho, K.-J. Yoon, and J. Lee (Berlin: Springer), 393–408. doi: 10.1007/978-3-642-33932-5\_37
- Gopher, D., and Braune, R. (1984). On the psychophysics of workload: why bother with subjective measures? *Hum. Fact. J. Hum. Fact. Ergon. Soc.* 26, 519–532. doi: 10.1177/001872088402600504
- Hansen, F. (2007). Human error: a concept analysis. *J. Air Transport.* 11, 61–77.
- Harms, L. (1991). Variation in drivers' cognitive load. Effects of driving through village areas and rural junctions. *Ergonomics* 34, 151–160. doi: 10.1080/00140139108967303
- Hart, S. G., and Staveland, L. E. (1988). "Development of NASA-TLX (Task Load Index): results of empirical and theoretical research," in *Advances in Psychology*, Vol. 52, eds P. A. Hancock and N. Meshkati (North-Holland: Human Mental Workload), 139–183. doi: 10.1016/S0166-4115(08)62386-9
- Holm, A., Lukander, K., Korpela, J., Sallinen, M., and Müller, K. M. I. (2009). Estimating brain load from the EEG. *Sci. World J.* 9, 639–651. doi: 10.1100/tsw.2009.83
- Horowitz, A. D., and Dingus, T. A. (1992). Warning signal design: a key human factors issue in an in-vehicle front-to-rear-end collision warning system. *Proc. Hum. Fact. Ergon. Soc. Annu. Meet.* 36, 1011–1013. doi: 10.1177/154193129203601320
- Kaplan, S., Guvensan, M. A., Yavuz, A. G., and Karalurt, Y. (2015). Driver behavior analysis for safe driving: a survey. *IEEE Trans. Intell. Transport. Syst.* 16, 3017–3032. doi: 10.1109/TITS.2015.2462084
- Kirk, R. E. (2015). "Experimental design," in *The Blackwell Encyclopedia of Sociology*, ed. G. Ritzer (Oxford: John Wiley & Sons, Ltd.), doi: 10.1002/9781405165518.wbeose087.pub2
- Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res. Rev.* 29, 169–195. doi: 10.1016/S0165-0173(98)00056-3
- Kohlmorgen, J., Dornhege, G., Braun, M., Blankertz, B., Müller, K. L.-R., Curio, G., et al. (2007). "Improving human performance in a real operating environment through real-time mental workload detection," in *Towards Brain-Computer Interfacing*, eds G. Dornhege, J. del R. Millá, T. Hinterberger, D. McFarland, and K.-Robert Mülle (Cambridge, MA: MIT Press).
- Kong, W., Lin, W., Babiloni, F., Hu, S., and Borghini, G. (2015). Investigating driver fatigue versus alertness using the granger causality network. *Sensors* 15, 19181–19198. doi: 10.3390/s150819181
- Kong, W., Zhou, Z., Jiang, B., Babiloni, F., and Borghini, G. (2017). Assessment of driving fatigue based on intra/inter-region phase synchronization. *Neurocomputing* 219, 474–482. doi: 10.1016/j.neucom.2016.09.057
- Lantieri, C., Lamperti, R., Simone, A., Costa, M., Vignali, V., Sangiorgi, C., et al. (2015). Gateway design assessment in the transition from high to low speed areas. *Transport. Res. Part F Traffic Psychol. Behav.* 34, 41–53. doi: 10.1016/j.trf.2015.07.017
- Lees, M. N., Cosman, J. D., Lee, J. D., Fricke, N., and Rizzo, M. (2010). Translating cognitive neuroscience to the Driver's operational environment: a neuroergonomic approach. *Am. J. Psychol.* 123, 391–411. doi: 10.5406/amerjpsyc.123.4.0391
- Lin, C.-T., Ko, L.-W., and Shen, T.-K. (2009). Computational intelligent brain computer interaction and its applications on driving cognition. *IEEE Computat. Intell. Mag.* 4, 32–46. doi: 10.1109/MCI.2009.934559
- Lin, C.-T., Wu, R.-C., Jung, T.-P., Liang, S.-F., and Huang, T.-Y. (2005). Estimating driving performance based on EEG spectrum analysis. *EURASIP J. Adv. Sig. Process.* 2005:521368. doi: 10.1155/ASP.2005.3165
- Maglione, A., Borghini, G., Aricò, P., Borgia, F., Graziani, I., Colosimo, A., et al. (2014). "Evaluation of the workload and drowsiness during car driving by using high resolution EEG activity and neurophysiologic indices," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Chicago, IL, 6238–6241. doi: 10.1109/EMBC.2014.6945054
- Michail, E., Kokonoz, A., Chouvarda, I., and Maglaveras, N. (2008). *EEG and HRV Markers Of Sleepiness and Loss of Control During Car Driving*. New York, NY: IEEE. doi: 10.1109/IEMBS.2008.4649724
- Moray, N. (ed). (2013). *Mental Workload: Its Theory and Measurement*. Berlin: Springer Science & Business Media, 8.
- Parasuraman, R., and Rizzo, M. (2008). *Neuroergonomics: The Brain at Work*, 1 Edn. New York, NY: Oxford University Press.
- Parasuraman, R., Sheridan, T. B., and Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: viable, empirically supported cognitive engineering constructs. *J. Cognit. Eng. Decis. Mak.* 2, 140–160. doi: 10.1518/155534308X284417
- Paxion, J., Galy, E., and Berthelon, C. (2014). Mental workload and driving. *Front. Psychol.* 5:1344. doi: 10.3389/fpsyg.2014.01344
- Petridou, E., and Moustaki, M. (2000). Human factors in the causation of road traffic crashes. *Eur. J. Epidemiol.* 16, 819–826. doi: 10.1023/A:1007649804201
- Philip, P., Sagaspe, P., Taillard, J., Valtat, C., Moore, N., Åkerstedt, T., et al. (2005). Fatigue, sleepiness, and performance in simulated versus real driving conditions. *Sleep* 28, 1511–1516. doi: 10.1093/sleep/28.12.1511
- Prinzel, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., and Pope, A. T. (2000). A closed-loop system for examining psychophysiological measures for adaptive task allocation. *Int. J. Aviat. Psychol.* 10, 393–410. doi: 10.1207/S15327108IJAP1004\_6
- Reason, J. (2000). Human error: models and management. *BMJ? Br. Med. J.* 320, 768–770. doi: 10.1136/bmj.320.7237.768
- Robert, J. H. G. (1997). Compensatory control in the regulation of human performance under stress and high workload: a cognitive-energetical framework. *Biol. Psychol.* 45, 73–93. doi: 10.1016/S0301-0511(96)05223-4
- Shankar, V., Mannering, F., and Barfield, W. (1995). Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. *Accident Anal. Prevent.* 27, 371–389. doi: 10.1016/0001-4575(94)00078-Z
- Siegel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*. New York, NY: McGraw-Hill.
- Subramanian, R. (2012). *Motor Vehicle Traffic Crashes as a Leading Cause Of Death in the United States, 2008 and 2009. HS-811 620. Traffic Safety Facts – Research Note. National Highway Traffic Safety Administration*. Available at: <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/811620>
- Summala, H., and Mikkola, T. (1994). Fatal accidents among car and truck drivers: effects of fatigue, age, and alcohol consumption. *Hum. Fact.* 36, 315–326. doi: 10.1177/001872089403600211
- Toppi, J., Borghini, G., Petti, M., He, E. J., De Giusti, V., He, B., et al. (2016). Investigating cooperative behavior in ecological settings: an EEG hyperscanning study. *PLoS One* 11:e0154236. doi: 10.1371/journal.pone.0154236
- Treat, J. R., Tumbas, N. S., McDonald, S. T., Shinar, D., and Hume, R. D. (1979). *Tri-level Study of the Caused of Traffic Accidents. Executive Summary*. Available at: <https://trid.trb.org/view.aspx?id=144095>
- Vecchiato, G., Borghini, G., Aricò, P., Graziani, I., Maglione, A. G., Cherubino, P., et al. (2016). Investigation of the effect of EEG-BCI on the simultaneous execution of flight simulation and attentional tasks. *Med. Biol. Eng. Comput.* 54, 1503–1513. doi: 10.1007/s11517-015-1420-6
- Verwey, W. B. (2000). On-line driver workload estimation. Effects of road situation and age on secondary task measures. *Ergonomics* 43, 187–209. doi: 10.1080/0014013001845558
- Vicente, K. J. (2013). *The Human Factor: Revolutionizing the Way People Live with Technology*. Hoboken: Taylor and Francis. Available at: <https://www.taylorfrancis.com/books/9781135877262>

- Wall, T. D., Michie, J., Patterson, M., Wood, S. J., Sheehan, M., Clegg, C. W., et al. (2004). On the validity of subjective measures of company performance. *Pers. Psychol.* 57, 95–118. doi: 10.1111/j.1744-6570.2004.tb02485.x
- Wang, H., Zhang, C., Shi, T., Wang, F., and Ma, S. (2015). Real-time EEG-based detection of fatigue driving danger for accident prediction. *Int. J. Neural Syst.* 25:1550002. doi: 10.1142/S0129065715500021
- World Health, and Organization. (2015). *Global Status Report on Road Safety 2015: Supporting a Decade of Action*. Geneva: WHO.
- World Health Organization (2013). *Global Status Report on Road Safety 2013: Supporting a Decade of Action*. Available at: <http://site.ebrary.com/id/10931312>
- Xing, Y., Lv, C., Cao, D., Wang, H., and Zhao, Y. (2018). Driver workload estimation using a novel hybrid method of error reduction ratio causality and support vector machine. *Measurement* 114, 390–397. doi: 10.1016/j.measurement.2017.10.002
- Zhang, H., Chavarriaga, R., Khaliliardali, Z., Gheorghe, L., Iturrate, I., and Millán, J. D. R. (2015). EEG-based decoding of error-related brain activity in a real-world driving task. *J. Neural Eng.* 12:066028. doi: 10.1088/1741-2560/12/6/066028

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Di Flumeri, Borghini, Aricò, Sciaraffa, Lanzi, Pozzi, Vignali, Lantieri, Bichicchi, Simone and Babiloni. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# EEG-Based Neurocognitive Metrics May Predict Simulated and On-Road Driving Performance in Older Drivers

Greg Rupp<sup>1\*</sup>, Chris Berka<sup>1</sup>, Amir H. Meghdadi<sup>1</sup>, Marija Stevanović Karić<sup>1</sup>, Marc Casillas<sup>1</sup>, Stephanie Smith<sup>1</sup>, Theodore Rosenthal<sup>2</sup>, Kevin McShea<sup>3</sup>, Emily Sones<sup>3</sup> and Thomas D. Marcotte<sup>3</sup>

<sup>1</sup> Advanced Brain Monitoring Inc., Carlsbad, CA, United States, <sup>2</sup> Systems Technology, Inc., Hawthorne, CA, United States,

<sup>3</sup> Department of Psychiatry, University of California, San Diego, San Diego, CA, United States

## OPEN ACCESS

### Edited by:

Karel Brookhuis,  
University of Groningen, Netherlands

### Reviewed by:

Jodi M. Gilman,  
Massachusetts General Hospital,  
Harvard Medical School,  
United States  
Berry Wijers,  
University of Groningen, Netherlands

### \*Correspondence:

Greg Rupp  
grupp@b-alert.com

**Received:** 15 June 2018

**Accepted:** 17 December 2018

**Published:** 15 January 2019

### Citation:

Rupp G, Berka C, Meghdadi AH, Karić MS, Casillas M, Smith S, Rosenthal T, McShea K, Sones E and Marcotte TD (2019) EEG-Based Neurocognitive Metrics May Predict Simulated and On-Road Driving Performance in Older Drivers. *Front. Hum. Neurosci.* 12:532. doi: 10.3389/fnhum.2018.00532

The number of older drivers is steadily increasing, and advancing age is associated with a high rate of automobile crashes and fatalities. This can be attributed to a combination of factors including decline in sensory, motor, and cognitive functions due to natural aging or neurodegenerative diseases such as HIV-Associated Neurocognitive Disorder (HAND). Current clinical assessment methods only modestly predict impaired driving. Thus, there is a need for inexpensive and scalable tools to predict on-road driving performance. In this study EEG was acquired from 39 HIV+ patients and 63 healthy participants (HP) during: 3-Choice-Vigilance Task (3CVT), a 30-min driving simulator session, and a 12-mile on-road driving evaluation. Based on driving performance, a designation of Good/Poor (simulator) and Safe/Unsafe (on-road drive) was assigned to each participant. Event-related potentials (ERPs) obtained during 3CVT showed increased amplitude of the P200 component was associated with bad driving performance both during the on-road and simulated drive. This P200 effect was consistent across the HP and HIV+ groups, particularly over the left frontal-central region. Decreased amplitude of the late positive potential (LPP) during 3CVT, particularly over the left frontal regions, was associated with bad driving performance in the simulator. These EEG ERP metrics were shown to be associated with driving performance across participants independent of HIV status. During the on-road evaluation, Unsafe drivers exhibited higher EEG alpha power compared to Safe drivers. The results of this study are 2-fold. First, they demonstrate that high-quality EEG can be inexpensively and easily acquired during simulated and on-road driving assessments. Secondly, EEG metrics acquired during a sustained attention task (3CVT) are associated with driving performance, and these metrics could potentially be used to assess whether an individual has the cognitive skills necessary for safe driving.

**Keywords:** EEG, event related potentials, sustained attention, driving, HIV, neurodegeneration, driving impairment test, on-road evaluation

## INTRODUCTION

Driving is an essential aspect of maintaining health, independence and quality of life as individuals age (Ball et al., 1998). Those who voluntarily avoid driving due to perceived age-related sensory or cognitive deficits often suffer substantial consequences such as decreased mobility, increased dependency, social isolation, depression, and higher incidence of nursing home placement



(Marottoli et al., 1997, 2000; Fonda et al., 2001; Ragland et al., 2005; Freeman et al., 2006; Czigler et al., 2008; Choi et al., 2013). Driving requires a myriad of cognitive functions including attention, visuospatial processing, psychomotor integration, adequate processing speed, and executive function (Kellison, 2009). Normal aging, in the absence of any neurological or psychiatric disease, can lead to declines in these cognitive abilities increasing the risk for an automobile collision (Brayne et al., 2000; Ball, 2009). However, the aging process and its effects on driving performance vary significantly between individuals (Ball, 2009). It has been suggested that specific age-related functional impairments, and not age itself, put one at risk for impaired driving (Ross et al., 2009). Overall, older drivers as a group incur the highest number of fatalities per mile driven compared to other age groups (although the physical frailty of older individuals contributes significantly to this mortality rate) (Tefft, 2017). In addition to normal aging, functional deficits associated with neurodegenerative diseases (NDDs) such as Alzheimer's (AD), Mild Cognitive Impairment (MCI), or HIV-associated neurocognitive disorders (HAND) may affect driving performance. While NDD patients are more likely to be at-risk drivers, research suggests that memory deficits alone may not necessarily lead to unsafe driving (Carr et al., 1998; Marcotte et al., 1999, 2004; Silverstein et al., 2002; Charlton et al., 2003; Duchek et al., 2003; Uc et al., 2004, 2005; Man-Son-Hing et al., 2007; Frittelli et al., 2009; Wadley et al., 2009; Kawano et al., 2012). Cognitive impairments that affect driving, such as visuospatial processing deficits often found in patients with MCI or HAND, may be subclinical and unobserved by the patient themselves or their friends and family (Cysique et al., 2009; Chiao et al., 2013). Therefore, driving impairment cannot be established using only age and/or a NDD diagnosis.

In the United States, legal requirements for elderly drivers vary greatly from state to state. Some states have no safety-related policies for older drivers, whereas other states may have limited requirements for elderly individuals. For example, license renewal in California for those over the age of 70 may require a vision and/or written test, and in rare cases an on-road evaluation is administered (Department of Motor Vehicles, 2018a). Other states like Connecticut and Delaware have no age-related safety policies in place. Driver's licenses in these states need to be renewed every 6–8 years for all drivers regardless of age, often with no functional assessment required (Department of Motor Vehicles, 2018b,c).

Physicians have a responsibility to identify patients of all ages that might be considered at-risk drivers. However, they are often reluctant to take action due to privacy concerns and/or the severe impact their intervention could have on the patient's quality of life that results from the loss of a driver's license. Currently, there is no definitive diagnostic test for physicians to administer that identifies at-risk drivers, but individuals deemed potentially high risk may be referred for neuropsychological testing. The relationships between on-road driving performance and standard neuropsychological tests are modest, particularly in patients with mild to moderate cognitive decline or those recovering from trauma, surgery or treatments such as chemotherapy (Withaar et al., 2000; Reger et al., 2004; Leproust et al., 2008;

Classen et al., 2009). The most reliable method of evaluating driving impairment is an on-road test with a DMV-certified driving examiner, but annual on-road driving evaluations for all seniors, or even just those with clinically diagnosed cognitive impairments, are neither practical nor economical (Schanke and Sundet, 2000; Kay et al., 2008; Versijpt et al., 2017). Therefore, there is a need for inexpensive and sensitive tests to predict on-road driving impairment.

This study investigated the use of simultaneous electroencephalogram (EEG) and electrocardiogram (ECG) measurement in a population of healthy participants and HIV+ patients (>55 years old) during a test of sustained attention and processing speed. The combination of EEG, ECG, and behavioral performance metrics derived from the 3CVT were previously proven highly sensitive and specific in quantifying daytime drowsiness associated either with sleep deprivation in healthy participants or in sleep disordered patients, predicting susceptibility to sleep deprivation, and assessing neurocognitive deficits in patients with Parkinson's disease (PD), AD, MCI, and sleep disorders (Westbrook et al., 2002; Berka et al., 2006, 2007, 2009; Pojman et al., 2009a,b; Johnson et al., 2010, 2011; Waninger et al., 2018). The 3CVT evaluates sustained attention, visuospatial processing speed, and decision-making. These cognitive abilities are relevant to driving performance and prior work suggests that EEG metrics obtained during 3CVT were sensitive to improvements in cognition as a result of successful interventions for both sleep deprivation and sleep disorders (Westbrook et al., 2002; Berka et al., 2006, 2007, 2009; Pojman et al., 2009a,b; Johnson et al., 2010, 2011; Stoiljkovic et al., 2018). In addition to 3CVT, EEG and ECG were also acquired during both a simulated driving scenario and an on-road driving evaluation to conduct an exploratory analysis to assess any potential real-time neurophysiological changes associated with driving performance. Specifically, differences in the early (P200) and late (LPP) components evoked by the 3CVT have been associated with differences in cognitive abilities such as selective attention, memory, and decision-making. Since the 3CVT EEG metrics were previously shown to be associated with neurocognitive deficits in cognitively impaired populations, the investigators hypothesized that these metrics could be useful in distinguishing Safe and Unsafe drivers.

Physiological (heart rate, heart rate variability, skin conductance, and respiration) and neurophysiological (EEG) measures have long been used to unobtrusively assess the psychophysiological correlates of driving performance during simulated and on-road driving. Characteristic changes in EEG Power Spectral Densities (PSDs) have been associated with real-time changes in driving performance, phasic task demands, multiple domains of workload, and drowsiness (Zwinkels et al., 1990; de Waard and Brookhuis, 1991; Brookhuis and de Waard, 1993; Rookhuis et al., 1993; Mitler et al., 1997; Lei and Roetting, 2011; Dijksterhuis et al., 2013). Similarly, heart rate and heart rate variability have proven useful in measuring dynamic changes in cognitive demand during driving (Brookhuis et al., 1991; Mulder, 2004; Mehler et al., 2009, 2012). Several recent reports suggest the potential utility of real-time EEG-based algorithms to detect driver drowsiness

and inattention. Continuous monitoring of EEG and heart rate data during driving provides excellent temporal resolution and offers the potential for identifying driver fatigue early enough to intervene and prevent sleep onset. Several recent reports suggest the potential utility of real-time EEG-based algorithms to detect driver drowsiness and inattention (Ajinoroozi et al., 2016; Perrier et al., 2016; Hajinoroozi et al., 2017). Several challenges remain for the implementation of integrated driver monitoring systems including: obtaining high quality EEG and ECG with unobtrusive sensor systems, validating and implementing the real-time algorithms to achieve accurate identification of fatigue or inattention, and determining the optimal approach to interventions during driving (Dong et al., 2010). Another important consideration is the generalizability of the algorithms across all age groups, as the majority of published results use algorithms that have been designed and tested on college age research participants. These EEG-based algorithms are used to monitor real-time changes during driving. To date, EEG metrics have not been used to predict driving performance in elderly individuals with or without cognitive impairment.

As normal aging is associated with changes in cognitive abilities related to driving, normal aging also affects EEG signals. Older individuals show a decrease in power in the alpha band (8–13 Hz) and decreased amplitude of ERP components, particularly the P300 and Late Positive Potential (LPP) (De Gennaro et al., 2005; Polich and Corey-Bloom, 2005; Olichney et al., 2008; Vecchio et al., 2013; López et al., 2014; Ishii et al., 2017). Older drivers are also more likely to exhibit EEG based signs of fatigue and distraction that increase risks of driving errors (Johansson, 1997). In patients diagnosed with Alzheimer's disease, the most commonly reported findings for resting-state EEG are: a shift of the power spectrum to slower frequencies (i.e., increased delta and theta specifically over the temporal-parietal regions; decreased alpha, beta, and gamma) (Bonanni et al., 2008; Jelic and Kowalski, 2009; Dauwels et al., 2010a,b; Tsolaki et al., 2014). Patients with AD also display prolonged latencies and diminished ERP amplitudes and these cognitive-evoked measures do tend to correlate better with severity of cognitive impairments (Polich and Corey-Bloom, 2005; Garn et al., 2014). The EEG power shifts and ERP differences in AD are primarily associated with memory related functions. Additionally, patients with HIV (with a subset of those potentially having HAND) exhibited decreased amplitude and increased latency of the P300 and the Late Positive Potential (LPP) components compared to healthy controls (Polich et al., 2000; Polich and Basho, 2002; Chao et al., 2004; Bauer, 2011; Olichney et al., 2011; Papaliagkas et al., 2011). To date, these studies have not directly examined the relationship between EEG metrics associated with aging or cognitive impairment and driving competencies.

This paper contributes to the field by: (1) establishing the link between neurophysiological measures obtained during computerized neurocognitive assessments and on-road driving performance, (2) evaluating older adults (>55 years old) and individuals with a condition that can lead to cognitive impairment (HIV+). As such, this research offers the potential to provide a standardized methodology for predicting driving impairment due to disease related causes or natural aging.

## MATERIALS AND METHODS

### Participants

Sixty-three healthy participants (HP) (age 55–87 years, mean =  $65 \pm 8.2$  years, 49.2% male) and 39 HIV+ patients (age 55–74 years, mean =  $61 \pm 4.7$  years, 87.1% male) were enrolled in the study. The groups did not differ in years of education (HIV+: 9–20 years of education, mean =  $15.5 \pm 2.9$ ; HP: 10–21 years of education, mean =  $15.6 \pm 2.7$ ). HIV+ patients were primarily recruited from the University of California, San Diego HIV Neurobehavioral Research Program (UCSD HNRP) and healthy participants from the surrounding San Diego community using flyers and handouts.

Participants were selected after an initial telephone screening to determine their eligibility including the capability to provide informed consent to cognitive testing, simulator testing, and an on-road driving evaluation. Participants were included only if they possessed a current driver's license which was confirmed by the California Department of Motor Vehicles (CA DMV) on the day of their visit.

Additional exclusion criteria were: a history of loss of consciousness >30 min, current substance dependence, psychosis, diagnosis of a cardiovascular, sleep, or pulmonary disorder, and central nervous system opportunistic infections or neurologic disease other than HIV infection, reported diagnoses of Attention Deficit Hyperactivity Disorder (ADHD) or anxiety related disorders. All HIV+ individuals were on anti-retroviral therapy to control viral load, and healthy participants were excluded for all medication except for over the counter drugs and drugs for hypertension, diabetes, arthritis (non-opioid pain medication), and mild to moderate depression. The HIV+ populations used for this study were taking the following medications: 15 on antidepressants, eight on benzodiazepines, two on antipsychotics, three on anxiolytics, three on narcotics, and one on an anticoagulant. Urine toxicology (7-panel) and breathalyzer evaluations were also collected from all participants prior to starting the study visit. If either test was positive or the participant acted in a manner suggesting intoxication, he/she was rescheduled, or withdrawn from the study.

Three participants who signed informed consent forms and began the study protocol were excluded from all analyses due to a positive urine test for methamphetamine, and one additional participant was excluded due to being severely cognitively impaired despite a negative HIV status. Protocols were approved by both the UCSD IRB and Sharp IRB (IRBANA).

### Procedures

All participants completed neuropsychological (NP) testing and Advanced Brain Monitoring's (ABM) 3-Choice Vigilance Task (3CVT) as well as driving simulations (a screening drive, and subsequent challenge drive). A subset of the participants from the HIV+ (N=20) and HP (N=30) groups also completed an on-road driving evaluation (see below). EEG was collected concurrently using ABM's STAT™ X10 EEG sensor headset during all three tasks: 3CVT, simulated driving, and the on-road driving evaluation. The X10 is a battery-powered, lightweight, easy-to-apply wireless EEG system that acquires 9 channels

of EEG (Fz, F3, F4, Cz, C3, C4, P3, P4, POz, referenced to linked mastoids), and ECG. It uses passive, Ag/AgCl electrodes printed on PET strip flex circuit cables. A piece of single-use foam filled with conductive cream (Synapse by Kustomer Kinetics) was attached to the strip over each electrode site in order to make contact with the scalp. Impedances were measured and all channels were considered acceptable at or below 40 kOhms. Amplification and the A/D conversion was done adjacent to the electrode sites, allowing for high-quality data to be collected with higher than traditional impedance cut-offs. Data were sampled at 256 Hz with a high band pass at 0.1 Hz and a low band pass, fifth order filter, at 100 Hz obtained digitally with sigma-delta 16-bit A/D converters. Data were transmitted wirelessly via Bluetooth to a host computer, where acquisition software then stored the psychophysiological data.

### Cognitive and Medical Assessment

Cognitive status was successfully obtained through NP testing for 85 of the 102 participants (29 HIV+ and 56 HP), determined using either the HNRP NP assessment battery (56% of cohort) or the NIH Toolbox Cognition module (44% of cohort) (Berka and Marcotte, 2017). For this subset of participants, 34% of the HIV+ group was classified as impaired and 27% of the healthy participants were classified as impaired based on the NP testing, meaning there were no group differences in cognitive status due to HIV status. Impairment was defined for the toolbox as a T score of <40 on two of the tests, and for the NP assessment as a global deficit score of <0.5. For all participants, HIV status was confirmed through a finger stick blood test.

### 3CVT and EEG Measures

All participants were administered 3CVT, with concurrent EEG recording to assess neurocognitive functions. The 3CVT incorporates features of the most common measures of sustained attention, such as the Continuous Performance Test, Wilkinson Reaction Time, and the PVT-192 (Riccio et al., 2001; Sateia, 2003). The 3CVT requires subjects to discriminate one primary Target (triangle shape ▲, 70% of trials) from Non-Target (triangle shape upside down ▼, 15% of trials). The remaining 15% of the trials were used as Distracters (presenting a diamond shape: ◆) to increase the task complexity but are not included in the final Event Related Potential analysis. The test is 20-min long, during which 376 images are presented for a duration of 0.2 s each. A training period is provided prior to the start to minimize practice effects (Levendowski et al., 2000, 2001). The 3CVT challenges the participant's ability to sustain attention by increasing the inter-stimulus interval (ISI) across four, 5-min quartiles. During the first quartile, the ISI ranges between 1.5 and 3 s, increasing up to 6 s during the second quartile, and up to 10 s during the third and fourth quartiles.

### ERP Measures

For the 3CVT task, raw EEG signals were filtered between 0.1 and 50 Hz using a Hamming windowed Sinc FIR filter (0.1 Hz transition band). For each event type, EEG data were epoched from 1 s before and 2 s after the stimulus onset. The baseline

was adjusted using data from 100 ms before the stimulus onset. Artifacts were detected and excluded using automated algorithms (EEGLAB software) (Delorme and Makeig, 2004). Outliers were detected based on kurtosis of signal distribution (kurtosis >5 standard deviation), joint probability of values in an epoch given the whole data set (thresholded at 5 standard deviation), and unusual spectral patterns of epochs (with power spectrum 35 dB higher or lower than the baseline in the frequency range of 20–30 Hz). To exclude trials contaminated by ocular artifacts, trials were rejected if the absolute value of the EEG amplitude in any channel exceeded 100 microvolts during a window of 50 ms pre-stimulus onset to 750 ms post-stimulus onset. A minimum of 15 clean trials for each of the stimulus subtypes in 3CVT (Target and Non-Target) were required to be included in the analysis of that subtype. Grand average ERPs in each condition and trial type were calculated using a weighted average with the number of ERPs in each condition as the weights. For each participant, ERPs were measured using the average of the signal during a window of 180–220 ms post-stimulus onset for the P200 component, and the late positive potential (LPP) was measured using the average of the signal during a window 300–700 ms post-stimulus onset.

### Simulated Driving

Participants completed two simulated driving scenarios: an initial screening and a challenge. Seventy-eight percent of eligible participants were able to complete both scenarios. The remaining 22% were unable to complete both scenarios, primarily due to mild to severe motion sickness. To mitigate motion sickness, the driving scenario was split into three sessions with breaks in between. A STISIM M300WS Console driving simulator (System Technology Inc., Hawthorne, CA, USA) was used for both sets of driving simulations (**Figure 1**). The screening drive is a practice session of approximately 15 min given in order to familiarize participants with the driving simulator. Following the Screening Drive, participants began the Challenge Drive, which is a longer (3, 10-min segments, 30 min total), more complex drive assessing a range of abilities. Participants were instructed to complete the Challenge Drive while following traffic laws. The Challenge Drive was designed to be a surrogate for measuring on-road driving performance.

The Challenge Drive consisted of monotonous, uneventful, and low-load driving scenarios as well as highly demanding events such as busy intersections, crash avoidance, and unprotected turns. Busy sections were interspersed throughout the simulation run and lasted for 4–5 min. For example, one complex segment required the driver to avoid and pass slow moving cars while driving through dense fog. Once the fog lifted, the driver entered a city scene where a van was parked in the left lane and two pedestrians suddenly stepped into the road from in between two parked cars. Other highly engaging events included passing through a narrow construction zone with many barriers, avoiding cars suddenly entering the roadway, making left turns in front of oncoming traffic, passing slow moving trucks on a two-lane highway with oncoming traffic, and avoiding pedestrians stepping into the roadway without warning. The non-challenging





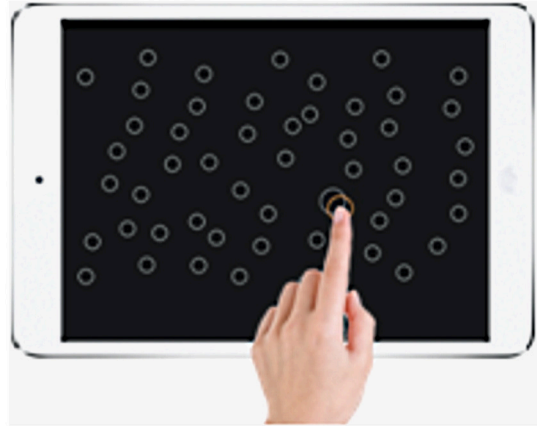
**FIGURE 1** | STISIM M300WS console with participant (written informed consent obtained).

sections consisted of stretches of highway where no other cars were present and no challenging events were triggered.

The Challenge Drive also contained a divided attention task called the Surrogate Reference Task (SuRT), aimed at examining distracted driving. The SuRT was initiated by an auditory cue (phone ringing) and required the participant to look down and to their right, forcing them to take their eyes entirely off the roadway to perform this secondary task, much like using a GPS or infotainment system. Participants were required to identify a circle that was different in size from other circles on the screen of a tablet (**Figure 2**). The easy, medium, and hard trials of this task were differentiated by the difference in size between the target and distractor circles. The target circle radii remained 20.7 mm for all three trials, while the distractor circle radii increased from 10.4 to 13.8 to 17.4 mm. Throughout this task, the simulation consisted of a two lane freeway without turns, a speed limit of 65 MPH, and no cars in either direction. Outcomes of interest included swerving [standard deviation of lateral position (SDLP)], speed maintenance (including variability) as well as accuracy and reaction time on the secondary task.

### On-Road Driving Evaluation

A subset of 50 participants (age 55–79 years, mean = 62 ± 6.6, 66% male, 40% HIV+) who completed the neurocognitive testbed and the driving simulator were selected to complete the on-road drive. Only 50 were selected due to time and budget restraints; selected participants must have completed the 3CVT and driving simulator scenarios. The on-road driving route was approximately 12 miles and required, on average, 45 min to complete (**Supplementary Figure 1**). It was conducted by the Sharp Rehabilitation Services Driving Program using a standardized approach with excellent inter-rater reliability (Cohen's  $K = 0.86$ ) and established sensitivity to HIV-related driving changes. A DMV-certified driving examiner was positioned in the front passenger seat of a dual-brake automobile; an occupational therapist (OT) and ABM technician (taking



**FIGURE 2** | Example of a participant identifying correct, target circle.

detailed notes about the driving safety and performance as well as monitoring the EEG signals) observed the drive in the rear seats. Participants were instructed to drive through residential and commercial areas, across controlled and uncontrolled intersections, and on freeways (including multiple merges). The participants followed single and multi-step directions (e.g., “Make the next available right turn... In three traffic lights, make a left turn”) throughout the duration of the drive.

### Evaluating Driving Performance

In order to evaluate driving performance participants were divided into groups of “Good” or “Poor” drivers based on performance in the simulator and “Safe” or “Unsafe” drivers based on on-road performance. The following sections describe this group assignment process.

#### On-Road Performance

Both the driving examiner and OT evaluated the drive in two ways. First, 186 scoring criteria for correctly performing traffic checks, maintaining lane position and speed, yielding when appropriate, etc. were assigned either a zero for pass, or a one for fail. Second, participants were given an overall score of 1 (excellent) through 5 (recommends they should not be driving) (**Supplementary Table 1**).

Each evaluator independently completed the pass/fail scores during the drive, and assigned an overall score after the conclusion of the drive. The driving instructor and OT would then arrive at a consensus evaluation for the overall score as well as a consensus regarding individual pass/fails. In addition, the OT documented critical errors in the form of physical or verbal interventions. Physical interventions included using the passenger-side brake and grabbing the wheel, while verbal interventions included any additional instructions or warnings that were not part of the scripted directions. Each driver was designated Safe or Unsafe based on the consolidated raters scores, comments, critiques, observations, and critical errors. Thirty-five of the 50 drivers were designated Safe (70%) and 15 drivers were designated Unsafe (30%).



## Driving Simulator Performance

Individual mistakes over the course of the challenge drive were counted and given weights to generate a weighted score as follows:

- 3 pts for a collision with another vehicle
- 2 pts for running stop signs or red lights
- 1 pt for speed exceedances and lane marker collisions (e.g., a traffic cone in a construction zone)
- 0.5 pts for crossing over the center dividing line or crossing into the right shoulder without causing a collision.

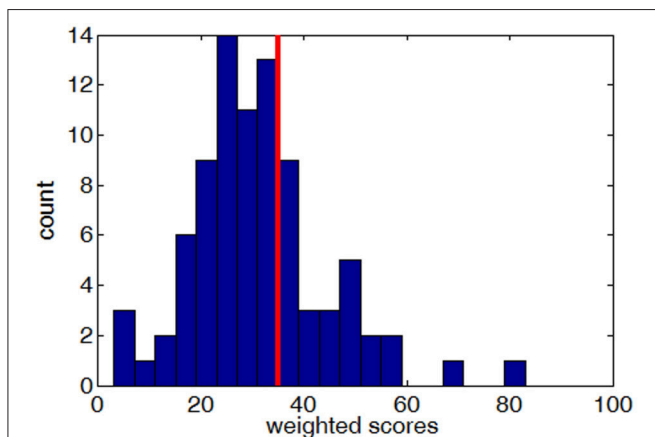
Using these weights, a total weighted score was computed for each participant who completed the simulated drive. This weighted score was used to divide drivers into either Good or Poor groups. Drivers with a weighted score of 35 or more designated as Poor. This threshold of 35 was chosen to result in 70/30% Good/Poor ratio to match the Safe/Unsafe ratio observed during the on-road drive (see On-Road Performance). **Figure 3** shows the distribution of weighted scores for all participants who completed the simulated driving scenario.

## Predicting Driving Performance

A linear discriminant function (LDF) was designed to classify Safe vs. Unsafe drivers using EEG ERP measures (P200 and LPP for both Target and Non-Target trials across all channels) obtained during the 3CVT test. The variables used for the LDF were selected through a step-wise algorithm in a logistic regression analysis. The classifier was evaluated using a leave-one-out cross validation method.

## RESULTS

EEG and behavioral measures were computed for all three tasks (3CVT, simulated driving, and on-road evaluation). Performance in the driving simulator was used to group subjects into either



**FIGURE 3 |** The distribution of weighted scores (higher scores indicate worse performance) across all subjects who completed the simulated drive, with the red line showing the cut-off threshold of 35.

Good or Poor (section Driving Simulator Performance), and on-road driving performance was used to designate subjects as either Safe or Unsafe (section On-Road Performance). To investigate the relationship between each behavioral/EEG measure and driving performance, these measures were averaged across the Safe (or Good) groups and were compared to the average of the Unsafe (or Poor) groups.

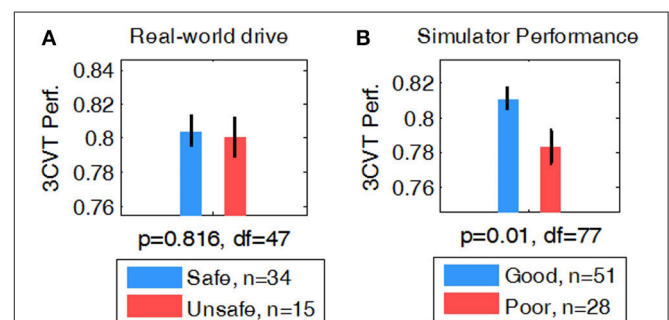
To investigate the relationships between HIV seropositivity and driving performance, chi-square tests of independence were performed for simulated and on-road driving groups. The proportion of Good vs. Poor (60.6 vs. 39.4%) drivers in the HIV+ group was not significantly different than that of the HP group (69.2 vs. 30.8%) [ $\chi^2$  (1,  $n = 85$ ) = 0.34,  $p = 0.56$ ]. Similarly, the proportion of Safe vs. Unsafe (60.0 vs. 40.0%) drivers in the HIV+ group was not significantly different than that of the HP group (76.6 vs. 23.4%) [ $\chi^2$  (1,  $n = 50$ ) = 0.89,  $p = 0.34$ ]. Therefore, driving performance both in the simulator and on-road was determined to be independent of HIV status in this population.

## Behavioral Measures

Behavioral measures included simulated driving performance, on-road driving performance, and Reaction Time (RT)/Accuracy for the 3CVT, as described in sections Driving Simulator Performance, On-Road Performance, and 3CVT and EEG Measures, respectively.

## 3CVT Behavioral Measures as Predictors of Driving Performance

Behavioral measures during the 3CVT attention task were computed for each participant including RT, Accuracy (percent correct), and a combined measure of performance (F-measure, i.e., a harmonic mean of normalized accuracy and reaction time) (Stikic et al., 2011). A student's t-test was used to determine whether group averages of 3CVT behavioral measures were different for Safe/Unsafe (on-road drive) and Good/Poor (driving simulator) drivers. F-measure showed no significant difference in performance between Safe and Unsafe drivers ( $p = 0.81$ ,  $df = 47$ ) (**Figure 4A**). However, Good drivers in the simulator had significantly higher performance compared to Poor drivers ( $p < 0.01$ ,  $df = 77$ ) (**Figure 4B**).



**FIGURE 4 |** Comparison of F-measure for (A) Safe vs. Unsafe and (B) Good vs. Poor.

## Driving Simulator

Throughout the driving simulation, there was high variability between subjects in speed, speed deviation, SDLP, and time to collision as individuals navigated the various complex segments with varying approaches. For example, **Supplementary Table 2** shows the high variance of speed between subjects for each block. Although participants were instructed to follow the rules of the road, the completion time for each segment of the driving scenario varied widely between participants. Because of the high between- and within- subject variability of these metrics, driving performance in the simulator was quantitatively computed using the variables described in section Driving Simulator Performance. To assess the relationship between on-road driving performance and simulator performance, a chi-squared test of independence was performed. 72.7% of Safe drivers were Good in the simulator and 71.4% of Unsafe drivers were Poor in the simulator [ $\chi^2(1, n = 47) = 6.23, p = 0.01$ ].

### SuRT Performance in Driving Simulator to Predict Simulator/On-Road Driving Performance

The mean Number Correct and mean Reaction Time for each of the three difficulty levels of the secondary task are illustrated in **Figure 5**. Students' *t*-tests revealed that no significant difference in Number Correct from easy to medium was present, but Number Correct did differ significantly between medium and hard (*t*-test,  $df = 141, p < 0.01$ ), and easy to hard (*t*-test,  $df = 141, p < 0.01$ ). Mean Reaction Time significantly increased from easy to medium (*t*-test,  $df = 143, p < 0.05$ ) and medium to hard (*t*-test,  $df = 141, p < 0.01$ ).

Ideal driving behavior during the SuRT would be characterized by a low rate of swerving (low SDLP), an average speed close to the speed limit (65 MPH), and a low rate of speed deviation. SDLP significantly increased from easy to

hard (*t*-test,  $df = 141, p < 0.01$ ) and from medium to hard (*t*-test,  $df = 141, p < 0.01$ ). Speed deviation significantly increased from easy to hard (*t*-test,  $df = 141, p < 0.05$ ) and medium to hard (*t*-test,  $df = 141, p < 0.01$ ). Average Speed decreased from medium to hard ( $df = 141, p < 0.05$ ).

While the SuRT task proved to be useful in measuring the effect of multitasking on driving behavior, neither SuRT driving performance nor secondary task performance were significantly different for Good vs. Poor (simulator) or for Safe vs. Unsafe (on-road) drivers.

## On-Road Drive

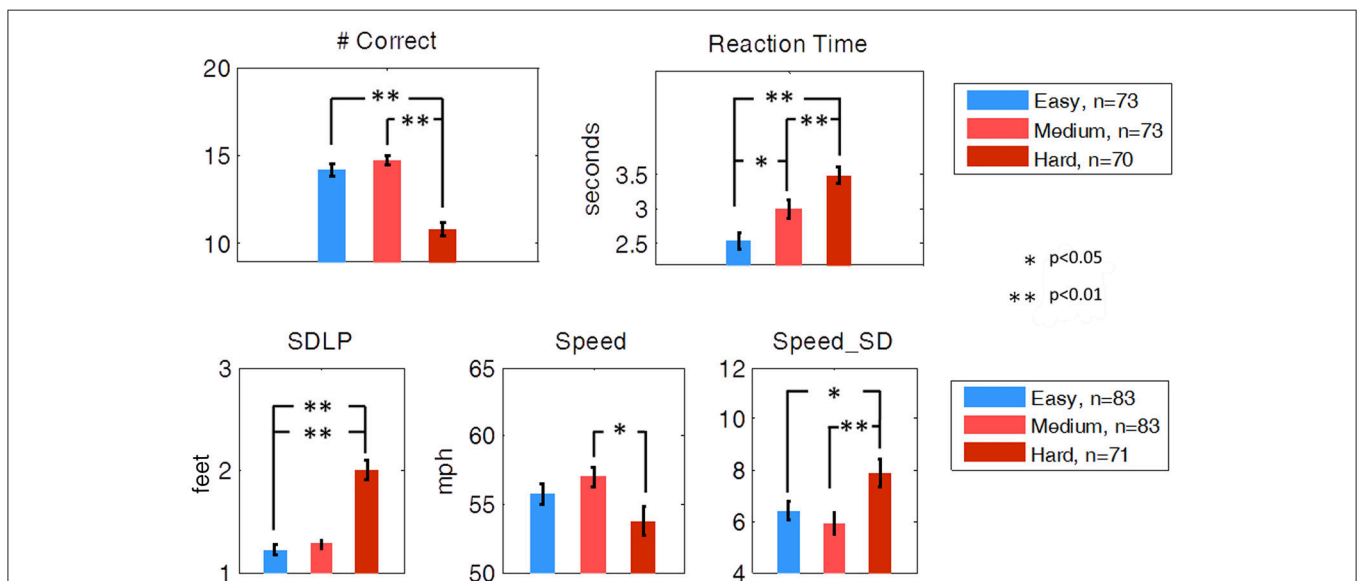
The overall Safe and Unsafe driver's scores were computed as described in section On-Road Performance and were used for group comparisons.

## Association Between 3CVT EEG ERP Measures and Driving Performance

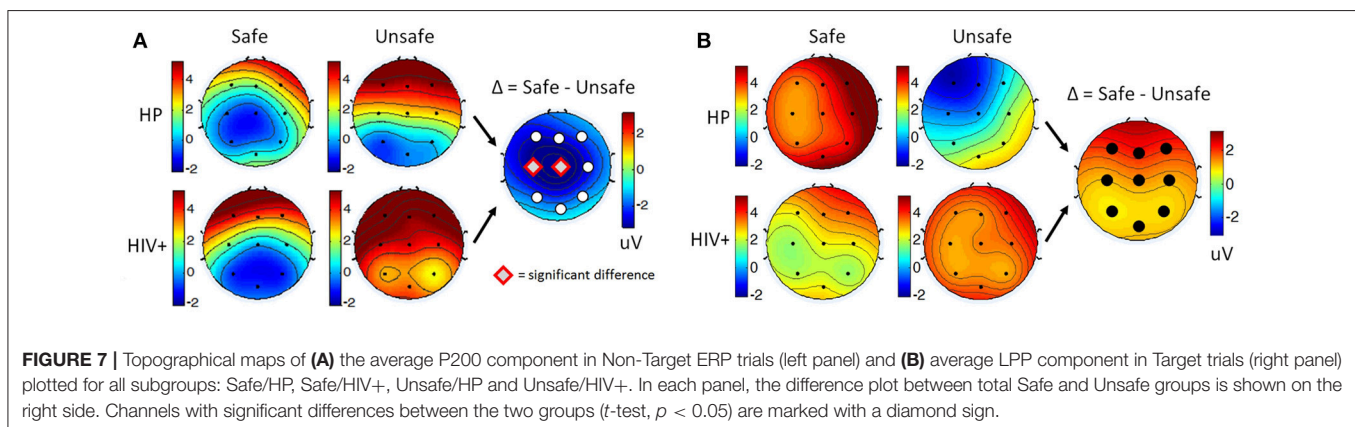
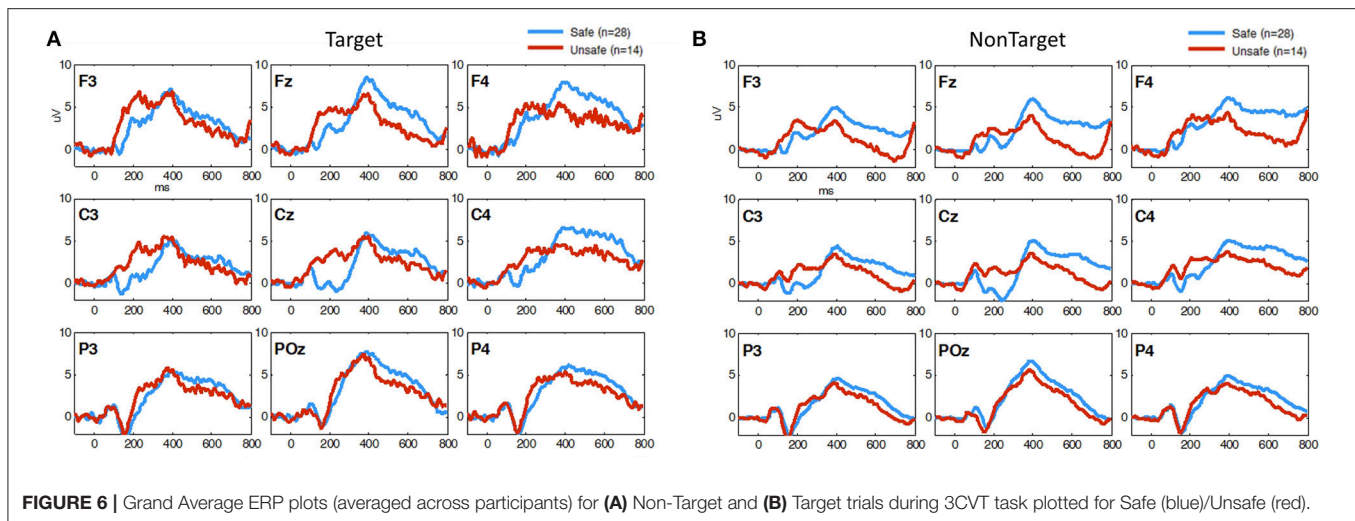
EEG measures obtained during 3CVT were compared for each group in order to discover any potential associations between 3CVT EEG measures and driving performance measures.

**Figure 6** shows the grand average ERPs for 3CVT Non-Target trials (left) and Target trials (right) plotted to compare the Safe and Unsafe drivers. On average, Unsafe drivers exhibit higher amplitudes at 200 ms post-stimulus onset and lower amplitude from 300 to 700 ms post-stimulus onset.

For each participant, ERPs were measured using the average of the signal during a window of 180–220 ms post-stimulus onset for the P200 component, and the late positive potential (LPP) was measured using the average of the signal during a window 300–700 ms post-stimulus onset. Safe drivers exhibited a significantly smaller P200 over the left central region for Non-Target trials compared to Unsafe drivers (**Figure 7A**). HP Safe drivers



**FIGURE 5 |** Secondary task performance and driving performance during the SuRT Easy, Medium, and Hard task. Participants performed worse, as indicated by all metrics except average speed, on the most difficult SuRT task.



exhibited a significantly larger LPP over the left frontal region compared to HP Unsafe drivers for Target trials (**Figure 7B**). There was no significant difference between HIV Safe and HIV Unsafe in terms of LPP amplitude (**Figure 7B**). Additionally, there was no significant difference in LPP amplitude when comparing Safe and Unsafe drivers from both groups. **Table 1** summarizes the significant findings. The difference in the P200 and LPP components between Safe and Unsafe drivers are listed for both trial types (Target and Non-Target) and for all channels in **Supplementary Table 3**.

**Figure 8** shows the grand average ERPs for 3CVT Non-Target trials (left) and Target trials (right) plotted to compare the Good and Poor drivers in the simulator. On average, Poor drivers exhibit higher amplitudes at 200 ms post-stimulus onset, and lower LPP amplitude from 300 to 700 ms post-stimulus onset.

Overall, Poor drivers had a significantly higher P200 over left frontal-central channels (**Figure 9A**) and a significantly lower LPP amplitude over left frontal channels (**Figure 9B**) compared to Good drivers. **Table 2** summarizes the P200 findings and **Table 3** summarizes the LPP findings for all significant channels. The difference in the P200 and LPP components between Good

and Poor drivers are listed for both trial types (Target and Non-Target) and for all channels in **Supplementary Table 4**.

## EEG Measures During the Simulator and On-Road Drive

EEG was acquired during the simulated driving scenario as well as the on-road drive in order to identify any possible real-time neurophysiological differences associated with driving performance. However, there were no significant findings.

## EEG and Behavioral Measures in Relation to Cognitive Status

Cognitive status (impaired vs. unimpaired, see section Cognitive and Medical Assessment) was not correlated with any of the behavioral, EEG, and driving performance measures included in this study.

## Classifier for Predicting On-Road Driving Performance

At the operating point the true positive rate and false positive rate of the classifier were 0.85 and 0.23, respectively. The area under

ROC curve was also used as an overall measure of classification performance. The results ( $AUC = 0.88$ ) were compared with another LDF using only performance measures obtained from the driving simulator as the predictors (see Driving Simulator

Performance) resulting in  $AUC = 0.73$ . The true positive and false positive rate at the operating point of this second classifier was 0.64 and 0.21, respectively. **Figure 10** shows the ROC curve for both classifiers. The higher performance of the EEG-based classifier, as opposed the classifier based on simulator data, demonstrates the power of EEG measures during an attention task in predicting on-road driving performance.

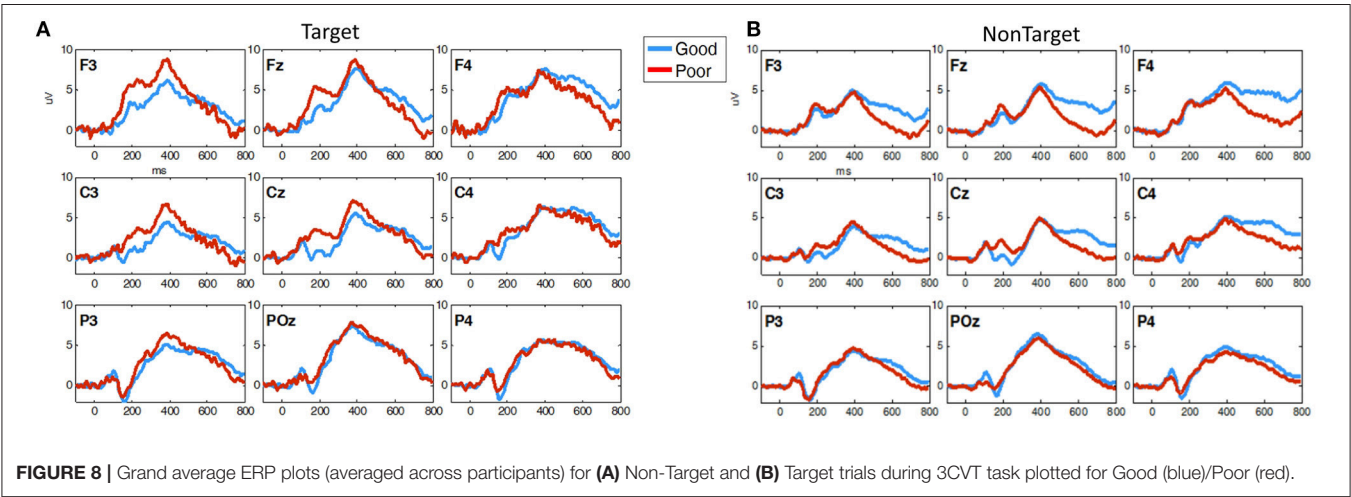
**TABLE 1 |** Average P200 components for all groups and subgroups based on on-road driving performance.

| Condition | Group                         | n  | P200 avg (NonTarget)        |                             |
|-----------|-------------------------------|----|-----------------------------|-----------------------------|
|           |                               |    | Mean $\pm$ SEM ( $\mu V$ )  |                             |
|           |                               |    | Cz                          | C3                          |
| HP        | Safe                          | 19 | $-1.07 \pm 0.96$            | $-0.31 \pm 0.80$            |
|           | Unsafe                        | 7  | $2.76 \pm 1.36$             | $2.30 \pm 0.95$             |
| HIV+      | Safe                          | 9  | $1.32 \pm 1.89$             | $2.37 \pm 1.35$             |
|           | Unsafe                        | 7  | $3.34 \pm 1.94$             | $3.97 \pm 1.98$             |
| All       | Safe                          | 28 | $-0.30 \pm 0.90$            | $0.55 \pm 0.72$             |
|           | Unsafe                        | 14 | $3.05 \pm 1.14$             | $3.14 \pm 1.08$             |
| HP        | $\Delta = \text{Safe-Unsafe}$ | 26 | <b><math>-3.83^*</math></b> | $-2.62$                     |
| HIV+      | $\Delta = \text{Safe-Unsafe}$ | 16 | $-2.02$                     | $-1.60$                     |
| All       | $\Delta = \text{Safe-Unsafe}$ | 42 | <b><math>-3.35^*</math></b> | <b><math>-2.59^*</math></b> |

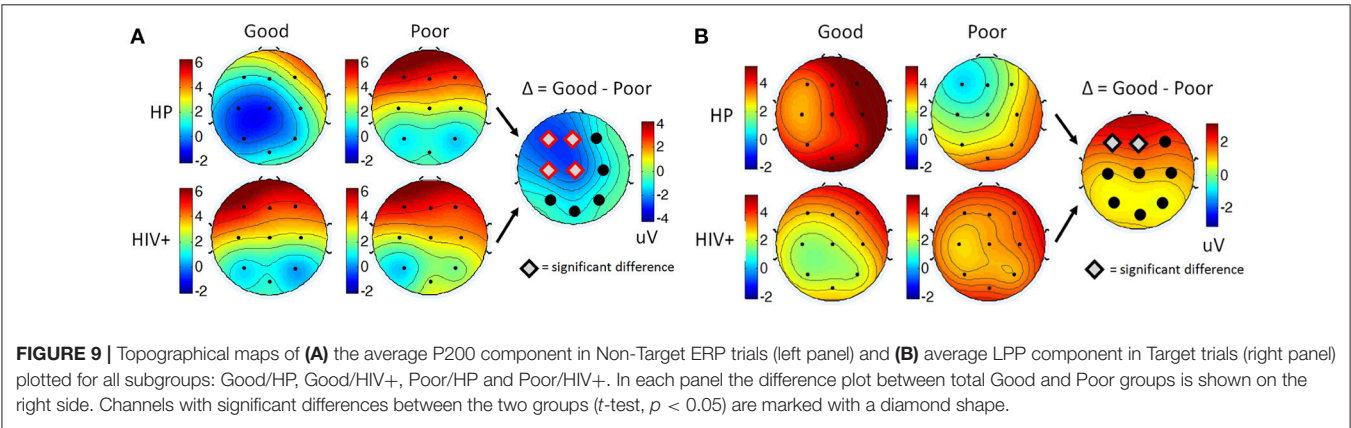
Significant differences (*t*-test,  $p < 0.05$ ) are marked with asterisk.

## DISCUSSION

Evidence from the present study revealed an association between on-road driving performance and EEG ERP data obtained during a short neurocognitive test of sustained attention (3CVT). The 3CVT EEG ERP measures were related to driving performance during a driving simulator task as well as an on-road driving evaluation. Unsafe on-road drivers and Poor drivers in the simulator both exhibited significantly larger P200 amplitude over the left frontal-central region compared to Safe (on-road) and Good (simulator) drivers, respectively. While this finding was observed for Target (frequent) and Non-Target (less frequent) trials, it was largest in response to Non-Target trials during the 3CVT. The P200 component is believed to index automatic, stimulus-driven allocation of attention to stimuli and may



**FIGURE 8 |** Grand average ERP plots (averaged across participants) for (A) Non-Target and (B) Target trials during 3CVT task plotted for Good (blue)/Poor (red).



**FIGURE 9 |** Topographical maps of (A) the average P200 component in Non-Target ERP trials (left panel) and (B) average LPP component in Target trials (right panel) plotted for all subgroups: Good/HP, Good/HIV+, Poor/HP and Poor/HIV+. In each panel the difference plot between total Good and Poor groups is shown on the right side. Channels with significant differences between the two groups (*t*-test,  $p < 0.05$ ) are marked with a diamond shape.



**TABLE 2 |** Average P200 amplitude for all groups and subgroups based on simulator driving performance.

|      |               |  | <i>n</i> | P200 avg (NonTarget)<br>Mean ± SEM (µV) |                |                |                |
|------|---------------|--|----------|---|----------------|----------------|----------------|
|      |               |  |          | Cz                                      | C3             | Fz             | F3             |
| HP   | Good          |  | 30       | −0.84 ± 0.73                            | −0.63 ± 0.58   | 1.15 ± 0.79    | 1.19 ± 0.68    |
|      | Poor          |  | 13       | 2.92 ± 0.90                             | 2.88 ± 0.78    | 4.70 ± 1.01    | 5.7 ± 1.06     |
| HIV+ | Good          |  | 17       | 2.90 ± 0.89                             | 3.19 ± 0.62    | 4.48 ± 1.13    | 5.55 ± 0.99    |
|      | Poor          |  | 10       | 3.54 ± 1.89                             | 2.91 ± 1.60    | 5.03 ± 1.66    | 5.39 ± 1.84    |
| All  | Good          |  | 47       | 0.51 ± 0.61                             | 0.75 ± 0.51    | 2.34 ± 0.68    | 2.77 ± 0.63    |
|      | Poor          |  | 23       | 3.19 ± 0.92                             | 2.89 ± 0.80    | 4.84 ± 0.90    | 5.56 ± 0.97    |
| HP   | Δ = Good-Poor |  | 43       | <b>−3.77**</b>                          | <b>−3.50**</b> | <b>−3.57**</b> | <b>−4.51**</b> |
| HIV+ | Δ = Good-Poor |  | 27       | −0.64                                   | 0.28           | −0.55          | 0.16           |
| All  | Δ = Good-Poor |  | 70       | <b>−2.68*</b>                           | <b>−2.14*</b>  | <b>−2.50*</b>  | <b>−2.80*</b>  |

Significant differences (t-test, \**p* < 0.05, \*\**p* < 0.01) are marked with an asterisk.

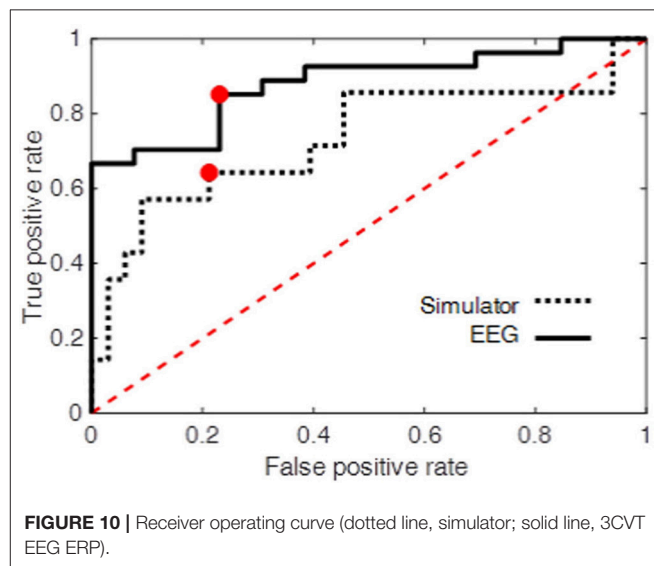
**TABLE 3 |** Average LPP amplitude for all groups and subgroups based on simulator driving performance.

|      |               | <i>n</i> | LPP (Target) Mean ± SEM (uV) |              |
|------|---------------|----------|------------------------------|--------------|
|      |               |          | Fz                           | F3           |
| HP   | Good          | 32       | 3.94 ± 0.70                  | 2.89 ± 0.60  |
|      | Poor          | 16       | 0.49 ± 1.58                  | −0.12 ± 1.34 |
| HIV+ | Good          | 19       | 2.74 ± 0.78                  | 2.76 ± 0.77  |
|      | Poor          | 11       | 2.56 ± 1.34                  | 2.04 ± 1.61  |
| All  | Good          | 51       | 3.49 ± 0.52                  | 2.84 ± 0.47  |
|      | Poor          | 27       | 1.33 ± 1.08                  | 0.75 ± 1.03  |
| HP   | Δ = Good-Poor | 48       | <b>3.45*</b>                 | <b>3.02*</b> |
| HIV+ | Δ = Good-Poor | 30       | 0.19                         | 0.72         |
| All  | Δ = Good-Poor | 78       | <b>2.16*</b>                 | <b>2.08*</b> |

Significant differences (t-test, \**p* < 0.05, \*\**p* < 0.01) are marked with an asterisk.

reflect biases for preferential processing of particular types of stimuli (Eldar et al., 2010; Gole et al., 2012; McIntosh et al., 2015). In this study, the association between P200 amplitude and driving performance may be linked to deficits in selective attention. Bad drivers exhibit impaired ability to maintain focus, improper allocation of attention, and are more easily distracted. In a separate study in which 3CVT EEG ERP biomarkers were evaluated in patients with a neurodegenerative disease affecting memory (amnesic MCI), no P200 differences were observed compared to healthy controls (Waninger et al., 2018). These amnesic MCI patients did not present with noticeable attentional deficits.

Additionally, Unsafe on-road drivers and Poor drivers in the simulator both exhibited a lower LPP amplitude over the frontal region, particularly for Target trials, compared to Safe (on-road) and Good (simulator) drivers, respectively. The late positive potential (LPP) has been shown to reflect feature evaluation, memory matching, and decision making (Withaar et al., 2000; Reger et al., 2004; Meghdadi et al., under review). Multiple reports suggest reduced amplitude of the LPP is associated with



cognitive decline (Schanke and Sundet, 2000; Charlton et al., 2003; Kay et al., 2008; Cysique et al., 2009; Versijpt et al., 2017; Department of Motor Vehicles, 2018a,c; Meghdadi et al., under review) and normal aging (Polich and Corey-Bloom, 2005; Babiloni et al., 2006, 2010; Olichney et al., 2008; López et al., 2014; Ishii et al., 2017). The association of bad driving and reduced amplitude of the LPP reported in the present study is consistent with previous studies that reported a correlation between LPP reduction and severity of cognitive impairment (Polich and Corey-Bloom, 2005; Garn et al., 2014).

The current study included healthy participants (HP) as well as HIV+ participants with well-controlled immune function as a result of antiretroviral therapy. Although current antiretrovirals are increasing the longevity and overall health of HIV+ individuals, HAND is still prevalent and may affect driving performance. The present study included only participants over the age of 55 due to the high likelihood

of age-related decline in driving performance. There were no significant differences observed in driving performance between the HIV+ and healthy groups. In fact, the proportion of bad drivers was equivalent for both groups. Bad drivers (Unsafe or Poor) exhibited an increase in P200 amplitude independent of HIV status with highest observed P200 amplitude in HIV+ Unsafe (or Poor) drivers and lowest P200 amplitude in HP Safe (or Good) drivers. Cognitive status as measured by standard neuropsychological testing (see Cognitive and Medical Assessment) did not correlate with P200 amplitude.

Additionally, group differences were observed in the LPP during 3CVT, with the association between bad driving performance and the reduced amplitude of the LPP only significant for the HP group. While bad drivers (Unsafe or Poor) in the HP group show a significant decrease in LPP compared to HP Safe or Good drivers, this reduction was not observed for the HIV+ group. This may be because the LPP has already been significantly reduced as a result of HIV seropositivity (Hillyard et al., 1973; Polich et al., 2000; Olichney et al., 2011; Papaliagkas et al., 2011).

The classifier used both P200 and LPP metrics to predict drivers as either Safe or Unsafe. However, variables selected by the stepwise feature selection and the results from 3CVT ERP data of the present study suggest the P200 is a stable and reliable predictor of driving performance. Preliminary results suggest this P200 effect is consistently observed across other tests of focused and divided attention (Meghdadi et al., under review).

While EEG measures acquired during the 3CVT sustained attention task were highly associated with driving performance, analysis of the EEG measures acquired in the driving simulator and on-road drive did not significantly predict driving performance. The complexity of the driving scenarios and varying driving strategies employed by participants did not allow for precise event locked EEG analyses as was the case for 3CVT. Although participants were instructed to follow the rules of the road, the completion time for each segment of the driving scenario and on-road drive varied widely between participants. The only highly controlled segment of the either task was the SuRT task performed during the simulated driving scenario. SuRT task difficulty was inversely correlated with SuRT driving and secondary task performance. However, neither was correlated with overall simulated or on-road driving performance.

In this study, EEG ERPs observed during attention tasks and their relation to driving performance provide the basis for an inexpensive, fast, and reliable screening exam for elderly drivers using only EEG acquired concurrently during attention tasks. Performance in the driving simulator alone provided only a reasonable prediction of on-road driving performance but was not nearly as accurate as the 3CVT EEG-based classifier.

Driving is an essential aspect of maintaining independence, but driving ability can begin to deteriorate as people age. Through natural aging or disease-related causes, functional impairments can impede elderly drivers from driving safely.

ERP measures (P200 and LPP) described in this study are shown to reliably predict driving performance in both healthy and HIV+ individuals across a broad age spectrum (55–87 years old). A diagnosis of a neurodegenerative disease (MCI, PDD, HAND, AD, etc.) alone does not necessarily mean an individual is too impaired to drive safely. In the present study, standard neuropsychological testing was not predictive of driving performance. Currently, there is no sensitive test to determine if an individual is actually impaired except for an on-road drive with a driving examiner. To address this unmet need, a portable EEG system could be used to perform a short and inexpensive neurocognitive test to obtain ERP data for any patient. This ERP data could in turn be fed into a classifier to determine whether or not an individual requires an on-road driving evaluation (classifier responded Unsafe or Safe). While there is a false positive rate of 23%, this approach offers a much better alternative than requiring on-road evaluations for all older or cognitively impaired drivers. Additionally, the model will be improved and refined by increasing the size of the dataset with other populations currently being studied.

Future research is required to fully describe the P200 effect by implementing different types of tasks designed to activate neural circuitry associated with varying aspects of attention and cognition. In the field of driving assessment, further experiments with larger and more diverse populations (including drivers with a variety of neurodegenerative diseases) are needed. A more in-depth analysis of driving performance is also needed to further understand the specific functional deficits associated with increased P200 amplitude.

## DATA AVAILABILITY STATEMENT

A link to download the de-identified data (.edf files) will be made available upon request.

## ETHICS STATEMENT

The protocols in the study were approved by both the University of California, San Diego (UCSD) IRB and Sharp IRB (IRBANA) with written informed consent of all subjects. The authors only received de-identified data. HIPPA guidelines were followed throughout the study to protect patient privacy.

## AUTHOR CONTRIBUTIONS

TM and CB conceived the present project, and CB supervised the project. GR wrote the manuscript with the help of CB, AM, SS, and TM. ES and KM collected the in-lab data, and GR collected the data in the field. AM, MK, MC, and GR analyzed the data. TM, ES, KM, and TR designed and implemented the simulated drive and TM designed the on-road drive. CB, AM, GR, MC, and MK interpreted the results.

## FUNDING

This work was supported by the NIH [Grant Number: 5R42MH097303].

## ACKNOWLEDGMENTS

The authors would like to thank the following people for their contributions to various aspects of this project: Vedeline

Torreón, Bradly Stone, Rudy Chang, Robin Johnson, Kyla Manawatao, and Josh Miller.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2018.00532/full#supplementary-material>

## REFERENCES

- Ajinoroozi, M., Mao, Z., Jung, T.-P., and Lin, C.-T., Huang, Y. (2016). EEG-based prediction of driver's cognitive performance by deep convolutional neural network. *Signal Process.* 47, 549–555. doi: 10.1016/j.image.2016.05.018
- Babiloni, C., Binetti, G., Cassarino, A., Dal Forno, G., Del Percio, C., Ferreri, F., et al. (2006). Sources of cortical rhythms in adults during physiological aging: a multicentric EEG study. *Hum. Brain Mapp.* 27, 162–172. doi: 10.1002/hbm.20175
- Babiloni, C., Visser, P. J., Frisoni, G., De Deyn, P. P., Bresciani, L., Jelic, V., et al. (2010). Cortical sources of resting EEG rhythms in mild cognitive impairment and subjective memory complaint. *Neurobiol. Aging* 31, 1787–1798. doi: 10.1016/j.neurobiolaging.2008.09.020
- Ball, K. (2009). Cognitive training: impact on driving and everyday function. *Eye Auto.*
- Ball, K., Owsley, C., Stalvey, B., Roenker, D. L., Sloane, M. E., and Graves, M. (1998). Driving avoidance and functional impairment in older drivers. *Accid. Anal. Prev.* 30, 313–322. doi: 10.1016/S0001-4575(97)00102-4
- Bauer, L. O. (2011). Interactive effects of HIV/AIDS, body mass, and substance abuse on the frontal brain: a P300 study. *Psychiatry Res.* 185, 232–237. doi: 10.1016/j.psychres.2009.08.020
- Berka, C., Ayappa, I., Burschtin, O., Piyathilake, H., Rapoport, D. M., Westbrook, P., et al. (2009). "High throughput brain-behavior assay. Quantification of eeg and performance in patients referred for assessment of daytime drowsiness," in *Sleep* (Westchester, NY: American Academy of Sleep Medicine One Westbrook Corporate CTR). A161–A161.
- Berka, C., Levendowski, D. J., and Davis, G. (2006). "Nicotine administration and withdrawal effects on EEG metrics of attention, memory and workload: implications for cognitive resource allocation," in *Augmented Cognition: Past, Present and Future, Foundations of Augmented Cognition* (Arlington, VA: Strategic Analysis, Inc), 174–183.
- Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., et al. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviat. Space Environ. Med.* 78(Suppl. 5), B231–B244.
- Berka, C., and Marcotte, T. (2017). *A Novel Approach to Assessing Cognitive State During Real-World Tasks*.
- Bonanni, L., Thomas, A., Tiraboschi, P., Perfetti, B., Varanese, S., and Onofri, M. (2008). EEG comparisons in early Alzheimer's disease, dementia with Lewy bodies and Parkinson's disease with dementia patients with a 2-year follow-up. *Brain* 131, 690–705. doi: 10.1093/brain/awm322
- Brayne, C., Dufouil, C., Ahmed, A., Denning, T. R., Chi, L. Y., McGee, M., et al. (2000). Very old drivers: findings from a population cohort of people aged 84 and over. *Int. J. Epidemiol.* 29, 704–707. doi: 10.1093/ije/29.4.704
- Brookhuis, K. A., de Vries, G., and de Waard, D. (1991). The effects of mobile telephoning on driving performance. *Accid. Anal. Prev.* 23, 309–316. doi: 10.1016/0001-4575(91)90008-S
- Brookhuis, K. A., and de Waard, D. (1993). The use of psychophysiology to assess driver status. *Ergonomics* 36, 1099–1110. doi: 10.1080/00140139308967981
- Carr, D. B., LaBarge, E., Dunnigan, K., and Storandt, M. (1998). Differentiating drivers with dementia of the Alzheimer type from healthy older persons with a traffic sign naming test. *J. Gerontol. A Biol. Sci. Med. Sci.* 53, M135–M139. doi: 10.1093/gerona/53A.2.M135
- Chao, L. L., Lindgren, J. A., Flenniken, D. L., and Weiner, M. W. (2004). ERP evidence of impaired central nervous system function in virally suppressed HIV patients on antiretroviral therapy. *Clin. Neurophysiol.* 115, 1583–1591. doi: 10.1016/j.clinph.2004.02.015
- Charlton, J. L., Koppel, S., Odell, M., Devlin, A., Langford, J., O'Hare, M., et al. (2003). *Influence of Chronic Illness on Crash Involvement of Motor Vehicle Drivers*. Clayton, VIC: Monash University. Accident Research Centre (MUARC).
- Chiao, S., Rosen, H. J., Nicolas, K., Wendelken, L. A., Alcantar, O., Rankin, K. P., et al. (2013). Deficits in self-awareness impact the diagnosis of asymptomatic neurocognitive impairment in HIV. *AIDS Res. Hum. Retroviruses* 29, 949–956. doi: 10.1089/aid.2012.0229
- Choi, M., Adams, K. B., and Kahana, E. (2013). Self-regulatory driving behaviors: gender and transportation support effects. *J. Women Aging* 25, 104–118. doi: 10.1080/08952841.2012.720212
- Classen, S., Levy, C., McCarthy, D., Mann, W. C., Lanford, D., and Waid-Ebbs, J. K. (2009). Traumatic brain injury and driving assessment: an evidence-based literature review. *Am. J. Occup. Ther.* 63, 580–591. doi: 10.5014/ajot.63.5.580
- Cysique, L. A., Vaida, F., Letendre, S., Gibson, S., Cherner, M., Woods, S. P., et al. (2009). Dynamics of cognitive change in impaired HIV-positive patients initiating antiretroviral therapy. *Neurology* 73, 342–348. doi: 10.1212/WNL.0b013e3181ab2b3b
- Czigler, B., Csikós, D., Hidasi, Z., Anna Gaál, Z., Csibri, E., Kiss, E., et al. (2008). Quantitative EEG in early Alzheimer's disease patients - power spectrum and complexity features. *Int. J. Psychophysiol.* 68, 75–80. doi: 10.1016/j.ijpsycho.2007.11.002
- Dauwels, J., Vialatte, F., and Cichocki, A. (2010a). Diagnosis of Alzheimer's disease from EEG signals: where are we standing? *Curr. Alzheimer Res.* 7, 487–505.
- Dauwels, J., Vialatte, F., Musha, T., and Cichocki, A. (2010b). A comparative study of synchrony measures for the early diagnosis of Alzheimer's disease based on EEG. *Neuroimage* 49, 668–693. doi: 10.1016/j.neuroimage.2009.06.056
- De Gennaro, L., Vecchio, F., Ferrara, M., Curcio, G., Rossini, P. M., and Babiloni, C. (2005). Antero-posterior functional coupling at sleep onset: changes as a function of increased sleep pressure. *Brain Res. Bull.* 65, 133–140. doi: 10.1016/j.brainresbull.2004.12.004
- de Waard, D., and Brookhuis, K. A. (1991). Assessing driver status: a demonstration experiment on the road. *Accid. Anal. Prev.* 23, 297–307. doi: 10.1016/0001-4575(91)90007-R
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009
- Department of Motor Vehicles (2018a). *State of California Department of Motor Vehicles*. Available online at: [www.dmv.ca.gov](http://www.dmv.ca.gov)
- Department of Motor Vehicles (2018b). *State of Connecticut Department of Motor Vehicles*. Available online at: [www.ct.gov/dmv](http://www.ct.gov/dmv)
- Department of Motor Vehicles (2018c). *State of Delaware Division of Motor Vehicles*. Available online at: [www.dmv.de.gov](http://www.dmv.de.gov)
- Dijksterhuis, C., de Waard, D., Brookhuis, K. A., Mulder, B. L., and de Jong, R. (2013). Classifying visuomotor workload in a driving simulator using subject specific spatial brain patterns. *Front. Neurosci.* 7:149. doi: 10.3389/fnins.2013.00149
- Dong, Y., Hu, Z., Uchimura, K., and Murayama, N. (2010). *Driver Inattention Monitoring System for Intelligent Vehicles: A Review*. IEEE Transactions on Intelligent Transportation Systems.

- Duchek, J. M., Carr, D. B., Hunt, L., Roe, C. M., Xiong, C., Shah, K., et al. (2003). Longitudinal driving performance in early-stage dementia of the Alzheimer type. *J. Am. Geriatr. Soc.* 51, 1342–1347. doi: 10.1046/j.1532-5415.2003.51481.x
- Eldar, S., Yankelevitch, R., Lamy, D., and Bar-Haim, Y. (2010). Enhanced neural reactivity and selective attention to threat in anxiety. *Biol. Psychol.* 85, 252–257. doi: 10.1016/j.biopsycho.2010.07.010
- Fonda, S. J., Wallace, R. B., and Herzog, A. R. (2001). Changes in driving patterns and worsening depressive symptoms among older adults. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 56, S343–S351. doi: 10.1093/geronb/56.6.S343
- Freeman, E. E., Gange, S. J., Muñoz, B., and West, S. K. (2006). Driving status and risk of entry into long-term care in older adults. *Am. J. Public Health* 96, 1254–1259. doi: 10.2105/AJPH.2005.069146
- Frittelli, C., Borghetti, D., Iudice, G., Bonanni, E., Maestri, M., Tognoni, G., et al. (2009). Effects of Alzheimer's disease and mild cognitive impairment on driving ability: a controlled clinical study by simulated driving test. *Int. J. Geriatr. Psychiatry* 24, 232–238. doi: 10.1002/gps.2095
- Garn, H., Waser, M., Deistler, M., Schmidt, R., Dal-Bianco, P., Ransmayr, G., et al. (2014). Quantitative EEG in Alzheimer's disease: cognitive state, resting state and association with disease severity. *Int. J. Psychophysiol.* 93, 390–397. doi: 10.1016/j.ijpsycho.2014.06.003
- Gole, M., Schäfer, A., and Schienle, A. (2012). Event-related potentials during exposure to aversion and its anticipation: the moderating effect of intolerance of uncertainty. *Neurosci. Lett.* 507, 112–117. doi: 10.1016/j.neulet.2011.11.054
- Hajinoroozi, M., Jianqiu, Z., and Yufei, H. (2017). Prediction of fatigue-related driver performance from EEG data by deep riemannian model. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2017, 4167–4170. doi: 10.1109/EMBC.2017.8037774
- Hillyard, S. A., Hink, R. F., Schwent, V. L., and Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science* 182, 177–180. doi: 10.1126/science.182.4108.177
- Ishii, R., Canuet, L., Aoki, Y., Hata, M., Iwase, M., Ikeda, S., et al. (2017). Healthy and pathological brain aging: from the perspective of oscillations, functional connectivity, and signal complexity. *Neuropsychobiology* 75, 151–161. doi: 10.1159/000486870
- Jelic, V., and Kowalski, J. (2009). Evidence-based evaluation of diagnostic accuracy of resting EEG in dementia and mild cognitive impairment. *Clin. EEG Neurosci.* 40, 129–142. doi: 10.1177/155005940904000211
- Johansson, K. (1997). "Traffic Safety in an aging society," in *Transportation, Traffic Safety and Health*, eds H. V. Holst, Nygren, Ake, Thord, and Roland. Springer. doi: 10.1007/978-3-662-03409-5\_9
- Johnson, R., Behneman, A., and Mills, P. (2010). Mitigation of sleep deprivation through omega-3 fatty acids: neurocognitive, inflammatory, EEG, and EKG evidence. *Soc. Neurosci.*
- Johnson, R. R., Popovic, D. P., Olmstead, R. E., Stikic, M., Levendowski, D. J., and Berka, C. (2011). Drowsiness/alertness algorithm development and validation using synchronized EEG and cognitive performance to individualize a generalized model. *Biol. Psychol.* 87, 241–250. doi: 10.1016/j.biopsycho.2011.03.003
- Kawano, N., Iwamoto, K., Ebe, K., Suzuki, Y., Hasegawa, J., Ukai, K., et al. (2012). Effects of mild cognitive impairment on driving performance in older drivers. *J. Am. Geriatr. Soc.* 60, 1379–1381. doi: 10.1111/j.1532-5415.2012.04021.x
- Kay, L., Bundy, A., Clemson, L., and Jolly, N. (2008). Validity and reliability of the on-road driving assessment with senior drivers. *Accid. Anal. Prev.* 40, 751–759. doi: 10.1016/j.aap.2007.09.012
- Kellison, M. R. I. (2009). *Neuropsychology of Everyday Functioning- Chapter 7 The Brain on the Road*. New York, NY: The Guilford Press.
- Lei, S., and Roetting, M. (2011). Influence of task combination on EEG spectrum modulation for driver workload estimation. *Hum. Factors* 53, 168–179. doi: 10.1177/0018720811400601
- Leproust, S., Lagarde, E., and Salmi, L. R. (2008). Systematic screening for unsafe driving due to medical conditions: still debatable. *BMC Public Health* 8:27. doi: 10.1186/1471-2458-8-27
- Levendowski, D., Berka, C., Olmstead, R., Konstantinovic, Z. R., Davis, G., Lumicao, M. N., et al. (2001). Electroencephalographic indices predict future vulnerability to fatigue induced by sleep deprivation. *Sleep* 24, A243–A244.
- Levendowski, D., Olmstead, R. E., Konstantinovic, Z. R., Berka, C., and Westbrook, P. (2000). Detection of electroencephalographic indices of drowsiness in real-time using a multi-level discriminant function analysis. *Sleep* 23, A243–A244.
- López, M. E., Bruña, R., Aurteneixe, S., Pineda-Pardo, J. Á., Marcos, A., Arrazola, J., et al. (2014). Alpha-band hypersynchronization in progressive mild cognitive impairment: a magnetoencephalography study. *J. Neurosci.* 34, 14551–14559. doi: 10.1523/JNEUROSCI.0964-14.2014
- Man-Son-Hing, M., Marshall, S. C., Molnar, F. J., and Wilson, K. G. (2007). Systematic review of driving risk and the efficacy of compensatory strategies in persons with dementia. *J. Am. Geriatr. Soc.* 55, 878–884. doi: 10.1111/j.1532-5415.2007.01177.x
- Marcotte, T. D., Heaton, R. K., Wolfson, T., Taylor, M. J., Alhassoon, O., Arfaa, K., et al. (1999). The impact of HIV-related neuropsychological dysfunction on driving behavior. The HNRC group. *J. Int. Neuropsychol. Soc.* 5, 579–592. doi: 10.1017/S1355617799577011
- Marcotte, T. D., Wolfson, T., Rosenthal, T. J., Heaton, R. K., Gonzalez, R., Ellis, R. J., et al. (2004). A multimodal assessment of driving performance in HIV infection. *Neurology* 63, 1417–1422. doi: 10.1212/01.WNL.0000141920.33580.5D
- Marottoli, R. A., de Leon, C. F. M., Glass, T. A., Williams, C. S., Cooney, L. M., and Berkman, L. F. (2000). Consequences of driving cessation: decreased out-of-home activity levels. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 55, S334–S340. doi: 10.1093/geronb/55.6.S334
- Marottoli, R. A., Mendes de Leon, C. F., Glass, T. A., Williams, C. S., Cooney, L. M., Berkman, L. F., et al. (1997). Driving cessation and increased depressive symptoms: prospective evidence from the New Haven EPESE. Established populations for epidemiologic studies of the elderly. *J. Am. Geriatr. Soc.* 45, 202–206. doi: 10.1111/j.1532-5415.1997.tb04508.x
- McIntosh, R. C., Tartar, J. L., Widmayer, S., and Rosselli, M. (2015). Negative attention bias and processing deficits during the cognitive reappraisal of unpleasant emotions in HIV+ women. *J. Neuropsychiatry Clin. Neurosci.* 27, e32–e39. doi: 10.1176/appi.neuropsych.13090222
- 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. *Hum. Factors* 54, 396–412. doi: 10.1177/0018720812442086
- Mehler, B., Reimer, B., Coughlin, J. F., and Dusek, J. A. (2009). Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transport. Res. Record* 2138, 6–12. doi: 10.3141/2138-02
- Mitler, M. M., Miller, J. C., Lipsitz, J. J., Walsh, J. K., and Wylie, C. D. (1997). The sleep of long-haul truck drivers. *N. Engl. J. Med.* 337, 755–761. doi: 10.1056/NEJM199709113371106
- Mulder, L. J. M. (2004). "Estimating Mental Effort Using Heart Rate and Heart Rate Variability," in *Handbook of Human Factors and Ergonomics Methods*, eds K. A. Brookhuis, A. Hedge, E. Salas, H. W. Hendrick (CRC Press), 1–20.
- Olichney, J. M., Taylor, J. R., Gatherwright, J., Salmon, D. P., Bressler, A. J., Kutas, M., et al. (2008). Patients with MCI and N400 or P600 abnormalities are at very high risk for conversion to dementia. *Neurology* 70, 1763–1770. doi: 10.1212/01.wnl.0000281689.28759.ab
- Olichney, J. M., Yang, J. C., Taylor, J., and Kutas, M. (2011). Cognitive event-related potentials: biomarkers of synaptic dysfunction across the stages of Alzheimer's disease. *J. Alzheimers Dis.* 26(Suppl 3), 215–228. doi: 10.3233/JAD-2011-0047
- Papaliagkas, V. T., Kimiskidis, V. K., Tsolaki, M. N., and Anogiannakis, G. (2011). Cognitive event-related potentials: longitudinal changes in mild cognitive impairment. *Clin. Neurophysiol.* 122, 1322–1326. doi: 10.1016/j.clinph.2010.12.036
- Perrier, J., Jongen, S., Vuurman, E., Bocca, M. L., Ramaekers, J. G., and Vermeeren, A. (2016). Driving performance and EEG fluctuations during on-the-road driving following sleep deprivation. *Biol. Psychol.* 121, 1–11. doi: 10.1016/j.biopsycho.2016.09.010
- Pojman, N., Behneman, D., Kintz, N., Johnson, R., Chung, G., Nagashima, S., et al. (2009b). "Characterizing the Psychophysiological Profile of Expert and Novice Marksmen," in *Foundations of Augmented Cognition (FAC)*, *Neuroergonomics and Operational Neuroscience, Lecture Notes in Computer Science*, Vol. 5638. eds D. D. Schmorrow, I. V. Estabrooke, and M. Grootjen (Berlin: Heidelberg: Springer), 524–532.
- Pojman, N., Johnson, R., Kintz, N., Behneman, L., Popovic, V., and Davis, G. (2009a). "Assessing fatigue using EEG classification metrics during neurocognitive testing," in *Sleep*, eds D. D. Schmorrow, I. V. Estabrooke, and



- M. Grootjen (Westchester, IL: American Academy of Sleep Medicine One Westbrook Corporate Center).
- Polich, J., and Basho, S. (2002). P3a and P3b auditory ERPs in HIV patients receiving anti-viral medication. *Clin. Electroencephalogr.* 33, 97–101. doi: 10.1177/155005940203300305
- Polich, J., and Corey-Bloom, J. (2005). Alzheimer's disease and P300: review and evaluation of task and modality. *Curr. Alzheimer Res.* 2, 515–525. doi: 10.2174/156720505774932214
- Polich, J., Ilan, A., Poceta, J. S., Mitler, M. M., and Darko, D. F. (2000). Neuroelectric assessment of HIV: EEG, ERP, and viral load. *Int. J. Psychophysiol.* 38, 97–108. doi: 10.1016/S0167-8760(00)00133-1
- Ragland, D. R., Satariano, W. A., and MacLeod, K. E. (2005). Driving cessation and increased depressive symptoms. *J. Gerontol. A Biol. Sci. Med. Sci.* 60, 399–403. doi: 10.1093/gerona/60.3.399
- Reger, M. A., Welsh, R. K., Watson, G. S., Cholerton, B., Baker, L. D., and Craft, S. (2004). The relationship between neuropsychological functioning and driving ability in dementia: a meta-analysis. *Neuropsychology* 18, 85–93. doi: 10.1037/0894-4105.18.1.85
- Riccio, C. A., Reynolds, C. R., and Lowe, P. A. (2001). *Clinical Applications of Continuous Performance Tests: Measuring Attention and Impulsive Responding in Children and Adults*. New York, NY: John Wiley & Sons, Inc.
- Rookhuis, K. A., De Vries, G., and De Waard, D. (1993). Acute and subchronic effects of the H1-histamine receptor antagonist ebastine in 10, 20 and 30 mg dose, and triprolidine 10 mg on car driving performance. *Br. J. Clin. Pharmacol.* 36, 67–70. doi: 10.1111/j.1365-2125.1993.tb05894.x
- Ross, L. A., Anstey, K. J., Kiely, K. M., Windsor, T. D., Byles, J. E., Luszcz, M. A., et al. (2009). Older drivers in Australia: trends in driving status and cognitive and visual impairment. *J. Am. Geriatr. Soc.* 57, 1868–1873. doi: 10.1111/j.1532-5415.2009.02439.x
- Sateia, M. J. (2003). Neuropsychological impairment and quality of life in obstructive sleep apnea. *Clin. Chest Med.* 24, 249–259. doi: 10.1016/S0272-5231(03)00014-5
- Schanke, A. K., and Sundet, K. (2000). Comprehensive driving assessment: neuropsychological testing and on-road evaluation of brain injured patients. *Scand. J. Psychol.* 41, 113–121. doi: 10.1111/1467-9450.00179
- Silverstein, N. M., Flaherty, G., and Tobin, T. S. (2002). *Dementia and Wandering Behavior: Concern for the Lost Elder*. New York, NY: Springer.
- Stikic, M., Larlus, D., Ebert, S., and Schiele, B. (2011). Weakly supervised recognition of daily life activities with wearable sensors. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 2521–2537. doi: 10.1109/TPAMI.2011.36
- Stoiljkovic, M., Kelley, C., Horvath, T. L., and Hajós, M. (2018). “Neurophysiological biomarkers for alzheimer's disease (ad) & mild cognitive impairment (mci) acquired during sleep & waking,” in AAT- AD/PD. Torino.
- Tefft, B. C. (2017). *Rates of Motor Vehicle Crashes, Injuries and Deaths in Relation to Driver Age, United States, 2014–2015*. Available online at: <https://aaafoundation.org/rates-motor-vehicle-crashes-injuries-deaths-relation-driver-age-united-states-2014-2> (Accessed April 26, 2018).
- Tsolaki, A., Kazis, D., Kompatsiaris, I., Kosmidou, V., and Tsolaki, M. (2014). Electroencephalogram and Alzheimer's disease: clinical and research approaches. *Int. J. Alzheimers. Dis.* 2014:349249. doi: 10.1155/2014/349249
- Uc, E. Y., Rizzo, M., Anderson, S. W., Shi, Q., and Dawson, J. D. (2004). Driver route-following and safety errors in early Alzheimer disease. *Neurology* 63, 832–837. doi: 10.1212/01.WNL.0000139301.01177.35
- Uc, E. Y., Rizzo, M., Anderson, S. W., Shi, Q., and Dawson, J. D. (2005). Driver landmark and traffic sign identification in early Alzheimer's disease. *J. Neurol. Neurosurg. Psychiatr.* 76, 764–768. doi: 10.1136/jnnp.2004.049338
- Vecchio, F., Babiloni, C., Lizio, R., Fallani, F. V., Blinowska, K., Verrienti, G., et al. (2013). Resting state cortical EEG rhythms in Alzheimer's disease: toward EEG markers for clinical applications: a review. *Suppl. Clin. Neurophysiol.* 62, 223–236. doi: 10.1016/B978-0-7020-5307-8.00015-6
- Versijpt, J., Tant, M., Beyer, I., Bier, J. C., Cras, P., De Deyn, P. P., et al. (2017). Alzheimer's disease and driving: review of the literature and consensus guideline from Belgian dementia experts and the Belgian road safety institute endorsed by the Belgian Medical Association. *Acta Neurol. Belg.* 117, 811–819. doi: 10.1007/s13760-017-0840-5
- Wadley, V. G., Okonkwo, O., Crowe, M., Vance, D. E., Elgin, J. M., Ball, K. K., et al. (2009). Mild cognitive impairment and everyday function: an investigation of driving performance. *J. Geriatr. Psychiatry Neurol.* 22, 87–94. doi: 10.1177/0891988708328215
- Waninger, S., Berka, C., Meghdadi, A., Karic, M. S., Stevens, K., Aguero, C., et al. (2018). Event-related potentials during sustained attention and memory tasks: utility as biomarkers for mild cognitive impairment. *Alzheimers Dement.* 10, 452–460. doi: 10.1016/j.dadm.2018.05.007
- Westbrook, P., Berka, C., Levendowski, J., Lumicao, M. N., Davis, G., and Olmstead, O. N. (2002). Biobehavioral quantification of alertness and memory in patients with sleep apnea. *Sleep* 25, A49–A50.
- Withaar, F. K., Brouwer, W. H., and van Zomeren, A. H. (2000). Fitness to drive in older drivers with cognitive impairment. *J. Int. Neuropsychol. Soc.* 6, 480–490. doi: 10.1017/S1355617700644065
- Zwinkels, J. C., Davidson, W. F., and Dodd, C. X. (1990). Optical properties of UV transmitting acrylics for use in a heavy water Cerenkov detector. *Appl. Opt.* 29, 3240–3248. doi: 10.1364/AO.29.003240

**Conflict of Interest Statement:** GR, CB, AM, MK, MC, SS are paid salaries or consulting fees by Advanced Brain Monitoring, and CB is a shareholder of Advanced Brain Monitoring, Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Rupp, Berka, Meghdadi, Karić, Casillas, Smith, Rosenthal, McShea, Sones and Marcotte. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Demonstrating Brain-Level Interactions Between Visuospatial Attentional Demands and Working Memory Load While Driving Using Functional Near-Infrared Spectroscopy

Jakob Scheunemann<sup>1,2†</sup>, Anirudh Unni<sup>1†</sup>, Klas Ihme<sup>3</sup>, Meike Jipp<sup>3</sup> and Jochem W. Rieger<sup>1\*</sup>

## OPEN ACCESS

### Edited by:

Bruce Mehler,  
Massachusetts  
Institute of Technology, United States

### Reviewed by:

Noman Naseer,  
Air University, Pakistan  
Sean Seaman,  
Touchstone Evaluations Inc.,  
United States  
John Lenneman,  
Toyota Collaborative Safety  
Research Center, United States  
Yu-Kai Wang,  
University of Technology Sydney,  
Australia

### \*Correspondence:

Jochem W. Rieger  
jochem.rieger@uni-oldenburg.de

<sup>†</sup>These authors have contributed  
equally to this work

**Received:** 19 June 2018

**Accepted:** 31 December 2018

**Published:** 23 January 2019

### Citation:

Scheunemann J, Unni A, Ihme K,  
Jipp M and Rieger JW (2019)  
Demonstrating Brain-Level  
Interactions Between Visuospatial  
Attentional Demands and Working  
Memory Load While Driving Using  
Functional Near-Infrared  
Spectroscopy.  
Front. Hum. Neurosci. 12:542.  
doi: 10.3389/fnhum.2018.00542

<sup>1</sup> Department of Psychology, University of Oldenburg, Oldenburg, Germany, <sup>2</sup> Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, <sup>3</sup> Institute of Transportation Systems, German Aerospace Center (DLR), Braunschweig, Germany

Driving is a complex task concurrently drawing on multiple cognitive resources. Yet, there is a lack of studies investigating interactions at the brain-level among different driving subtasks in dual-tasking. This study investigates how visuospatial attentional demands related to increased driving difficulty interacts with different working memory load (WML) levels at the brain level. Using multichannel whole-head high density functional near-infrared spectroscopy (fNIRS) brain activation measurements, we aimed to predict driving difficulty level, both separate for each WML level and with a combined model. Participants drove for approximately 60 min on a highway with concurrent traffic in a virtual reality driving simulator. In half of the time, the course led through a construction site with reduced lane width, increasing visuospatial attentional demands. Concurrently, participants performed a modified version of the *n*-back task with five different WML levels (from 0-back up to 4-back), forcing them to continuously update, memorize, and recall the sequence of the previous '*n*' speed signs and adjust their speed accordingly. Using multivariate logistic ridge regression, we were able to correctly predict driving difficulty in 75.0% of the signal samples (1.955 Hz sampling rate) across 15 participants in an out-of-sample cross-validation of classifiers trained on fNIRS data separately for each WML level. There was a significant effect of the WML level on the driving difficulty prediction accuracies [range 62.2–87.1%;  $\chi^2(4) = 19.9$ ,  $p < 0.001$ , Kruskal–Wallis *H* test] with highest prediction rates at intermediate WML levels. On the contrary, training one classifier on fNIRS data across all WML levels severely degraded prediction performance (mean accuracy of 46.8%). Activation changes in the bilateral dorsal frontal (putative BA46), bilateral inferior parietal (putative BA39), and left superior parietal (putative BA7) areas were most predictive to increased driving difficulty. These discriminative patterns diminished at higher WML levels indicating that visuospatial

attentional demands and WML involve interacting underlying brain processes. The changing pattern of driving difficulty related brain areas across WML levels could indicate potential changes in the multitasking strategy with level of WML demand, in line with the multiple resource theory.

**Keywords:** driver state assessment, mental workload, driver workload estimation, visual-motor coordination, visual attention, brain-level interactions, dual-task, fNIRS

## INTRODUCTION

Driving is a complex task, composed of multiple subtasks where different cognitive demands are concurrently imposed on the driver. For instance, one needs to be attentive toward unforeseen events, integrate information from within and outside the vehicle, and control the vehicle to keep it on the lane. All those tasks require cognitive resources of limited capacity (Wickens et al., 2008). Some of these tasks could possibly draw from the same shared resources, leading to a potential interaction between different subtasks.

Working memory plays an important role while driving since the driver has to continuously integrate and dynamically update information from internal and external traffic environments (De Waard, 1996; da Silva, 2014). For example, Wood et al. (2016) have associated increased working memory capacity with better ability to control visual attention while being less distracted in different driving tasks. Further, certain driving situations are associated with increased working memory demands, e.g., left turns at intersections (Guerrier et al., 1999) or driving within a dense city environment (Patten et al., 2006) as they require integration of more items into trajectory planning. Yet, working memory is a capacity-limited system (Baddeley, 2003; Cowan, 2010) and working memory overload deteriorates driving performance (Lavie, 2010). For example, it has been shown that increasing working memory load (WML) via a secondary task decreases driving performance on the lane change task (Ross et al., 2018). Interestingly, this effect was larger for people with less working memory capacity.

Besides working memory, driving requires visuospatial attention and visuomotor control (Vingerhoets and Stroobant, 1999; Lust et al., 2011; Benedetto et al., 2013). Visual attention is demanded because the driver needs to simultaneously integrate central and peripheral vision within a rapidly changing moving environment, while monitoring for unexpected critical events (Owsley and McGwin, 2010). Under decreased vision, more resources are allocated to lane keeping (Gao and Zhang, 2016). More specifically, Brooks et al. (2018) could link a decrement in driving performance in a lane-keeping task to increased peristimulus alpha activity, an indication for poor visuospatial attention. Further, when participants drove in a narrow road condition as compared to the ordinary driving task with normal lane widths, fNIRS measured increased activation in the prefrontal areas (Shimizu et al., 2009). This supports other findings showing that driving in narrowed lanes is more demanding (De Waard et al., 1995; Liu et al., 2016a) and associated with performance loss (Rosey and Auberlet, 2012).

Thus, narrowed lanes seem to increase visuospatial attention load necessary for controlling the vehicle safely.

In driving, different task demands interact with each other (Borghini et al., 2014; Matthews et al., 2015). On the behavioral level, there are various studies that have investigated the effect of cognitive load on driving performance. For example, a majority of the studies suggest that cognitive load actually improves driving performance indicated by improved lane keeping (He and McCarley, 2011; Cooper et al., 2013; for review see Engström et al., 2017). Yet, for studies in which driving difficulty was increased by exposing the car to crosswinds, an additional cognitive load task led to an improvement in lateral control in one study (He et al., 2014), but a drop in another study (Medeiros-Ward et al., 2014).

The interaction of workload and driving performance on the neural level was studied by Wang et al. (2018). In their driving study, using electroencephalography (EEG), car drifts were induced requiring the participant to make lane-keeping adjustments. Additionally, a mathematical calculation task was presented either right before, right after or simultaneously to the induced car drifts. Theta and alpha oscillations in frontal, parietal and occipital areas in the different dual-task conditions were compared to oscillations in single task conditions. While over-additive activation in the frontal theta oscillations were found for the simultaneous condition, all other location-band combinations revealed either additive or under-additive activation in dual-tasking. Vossen et al. (2016) studied the effect of WML on the temporal neural markers for visuospatial attention. Participants performed worse in a visuospatial attention task, in which participants had to react to specific cued visual stimuli in a traffic scenery, when they had to complete an additional verbal memory rehearsal task simultaneously. A further analysis of evoked response potentials (ERPs) from EEG showed that in the high WML conditions, there was a reduction and delay of neural markers only in the early stages of the visuospatial task associated with the initiation of spatial orienting. On the contrary, later stages of the visuospatial task responsible for retaining attentional focus and target selection revealed no differences in the high WML conditions.

The effect of an additional task on the primary driving task was also studied with functional magnetic resonance imaging (fMRI). In a driving simulator study, Just et al. (2008) found a decrease in parietal activation associated with spatial attention in normal driving when participants performed an additional listening comprehension task. As spatial attention and listening comprehension draw resources mostly from non-overlapping

cortical areas, the authors interpret the “diversion of attention as reflecting capacity limit on the amount of attention or resources that can be distributed across the two tasks” (p. 76). Similarly, in a more recent fMRI-driving simulator study, Choi et al. (2017) found a decrease in activation in the parietal areas and an increase in activation in the inferior frontal gyrus and the superior temporal gyrus associated with an additional listening comprehension task while driving. These results illustrate the complex interaction of how an additional task alters the neural activation associated with the primary driving task.

In a cognitive approach on dual-tasking, Wickens (2008) defined resources in his multiple resource theory of attention along four dimensions, namely stages of processing, codes of processing, modalities, and visual channels. The model assumes an interference in dual-tasking when tasks compete for the same resources. For each task, a computational model codes the amount of resources needed for each dimension. For any dimension, if all tasks combined require more resources than what is available, the model predicts interference and performance loss (Wickens, 2002). In an earlier study, the model was implemented to predict driving performance along nine different dual-task combinations consisting of different driving conditions (e.g., urban vs. rural routes) and additional different secondary tasks (e.g., visual vs. auditory backward reading of numbers; Horrey and Wickens, 2003). Performance loss in dual-tasking was successfully predicted by the model for latency of the secondary task and response times to critical road hazards.

An important aspect of the multiple resource theory is executive control, which describes the allocation of resources between tasks. Especially in situations of high dual-task demands, resources might be drawn away from a less prioritized task toward a task with higher priority. Hence, the amount of resources allocated to a subtask depends on the demands of the other subtask, in particular when the other subtask is prioritized. However, how these interactions happen on the brain level in real world tasks is largely unknown. Therefore, in this study, we aimed to investigate at the brain level, how different task demands in one cognitive domain affect the resource allocation for another cognitive domain, by comparing the specificity of predictive brain activation patterns across various dual-tasking scenarios. Specifically, we sought to explore how the assessment of visuospatial attentional driving demands from functional near-infrared spectroscopy (fNIRS) measurements depends on different WML levels.

Functional near-infrared spectroscopy has recently become popular in driving research as a measure of brain activity because it provides brain activations measures with reasonable anatomical and temporal resolution in relatively unconstrained applied settings (Liu et al., 2016b; Sibi et al., 2016). fNIRS uses near-infrared light to measure local concentration changes of deoxygenated hemoglobin (HbR) and oxygenated hemoglobin (HbO) from cortical brain areas which are seen as correlates of functional brain activity (Villringer et al., 1993; Sassaroli and Fantini, 2004). In comparison to HbO, HbR signals are considered to be less influenced by systemic physiological artifacts like cardiac pulsation, respiration, or Mayer wave fluctuations than HbO (Obrig et al., 2000;

Zhang et al., 2005, 2009; Huppert et al., 2009; Suzuki, 2017). Other studies additionally reported that HbR tends to correlate stronger with blood oxygenation level dependent (BOLD) response than HbO (MacIntosh et al., 2003; Huppert et al., 2006; Schroeter et al., 2006; Foy et al., 2016).

In comparison to fMRI, fNIRS has lower spatial (Cui et al., 2011; Mehta and Parasuraman, 2013; Pinti et al., 2018), but better temporal resolution (Huppert et al., 2006). Compared to EEG, fNIRS has lower temporal (Naseer and Hong, 2015), yet better spatial resolution (Scholkmann et al., 2014). Due to its robustness against motion artifacts and external electrical noise, fNIRS is suitable for applied settings (Masataka et al., 2015; Balardin et al., 2017) and has been used in actual driving (Yoshino et al., 2013a,b). fNIRS has shown to be sensitive toward changes in mental workload in the applied fields of simulated flight operation (Ayaz et al., 2012; Durantin et al., 2014), simulated urban rail driving (Li et al., 2018), as well as simulated (Unni et al., 2017; Xu et al., 2017) and actual car driving (Ahn Son et al., 2018). Further, fNIRS could detect elevated visual attention in curve driving, as indicated by increased activity in right premotor cortex, right frontal eye field, and bilateral prefrontal cortex (Oka et al., 2015). Thus, fNIRS is applicable in applied driving settings while providing independent measures of activity in functionally specific brain areas.

In this study [some data has already been published in Unni et al. (2017)], we used fNIRS brain activation measurements obtained during driving to predict two types of cognitive demands: visuospatial attentional demands and working memory demands, both modulated simultaneously. To manipulate visuospatial attentional driving difficulty, participants drove in a 360° Virtual-Reality (VR) driving simulator, half of the time through a construction site with a reduced lane width. At the same time, participants had to perform the primary driving task, which was a working memory speed regulation task (Unni et al., 2017) with five different WML levels. Recording almost whole-head fNIRS brain activation measurements, we aimed at predicting driving difficulty (i.e., driving outside and within construction sites with narrower lane widths) as a measure for visuospatial attentional demands. One of our central questions was whether it is possible to predict driving difficulty or whether task interactions between visuospatial attentional demands and WML levels at the brain level render this impossible. More precisely, we calculated decoding models for the prediction of driving difficulty from almost whole-head fNIRS for each WML level separately and a model which combined fNIRS data over all WML levels. A model with good prediction accuracy for driving difficulty can be interpreted such that there exist distinct neural correlates associated with increased driving difficulty. If there was no interaction between WML and visuospatial attention, a decoding model which combined fNIRS data over all WML levels would perform similarly well in predicting driving difficulty as using a decoding model for each WML level separately. However, if there was an interaction between visuospatial attentional demands and working memory demands, activation patterns associated with increased driving difficulty would differ over WML levels leading to better prediction accuracy for the separate models. Hence, the comparison of prediction



accuracies of the different decoding models characterizes the interaction between the visuospatial attention with working memory processing at the brain level. This is relevant for the development of brain-based driver assistive systems as well as for understanding the nature of the multitasking interactions at the brain level.

## MATERIALS AND METHODS

The experiment was implemented in a driving simulator where participants drove on a highway with varying concurrent traffic. Participants performed a driving task in a two factorial within participant design with factors driving difficulty manipulated by visuospatial attentional demands (two levels: non-construction and construction) and WML (five levels: 0–4 back). The driving difficulty was manipulated via changes of lane width and for WML manipulation, participants performed a digit-span *n*-back speed regulation task. The details of the tasks are provided below.

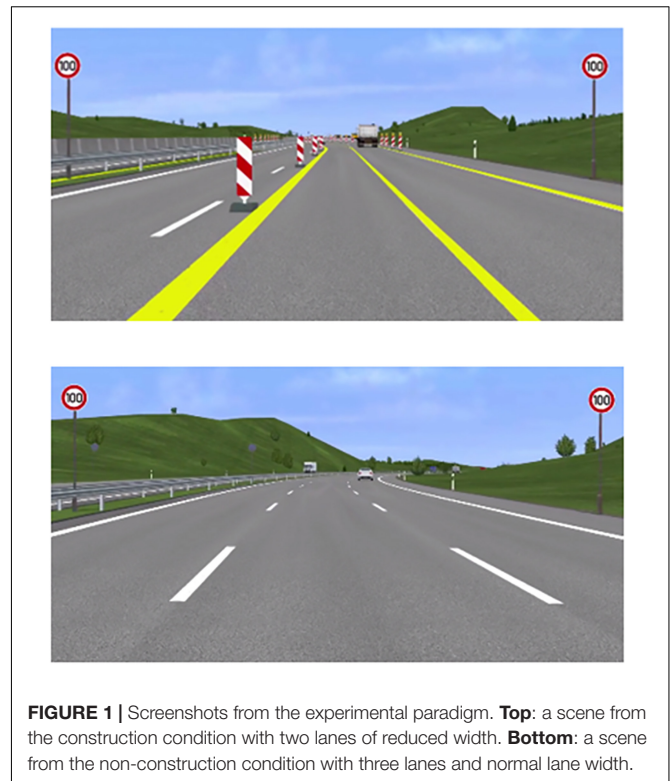
### Participants

Nineteen volunteers (17 males) aged 19–32 years (Mean  $\pm$  SD = 25.2  $\pm$  3.7) participated in the experiment. All participants possessed a valid German driving license at the time of the experiment. Participants gave informed consent prior to the experiment and received a financial reimbursement of 10 € per hour. The experiment was conducted according to the guidelines of the German Aerospace Center and was approved by the Ethics Committee of the Carl von Ossietzky University, Oldenburg.

### Experimental Set-Up

The experiment was set up in a VR-lab at the German Aerospace Research Center allowing a 360° full view (Fischer et al., 2014). During the experiment, participants were operating a realistic vehicle mock-up equipped with common throttle, brake pedal, steering wheel, and indicators. Participants drove on a simulated, slightly curvy highway (64 km in total; developed on the platform Virtual Test Drive, Vires Simulationstechnologie, Bad Aibling, Germany) with varying concurrent traffic. There were 15 vehicles set randomly in an area with a radius of 1000 m around the ego vehicle. Of those vehicles, 60% followed the direction of the ego vehicle; 35% were in the front, 35% in the back, 15% to the left and 15% to the right of the ego vehicle; and 45% were trucks, other 55% were cars.

While driving, fNIRS brain activation measurements were recorded from almost whole-head at a sampling frequency of 1.955 Hertz (Hz) from thirty-two optical emitters and detectors using two NIRScout systems (NIRx Medical Technologies, LLC, United States) in tandem mode. The system uses two wavelengths of 760 and 850 nm to calculate the relative concentration changes of HbO and HbR. We defined 78 fNIRS channels (emitter-detector combinations) in total with an average channel distance of about 3.5 cm. The exact channel locations are provided in Unni et al. (2017). Along with fNIRS data, steering wheel position and driving speed was also recorded at a sampling frequency of 50 Hz.



**FIGURE 1 |** Screenshots from the experimental paradigm. **Top:** a scene from the construction condition with two lanes of reduced width. **Bottom:** a scene from the non-construction condition with three lanes and normal lane width.

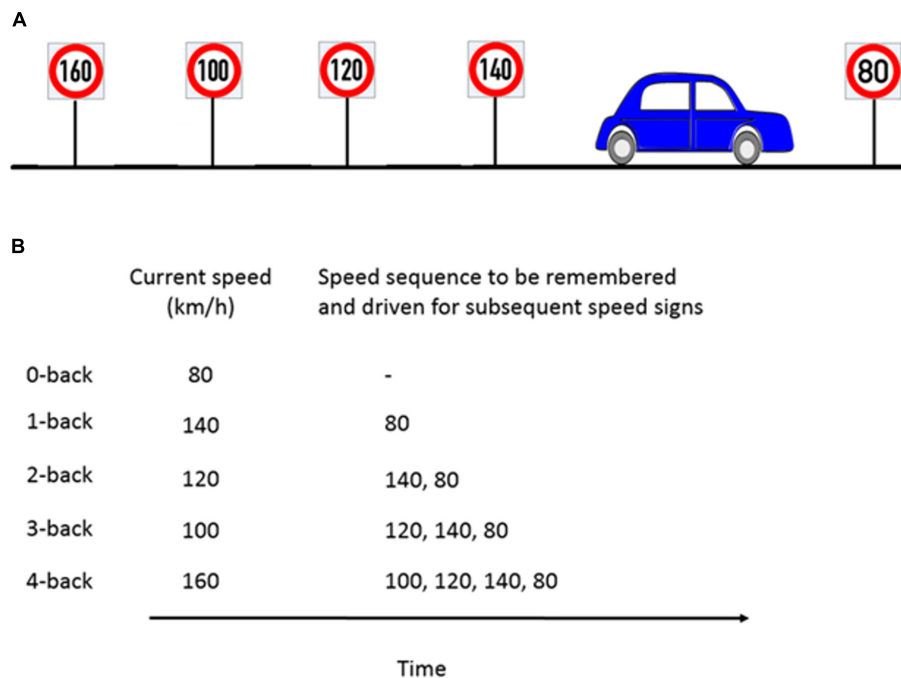
### Visuospatial Attention Manipulation

We manipulated the visuospatial attention demands for the driving task throughout the highway. For about half of the time, participants were driving within a construction site (labeled as *construction*). During the other half of the drive, participants were driving on a normal road without the construction site (labeled as *non-construction*).

The main differences between those two conditions were the number of available lanes and their widths. In the non-construction condition, there were three lanes available with a total width of 10.75 m, consisting of two lanes with a width of 3.5 m (left and center lane) and a slightly wider right lane with a width of 3.75 m. Driving in the construction site was more difficult where only two lanes were available. The widths of the lanes were also reduced along the construction sites with the left and right lanes having a width of 2.5 and 3.5 m, respectively, resulting in a total width of 6 m.

Further, the highway resembled the typical design of German highways. In the non-construction site, there were solid markings in white on the left and right of the road with dashed lines between the lanes. As typical for German highways, pylons marked the beginning and end of the construction sites and yellow markings highlighted the new lanes. The positions and design of the speed signs remained the same.

Screenshots from the experimental paradigm for both conditions can be seen in **Figure 1**. In both conditions, participants had to avoid collisions with other vehicles in ongoing traffic and overtake when it was deemed necessary to drive at the correct speed. Speed signs and WML levels varied at the same rate over both levels of driving difficulty.



**FIGURE 2 |** Example of the  $n$ -back experimental paradigm to manipulate cognitive workload. **(A)** Consider a scenario where the participant is about to pass the 80 km/h speed sign and the previous four speed signs were as shown in the schematic. **(B)** For the corresponding  $n$ -back task, participants had to memorize the last  $n$  speed signs and drive at the  $n$ -th speed sign which occurred previously. For example, at 1-back, the participant's target speed is the previous sign (140 km/h) and has to keep the current speed sign in memory (80 km/h). Figure taken from Unni et al., 2017.

## Working Memory Load Manipulation

The  $n$ -back task is considered to be a benchmark for WML manipulation in neurocognitive psychology (Kirchner, 1958). In a classical  $n$ -back task, a series of numbers, letters, or other stimuli are presented. Participants then have to compare the current stimulus with the stimulus  $n$  steps back and give a response whenever they are the same. We modified the classical  $n$ -back task to be applicable in the driving scenario by using speed signs as stimuli. Participants had to adjust their speed to the speed sign they passed  $n$  speed signs before. For a successful performance, it was necessary that participants continuously update, memorize and recall the previous  $n$  speed signs. Our experiment consisted of five different workload levels from  $n = 0$  (adjusting the speed to the current speed sign) to  $n = 4$  (adjusting the speed to the 4th previous speed sign). The task is illustrated in **Figure 2**. A detailed explanation of the WML speed regulation task can be found in Unni et al. (2017).

Participants had a 6 s window (3 s both before and after passing the sign) to adjust their speed to the target speed. A deviation up to  $\pm 5$  km/h from the target speed was judged as correct. Whenever the deviation was more than  $\pm 5$  km/h, a warning message 'Please pay attention to your speed' was displayed on the screen. This was done to motivate the participant to drive at the correct speed. This message appeared on the screen until the participant drove within the correct speed range. For every new  $n$ -back task, participants were instructed to stay at the speed of the first sign until they passed ' $n$ ' successive speed signs before they could begin with the  $n$ -back task. There were nine

different speed signs (60–140 km/h in steps of 10 km/h) presented in random order to avoid sequencing effects. At the beginning of a new  $n$ -back condition, participants were informed via a message displayed for 5 s on the VR-screen about the next  $n$ -back level to be accomplished.

## Experimental Procedure

The participants started with a 20 min training session where they drove each of the five different  $n$ -back levels twice. Then, the main experiment started, which lasted about 60 min with a break in the middle. In total, the participants performed each 3 min long  $n$ -back level four times, twice in each of the *construction* and *non-construction* conditions. The speed signs were distributed such that the participants passed a new speed sign roughly every 20 s with some temporal jitter. The construction and non-construction sites were alternating with every change in  $n$ -back level. The order of the  $n$ -back levels was pseudorandomized in such a way that the same  $n$ -back level was never driven twice in a row and each  $n$ -back level was performed twice in the construction and the non-construction conditions respectively. Also, the sequence of  $n$ -back levels repeated itself in reversed order after the break to avoid sequencing effects.

## Data Analysis

### Driving Behavior

To determine the effect of increasing WML levels, we calculated error rates in the speed regulation task. As a measure of performance in the working memory task, we calculated the

percentage of time segments in which the participant did not reach the target speed (<90% driving within the tolerance interval around the target speed). In line with the analysis for the fNIRS data described below, we have excluded those time segments (~8% of time segments over all participants) from the other analysis of driving behavior.

In order to check whether driving through the construction site was associated with changes in driving performance, we analyzed the steering reversal rate. Steering reversal rate was defined as the number of times the participant crossed the centered position of the wheel. Steering reversal rate usually increases with increased driving difficulty, as more corrections to the steering wheel position are required (Macdonald and Hoffmann, 1980). As a measure for increased driving difficulty, we calculated the difference in steering reversal rate between driving in the construction and non-construction condition for each *n*-back level.

Due to a problem in data recording in one participant, driving behavior is presented for only 14 participants.

### Working Memory Capacity

To ensure that all participants had comparable levels of working memory capacity, they first performed the memory updating task from the working memory capacity test battery by Lewandowsky et al. (2010). In this test, participants had to remember a set of digits which they had to update continuously through a series of simple arithmetic operations (single digit addition and subtraction). For every correct trial, participants received 1 point. The average total score was 38.4 (*SD* = 10.7) out of a maximum possible score of 60. One participant was excluded from the data analysis, because of a score more than two standard deviations below the mean.

### fNIRS Data Processing

We used the nirsLAB analysis package (Xu et al., 2014) for fNIRS pre-processing. Physiological artifacts (heartbeat, respiration, and Mayer waves) were reduced with a low-pass filter (finite impulse response with least-square error minimization) with a cut-off frequency of 0.1 Hz. We used the Gratzner Spectrum to obtain the molar extinction coefficients of HbO and HbR corresponding to wavelengths of 760 and 850 nm, respectively (Prahl et al., 1999). The corresponding molar extinction coefficients are  $\epsilon_{760} = [1486.59 \ 3843.71]$  and  $\epsilon_{850} = [2526.39 \ 1798.64] \text{ M}^{-1} \cdot \text{cm}^{-1}$  (nirsLAB, NIRx Medical Technologies). The differential path length factor takes into account the increased distance the light path travels from the emitter to the detector because of scattering and absorption effects. The differential path length factors for HbO and HbR were 7.25 and 6.38, respectively (Essenpreis et al., 1993). The relative concentration changes in hemoglobin (mmol/l) were calculated via the modified Beer–Lambert's law (Sassaroli and Fantini, 2004). For the modified Beer–Lambert's law calculation, the exact source-detector distance for each NIRS channel was computed by nirsLAB according to the corresponding distances between emitter and detector pairs on the NIRS cap.

We computed a channel-wise coefficient of variation (CV) which is a measure for the signal-to-noise ratio (SNR) from the

unfiltered raw data. CV is calculated as the ratio of the standard deviation and the mean of each NIRS channel over the entire duration of the experiment (Schmitz et al., 2005; Schneider et al., 2011). All channels with a CV greater than 20% were excluded from further analysis. On average, 64 channels per participant were included in the analysis (*SD* = 7). For the following fNIRS analysis, we have used the HbR signal.

In the fNIRS analysis, we excluded all consecutive time segments between two successive speed signs (~20 s) in which the participant didn't reach the target speed (~8% of time segments over all participants). This was done because we were not sure whether the participant was continuing to focus on the working memory task in those time segments or whether he or she had already given up at an earlier stage due to the inability to focus on the task due to cognitive overload. This is important, as disengagement from difficult tasks reduces the actual cognitive load and affects interpretability of results since workload would be significantly lower than what would be expected on basis of objective task requirements (Victor et al., 2005; Mehler et al., 2012).

A common method to increase the SNR is the application of a Principal Component Analysis (PCA) on the pre-processed fNIRS data (Virtanen et al., 2009). In a PCA, the fNIRS data is transformed to a new set of variables called 'principal components' (PCs) that are linearly uncorrelated and ordered according to the amount of variance explained in the data. It is presumed that motion artifacts contribute more to the variance than the neurophysiological signals and hence the first PC will mostly explain variance dominated by motion artifacts. Therefore, in order to remove motion artifacts, we deleted the first PC, which has shown to be a successful procedure in motion artifact reduction of fNIRS data (Cooper et al., 2012; Brigadoi et al., 2014). Besides motion artifacts, fNIRS data contains noise, for example random instrumental white noise. As we can assume this noise to have a Gaussian distribution, all PCs will contain noise of the same Gaussian distribution. As all PCs contain the same noise variance, first PCs, which explain most of the variance will have a better SNR than later PCs, which explain little variance but will be dominated by the same noise variance and therefore have a worse SNR. That is why we retained only PCs with high exploratory value before transforming the PCs back into the time-series fNIRS data. Based on the recommendation by Jolliffe (1972) on the Kaiser's rule (Kaiser, 1958), all components with eigenvalues larger than 0.7 were kept. With the procedure of deleting the first PC and all other PCs with an eigenvalue smaller than 0.7, 7.09 PCs (*SD* = 2.07) were retained on average over all 15 participants. As detailed in the section below, the PCs are calculated on training data in a cross-validation scheme. These retained PCs were then transformed back to the original space resulting in a less noisy time-series fNIRS data.

### Multivariate Cross-Validated Prediction of Driving Difficulty

Our goal was to predict the driving difficulty, i.e., whether the participant was in the construction or non-construction condition. First, we calculated binary multivariate logistic ridge regression models (Hastie et al., 2009) for the prediction of

driving difficulty from fNIRS data for each WML level, i.e., we calculated separate models for each of the five  $n$ -back levels for each participant. Second, we calculated one binary multivariate logistic ridge regression model to predict driving difficulty from fNIRS data combined over all WML levels for each participant. Both models used time-resolved fNIRS HbR pre-processed data from all the good channels at each timepoint (sampling frequency 1.955 Hz) as one signal sample. From each signal sample, channel-wise weights were used for the model, which were computed using the Glmnet toolbox (Qian et al., 2013). The output of the logistic regression model can be interpreted as a class probability. Consequently, we computed a model output for each signal sample. All samples with a model output of  $p \geq 0.5$  were assigned to the class construction. This allowed us to calculate the rates at which the model correctly classified different conditions.

In this study, we report model accuracy, which indicates the proportion of correctly classified samples as either construction or non-construction. The accuracy was calculated as follows:

$$\text{Accuracy (\%)} = \frac{TP_c + TP_{nc}}{TP_c + TP_{nc} + FP_c + FP_{nc}} * 100$$

Here, the  $TP$  refers to the true positives (number of samples correctly classified) and  $FP$  refers to the false positives (number of samples incorrectly classified) for the two conditions denoted by  $c$  for construction and  $nc$  for non-construction.

While classification accuracy is an intuitive concept to evaluate the performance of a model, it can be biased, e.g., by uneven data sets. In contrast, precision and recall are advantageous performance measures, insensitive to training set size differences (Rieger et al., 2008). Precision provides information about how precise the model is in assigning a particular sample to the respective empirical class ('construction' or 'non-construction'). On the other hand, recall is the proportion of samples belonging to a particular class ('construction' or 'non-construction') which were also assigned to the same class by the model. Here, we report the  $F1$ -scores which are a harmonic average of the precision and recall measures. A  $F1$ -score of 1 indicates perfect precision and recall (Shalev-Shwartz and Ben-David, 2014). The  $F1$ -score for the construction condition was calculated as follows:

$$F1\text{-score} = \frac{2 * TP_c}{2 * TP_c + FP_c + FP_{nc}}$$

In order to test the generalization of the logistic ridge regression model to new data and to avoid overfitting, an out-of-sample nested cross-validation procedure as suggested by Hastie et al. (2009) was used for model training and testing. The outer loop implemented a five-fold cross-validation where the pre-processed fNIRS time-series data was split into five consecutive blocks. In each fold, a different set of four blocks was used as training set to train the model while the left-out block was used to test the generalization of the model. In addition, an inner five-fold cross-validation loop was implemented on the training set where we first performed the PCA of the fNIRS time-series data to reduce noise, after which it was transformed back from PC-space to the original time-series space. Using the Glmnet toolbox (Qian et al., 2013), channel-wise weights for the logistic

regression model were found, for which the  $\lambda$  regularization parameter was optimized internally by Glmnet in the training phase. The cross-validation procedure avoids overfitting of the data to the model and provides an estimate of how well a decoding approach would predict new data in an online analysis (Reichert et al., 2014).

## Univariate Correlation Analysis

Interpreting the channel weights as indicators for brain areas involved with the experimental condition can be difficult as they result from a multivariate model and each weight can only be interpreted in the context of the whole model (Reichert et al., 2014; Weichwald et al., 2015; Holdgraf et al., 2017). To achieve better interpretability, we additionally fitted channel-wise, univariate logistic regression models of the fNIRS HbR data on the driving difficulty for each participant for the separate models. The fNIRS data was the same preprocessed data that was used for the multivariate analysis. To reduce noise and movement artifacts, we used a PCA the same way as for the multivariate analysis. We performed a PCA for each condition and participant, deleted the first and all PCs with an eigenvalue smaller than 0.7 and then transformed it back from PC-space to the original time-series space. To determine model fit, we used the method suggested by Tjur (2009), to calculate  $R^2$  as measure of the predictivity of a channel ( $R^2_{uvr}$ ). The  $Tjur R^2$  varies between 0 (no predictivity) and 1 (perfect predictivity).

We created averaged predictivity maps across all participants ( $Tjur R^2_{avg}$ ) for each fNIRS channel, illustrating the differences in brain activation between construction and non-construction site driving, separately for each  $n$ -back level. Those averages were calculated by weighting the single-subject's univariate coefficient of determination ( $R^2_{uvr}$ ) with prediction accuracy from the multivariate regression analysis:

$$Tjur R^2_{avg}(i) = \frac{\sum_{i, n=1}^{i, n} R^2_{uvr}(i) * \text{Accuracy}(n)}{\sum_1^n \text{Accuracy}(n)}$$

## RESULTS

### Participants

Four participants were excluded from the analysis, three of them due to a large number (> 50%) of noisy fNIRS channels and one due to low performance in the working memory capacity test. Thus, data from fifteen participants, all males, aged 19–32 years (Mean  $\pm$  SD = 25.6  $\pm$  3.96) are included in the following analysis.

### Driving Behavior

#### Steering Reversal Rate

Across all  $n$ -back levels, the steering reversal rate was higher in the construction condition than in the non-construction condition, indicating that the construction site increased driving difficulty (see **Table 1**). Additionally, this difference increased for higher  $n$ -back levels, with exception of the 3-back, indicating that



**TABLE 1 |** Steering Reversal Rate in Hertz.

|  | 0-back          | 1-back          | 2-back          | 3-back           | 4-back           | Mean             |
|--|-----------------|-----------------|-----------------|------------------|------------------|------------------|
| Construction                               | 0.012           | 0.012           | 0.017           | 0.013            | 0.018            | 0.014            |
| Non-construction                           | 0.011           | 0.008           | 0.012           | 0.009            | 0.010            | 0.010            |
| $t_{\text{construction-non-construction}}$ | $t(13) = 2.027$ | $t(13) = 8.123$ | $t(13) = 9.817$ | $t(13) = 15.571$ | $t(13) = 11.445$ | $t(13) = 13.821$ |
| Significance test                          | $p = 0.064$     | $p < 0.001$     | $p < 0.001$     | $p < 0.001$      | $p < 0.001$      | $p < 0.001$      |
| Bonferroni corrected $p^*$ -value          | $p^* = 0.318$   | $p^* < 0.001$   | $p^* < 0.001$   | $p^* < 0.001$    | $p^* < 0.001$    | $p^* < 0.001$    |

**TABLE 2 |** Differences in errors between driving difficulty conditions (construction–non-construction) calculated via paired-sample  $t$ -test and the effect size Cohen's  $d$ .

|  | 0-back                          | 1-back                         | 2-back                          | 3-back                         | 4-back                         | Mean                           |
|--|---------------------------------|--------------------------------|---------------------------------|--------------------------------|--------------------------------|--------------------------------|
| Construction                                     | 0.01                            | 0.05                           | 0.07                            | 0.21                           | 0.17                           | 0.10                           |
| Non-construction                                 | 0.01                            | 0.02                           | 0.13                            | 0.05                           | 0.08                           | 0.06                           |
| Differences between driving difficulty condition | $t(13) = -0.195$<br>$p = 0.849$ | $t(13) = 1.011$<br>$p = 0.331$ | $t(13) = -1.158$<br>$p = 0.268$ | $t(13) = 3.014$<br>$p = 0.010$ | $t(13) = 2.189$<br>$p = 0.046$ | $t(13) = 3.198$<br>$p = 0.007$ |
| Cohen's $d$                                      | -0.06                           | 0.52                           | -0.46                           | 1.10                           | 0.76                           | 0.85                           |
| Bonferroni corrected $p^*$ -value                | $p^* = 1.00$                    | $p^* = 1.00$                   | $p^* = 1.00$                    | $p^* = 0.050$                  | $p^* = 0.230$                  | $p^* = 0.035$                  |

driving difficulty increased with increasing WML levels ( $r = 0.65$ ,  $p < 0.001$ ). This is also supported by a two-factor analysis of variance (ANOVA) with the factors driving difficulty and WML level. For steering reversal rate we observed main effects for both driving difficulty [ $F(1,130) = 146.87$ ,  $p < 0.001$ ] and WML level [ $F(4,130) = 19.08$ ,  $p < 0.001$ ], as well as a significant interaction effect [ $F(4,130) = 10.49$ ,  $p < 0.001$ ]. For additional analysis on lane deviation (see **Supplementary Table S1**).

### Error Rates in WML Speed Regulation Task

We calculated the error rates (percentage of target speeds the participants failed to reach) in the WML speed regulation task in the construction and non-construction condition. A two-factor ANOVA with the factors driving difficulty and WML level revealed main effects of error rates for both driving difficulty [ $F(1,130) = 5.12$ ,  $p = 0.03$ ] and WML level [ $F(4,130) = 6.16$ ,  $p < 0.001$ ], as well as a significant interaction effect [ $F(4,130) = 3.54$ ,  $p < 0.01$ ]. **Figure 3** shows that for all  $n$ -back levels except for 2-back driving in the construction site was accompanied by more errors in the working memory speed regulation task as compared to driving in the non-construction site. This was especially true for the 3-back and 4-back levels (see **Table 2**). The reduced memory performance suggests that increased recruitment of cognitive resources required to meet increasing visuospatial attention demands for the lane-keeping task interacts with cognitive resource recruitment in the working memory task.

## FNIRS Results

### Prediction of Driving Difficulty

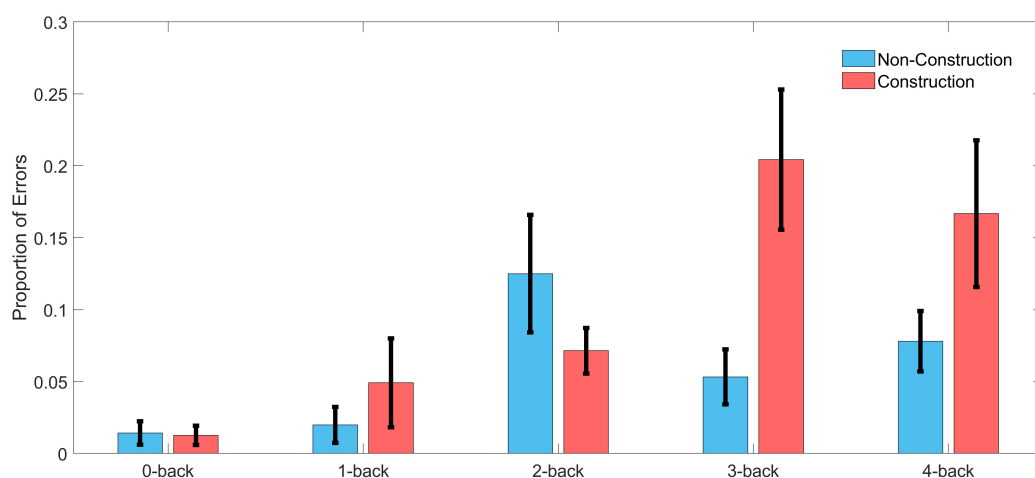
Our goal was to classify the driving difficulty from multivariate logistic ridge regression using pre-processed fNIRS signal samples (sampling frequency 1.955 Hz) in a cross-validation scheme with five equally sized blocks to avoid class size bias. We first calculated separate models for each WML level and each participant. With this procedure, we predicted driving difficulty correctly in 75.0% of the signal samples on average over WML levels and participants. The mean  $F1$ -score was 0.70. The similar

scores between  $F1$ -score and accuracy suggest that the model was not biased to a single class. There was a significant effect of the WML level on the prediction of driving difficulty as indicated by the rank-based non-parametric Kruskal–Wallis  $H$  test for both model accuracy [range: 62.2–87.1%;  $\chi^2(4) = 19.91$ ,  $p < 0.001$ ] and  $F1$ -scores [range: 0.57–0.86;  $\chi^2(4) = 15.46$ ,  $p < 0.01$ ]. Predictions were better for intermediate WML levels (1-back and 2-back) as illustrated in **Figure 4** for model accuracy and **Table 3** for  $F1$ -scores. This pattern of prediction accuracy holds for most individual participants: In 12 out of 15 participants, best model performance  $F1$ -scores were achieved for either 1-back or 2-back.

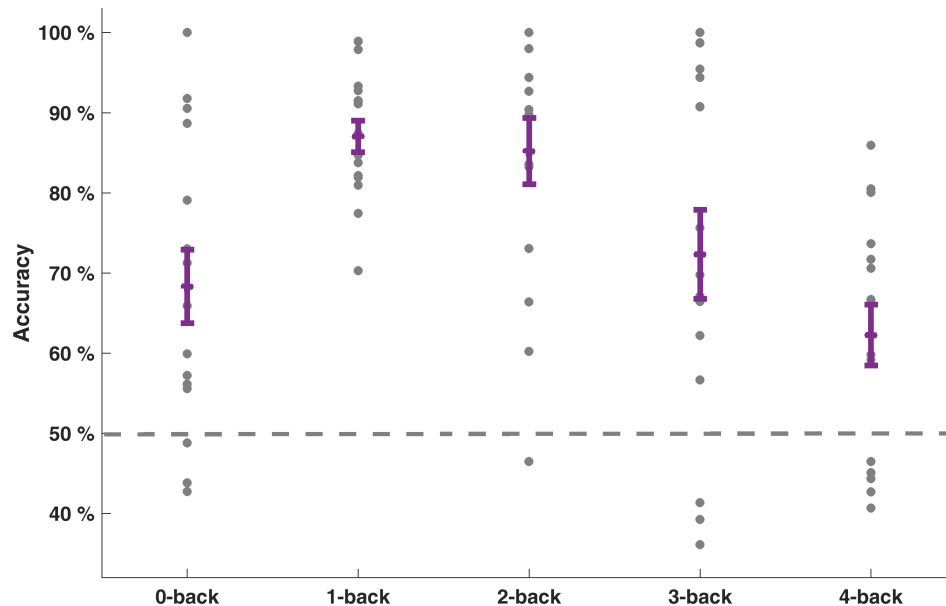
Prediction performance declined, when we used a decoding model that combined the fNIRS data over WML levels to classify driving difficulty. With this procedure, prediction was around chance level with a mean classification accuracy of

**TABLE 3 |**  $F1$ -scores of each classifier for predicting driving difficulty and means across participants and  $n$ -back levels (individual maxima bold).

| Participant | 0-back      | 1-back      | 2-back      | 3-back      | 4-back      | Mean        |
|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 01          | 0.23        | <b>0.77</b> | 0.09        | 0.32        | 0.70        | 0.42        |
| 02          | 0.70        | 0.80        | <b>0.88</b> | 0.23        | 0.72        | 0.67        |
| 03          | 0.62        | 0.85        | <b>0.98</b> | 0.91        | 0.54        | 0.78        |
| 04          | 0.89        | <b>0.98</b> | 0.47        | 0.90        | 0.66        | 0.78        |
| 05          | 0.55        | 0.93        | 0.90        | <b>1.00</b> | 0.21        | 0.72        |
| 06          | 0.90        | 0.86        | <b>0.88</b> | 0.71        | 0.84        | 0.84        |
| 07          | 0.44        | 0.78        | <b>0.94</b> | 0.48        | 0.67        | 0.66        |
| 08          | 0.42        | 0.99        | <b>1.00</b> | 0.74        | 0.31        | 0.69        |
| 09          | 0.77        | 0.68        | 0.89        | <b>0.94</b> | 0.52        | 0.76        |
| 10          | 0.67        | 0.87        | <b>1.00</b> | 0.23        | 0.84        | 0.72        |
| 11          | 0.50        | 0.82        | <b>1.00</b> | 0.99        | 0.33        | 0.73        |
| 12          | 0.73        | 0.88        | <b>1.00</b> | 0.67        | 0.56        | 0.77        |
| 13          | <b>1.00</b> | 0.93        | 0.16        | 0.94        | 0.77        | 0.76        |
| 14          | 0.24        | <b>0.91</b> | 0.77        | 0.27        | 0.44        | 0.53        |
| 15          | 0.89        | <b>0.91</b> | 0.57        | 0.31        | 0.40        | 0.61        |
| Mean        | <b>0.64</b> | <b>0.86</b> | <b>0.77</b> | <b>0.64</b> | <b>0.57</b> | <b>0.70</b> |



**FIGURE 3 |** Error rate in the speed regulation task for driving in construction and non-construction condition for each  $n$ -back level across all participants. Black lines indicate the standard error of the mean ( $n = 15$ ).



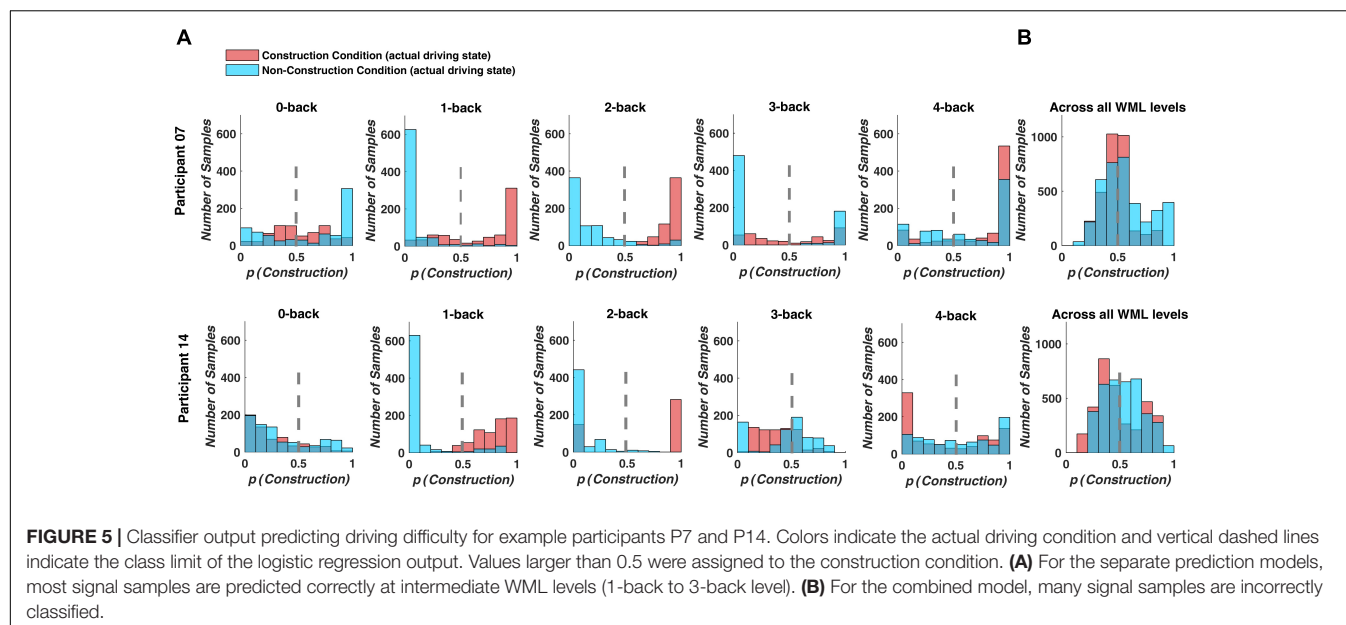
**FIGURE 4 |** Prediction accuracies of driving difficulty for the models separate for each WML level. Individual accuracy score is indicated as dots. Mean accuracy per WML level and its standard error of the mean are depicted in purple. Dashed line at 50% indicates the theoretical guessing level.

**TABLE 4 |** Accuracy and  $F1$ -score of each classifier across participants for the prediction of driving difficulty across all WML levels.

| Participant | P1   | P2   | P3   | P4   | P5   | P6   | P7   | P8   | P9   | P10  | P11  | P12  | P13  | P14  | P15  | Mean |
|-------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| Accuracy(%) | 49   | 54   | 34   | 41   | 61   | 51   | 44   | 41   | 52   | 46   | 42   | 43   | 55   | 42   | 48   | 47   |
| $F1$ -score | 0.44 | 0.44 | 0.40 | 0.42 | 0.57 | 0.50 | 0.42 | 0.18 | 0.44 | 0.31 | 0.35 | 0.55 | 0.52 | 0.38 | 0.37 | 0.42 |

46.8% and a mean  $F1$ -score of 0.419 over all participants (see Table 4). Figure 5 depicts example histograms of the classifier output for two participants. These results show that for separate models (Figure 5A), prediction of driving difficulty is clearly higher than in the combined model (Figure 5A), suggesting an interaction between brain networks modulated by increasing

driving difficulty and brain networks modulated by WML variations. Importantly, this interaction appears to be asymmetric as the reverse was not the case. Unni et al. (2017) demonstrated that WML level can be predicted from fNIRS measurements independent of changes in driving difficulty using data from the same experiment.



**TABLE 5 |** Comparison of mean accuracy for prediction of driving difficulty between predictions within WML levels and adjacent WML levels.

|                                      | 0c                             |      | 1c                             |      | 2c                             |      | 3c                             |      | 4c                             | Mean |
|--------------------------------------|--------------------------------|------|--------------------------------|------|--------------------------------|------|--------------------------------|------|--------------------------------|------|
| Discrimination of driving difficulty | 0nc                            | 1nc  | 1nc                            | 2nc  | 2nc                            | 3nc  | 3nc                            | 4nc  | 4nc                            |      |
| Within WML level                     | 0.68                           |      | 0.87                           |      | 0.85                           |      | 0.72                           |      | 0.62                           | 0.75 |
| Adjacent WML level                   |                                | 0.64 |                                | 0.63 |                                | 0.55 |                                | 0.49 |                                | 0.58 |
| Comparison                           | $t(14) = 0.601$<br>$p = 0.558$ |      | $t(14) = 5.913$<br>$p < 0.001$ |      | $t(14) = 6.311$<br>$p < 0.001$ |      | $t(14) = 2.736$<br>$p = 0.016$ |      | $t(14) = 5.854$<br>$p < 0.001$ |      |

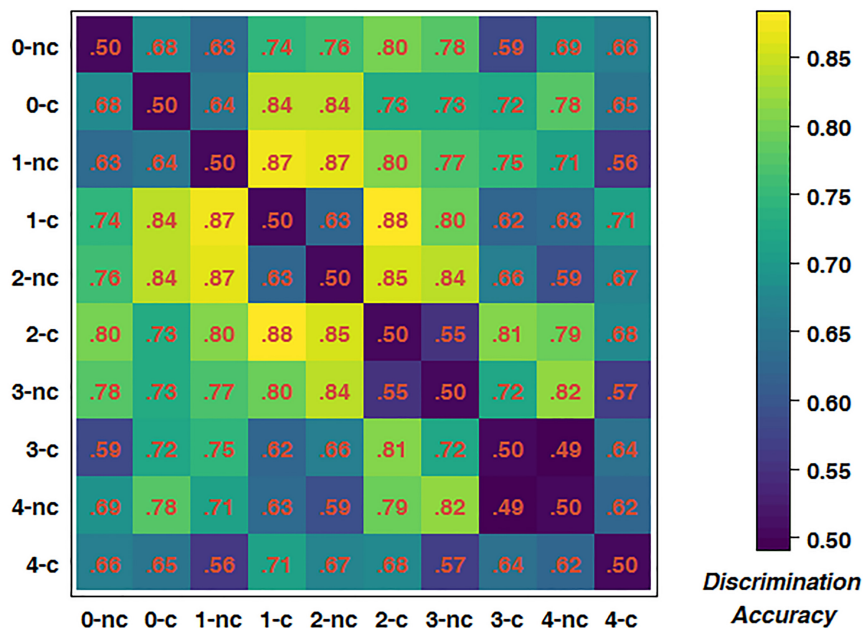
To further test for an interaction between driving difficulty and WML, we trained classifiers for all possible pairings of experimental conditions to obtain a dissimilarity matrix. As there were five WML levels and each WML level consisted of two different driving difficulty levels, there were ten conditions in total, resulting in 45 pairings. **Figure 6** depicts the mean dissimilarity matrix over all participants. Higher discrimination accuracies indicate more reliable changes in brain activations with increasing driving difficulty. In line with the previous analysis, the highest discrimination rates were achieved at intermediate WML levels. This is indicated by accumulation of pairs with higher discrimination rates (depicted by yellow color) in the central areas of the matrix. In addition, a closer analysis of the pattern along the first off-diagonal trace shows an alternating pattern of high and low discrimination accuracies. For example, the 2-back construction brain measurements could be better discriminated from 2-back non-construction than from 3-back non-construction [ $t(14) = 6.311$ ,  $p < 0.001$ ]. This pattern was consistent across other  $n$ -back levels and summarized in **Table 5**. The average prediction accuracy of driving difficulty within the same WML level was 75.0%, whereas the prediction accuracy of driving difficulty for adjacent WML levels was 57.8%, with this difference being significant [ $t(14) = 5.854$ ,  $p < 0.001$ ]. This shows that the driving difficulty became less discriminable by fNIRS data once the WML was increased slightly in the non-construction condition, a pattern that is expected when we

assume interactions of driving difficulty with varying WML at the brain level.

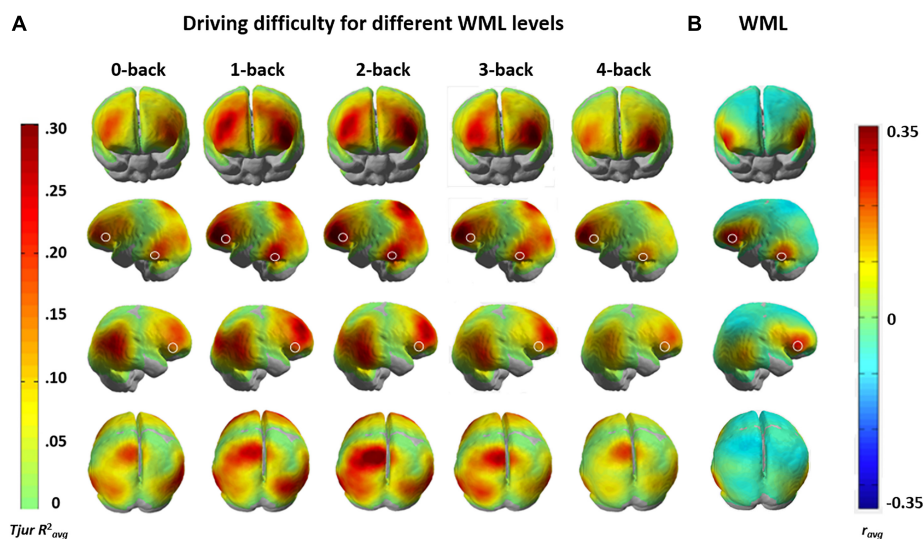
### Localization of Predictive Brain Areas

To gain further insights into the functional anatomy of brain areas associated with increased driving difficulty and their modulation by WML level variation, we calculated channel-wise univariate logistic regressions of HbR levels between the construction and non-construction conditions for each participant and each  $n$ -back level. **Figure 7A** shows the group-level brain maps depicting classification separability, derived as the weighted averaged channel-wise  $T_{jur} R^2$  coefficients ( $T_{jur} R^2_{avg}$ ) from the univariate logistic ridge regression model. The maps show that predictivity of fNIRS activation in the lateral dorsal frontal and parietal areas increases up to the 2-back WML level, while the predictivity of fNIRS activation decreases at higher WML levels (i.e., the 3-back and especially the 4-back levels). This follows the pattern of discriminability variation in the multivariate analysis. These results indicate that the loci of interaction between WML and driving difficulty are in the bilateral dorsal frontal (putative BA 46), bilateral inferior parietal (putative BA 39), and left superior parietal (putative BA 7) areas.

We compared the brain maps to the results from Unni et al. (2017) depicted in **Figure 7B**, where the same fNIRS data was used to predict WML levels independent of driving difficulty (average correlation between predicted and induced



**FIGURE 6 |** Dissimilarity matrix of predictions of all possible pairings of conditions. c, construction; nc, non-construction; corresponding number indicates WML level.



**FIGURE 7 | (A)** Weighted mean of channel-wise predictivity ( $Tjur R^2_{avg}$ ) for driving difficulty at the different WML levels. **(B)** Weighted averaged group-level univariate correlation ( $r_{avg}$ ) HbR brain maps showing brain areas sensitive to changes in WML independent of driving difficulty. White shapes mark WML prediction maxima in all maps. Data for the two analyses were recorded in the same session with concurrent manipulation of driving difficulty and WML.

WML  $r = 0.61$ ). The comparison of the anatomical locations of predictive maxima for WML predictions in **Figure 7B** (marked by white shapes) to **Figure 7A** suggests only partial overlap between the brain resources predictive to the different task demands. Variation of WML level was best predicted in bilateral inferior frontal gyrus (IFG; putative BA 45), an area more posterior to the lateral dorsal frontal areas (putative BA 46) predictive for driving difficulty. An occipito-temporal predictive region (putative BA 21) overlapped between WML and driving

difficulty predictors but appeared more left lateralized in WML prediction, which has a stronger language component. The bilateral inferior and left superior parietal areas (putative BA 39 and BA 7, respectively) which showed increased predictivity to driving difficulty seems to show reduced correlations in the WML level predictions (see **Supplementary Figure S1** for an annotation of putative Brodmann areas). This suggests that these areas are unique to the prediction of driving difficulty, likely involved in visuomotor attention (Jovicich et al., 2001; Caplan



**TABLE 6 |** Brain areas showing predictive maxima of the driving difficulty and WML levels and their corresponding MNI co-ordinates.

| Brain areas                   | Putative Brodmann Area (BA) | X   | Y   | Z  |
|-------------------------------|-----------------------------|-----|-----|----|
| <b>For driving difficulty</b> |                             |     |     |    |
| Right dorsal frontal          | 46                          | 38  | 52  | 30 |
| Left dorsal frontal           | 46                          | −38 | 50  | 30 |
| Right inferior parietal       | 39                          | 44  | −76 | 32 |
| Left inferior parietal        | 39                          | −52 | −66 | 18 |
| Left superior parietal        | 7                           | −18 | −64 | 70 |
| Left occipito-temporal        | 21                          | −68 | −38 | −4 |
| Right occipito-temporal       | 21                          | 68  | 38  | 4  |
| <b>For WML levels</b>         |                             |     |     |    |
| Right inferior frontal gyrus  | 45                          | 52  | 36  | 20 |
| Left inferior frontal gyrus   | 45                          | −52 | 38  | 20 |
| Left occipito-temporal        | 21                          | −68 | −38 | −4 |

et al., 2006) such as vigilance and tracking of moving objects (Culham et al., 1998), but nevertheless their predictivity depends on WML level.

We visualized the averaged brain map on the MNI 152 brain in the Neurosynth<sup>1</sup> and used MRICron<sup>2</sup> to determine MNI co-ordinates and the corresponding Brodmann areas for the brain areas depicting increased predictive discriminability of the driving difficulty. **Table 6** lists the brain areas and their corresponding MNI-co-ordinates of the predictive maxima of the driving difficulty and the WML levels.

## DISCUSSION

In this driving simulator study, we varied visuospatial attention demands by changing the lane widths, thus manipulating driving difficulty while participants performed a modified *n*-back WML speed regulation task. Using almost whole-head fNIRS brain activation measurements, we were able to predict the driving difficulty using a decoding model for each WML level separately. However, the predictions of driving difficulty degraded significantly when we tried to predict driving difficulty using a decoding model which combined fNIRS data over all WML levels.

In order to investigate possible interactions between visuospatial attention and WML, there were two experimental manipulations. To induce different demands in visuospatial attention, participants drove half of the time through a construction site with reduced lane-widths, increasing driving difficulty. At the same time, participants performed a modified *n*-back speed regulation task (0-back to 4-back) resulting in five different levels of WML. Our goal was to predict the driver's current driving difficulty from almost whole-head fNIRS brain activation measurements using a multivariate, cross-validated logistic ridge regression model. As we were interested in understanding if there exists an interaction between

visuospatial attention and WML on a brain level, we predicted driving difficulty with a decoding model which used fNIRS data separately for each WML level and with the same decoding model using fNIRS data combined over all WML levels to compare the decoding accuracies between the models. Our rationale was that if visuospatial attention and working memory had independent underlying brain processes, it should be possible to predict driving difficulty in a combined model across all WML levels. However, this was not the case. In fact, prediction accuracy for driving difficulty across all WML levels was at chance level. Yet, model accuracy improved when the prediction of driving difficulty was calculated separately for each WML level (mean accuracy = 75.0% over all WML levels). Further, there was a significant effect of the WML level on the prediction of driving difficulty.

Thus, we draw two conclusions. First, as driving difficulty could be predicted separately for each WML level, changes in driving difficulty lead to changes in neural correlates detectable by fNIRS. This means that the separate models were able to identify neural correlates specific to changes in driving difficulty for each WML level. Second, the chance level accuracy achieved while predicting driving difficulty in the combined model across different WML levels suggests that no neural correlates measurable with fNIRS changed with driving difficulty across different WML levels. This means, the changes in activation patterns due to changes in driving difficulty depended on the driver's current WML level. The interaction of the underlying brain processes is further supported by the additional comparisons of all possible combinations of predictions of driving difficulty separately across different WML levels. We showed that the construction condition could be better predicted when discriminated against the non-construction condition at the same WML level than when discriminated against a non-construction condition at the successive WML level. This suggests that an increase in WML recruits a neural network which reduces the discriminability of different levels of driving difficulty.

As fNIRS has good spatial resolution, it allowed us to determine brain areas predictive for visuospatial attention and to study a possible effect of WML on these brain areas. In order to identify potential brain areas associated with increased driving difficulty, we calculated group-level brain maps using univariate channel-wise logistic regression analysis to predict driving difficulty for each WML level. This analysis revealed the bilateral dorsal frontal (putative BA 46), bilateral inferior parietal (putative BA 39), and left superior parietal (putative BA 7) areas to be most sensitive to changes in driving difficulty. Nevertheless, these discriminative patterns diminished at higher WML levels indicating an interaction between visuospatial demands and WML levels.

The bilateral dorsal frontal areas (putative BA 46) are known to be involved in executive control of behavior (Kübler et al., 2006). In contrast, the bilateral inferior parietal (putative BA 39) and left superior parietal (putative BA 7) areas have been associated with visuomotor integration, spatial perception and orientation as well as in visual motion analysis (Andersen, 2011) and visuomotor attention (Jovicich et al., 2001; Caplan et al.,

<sup>1</sup><http://neurosynth.org>

<sup>2</sup><https://www.nitrc.org/projects/mricron>

2006) such as vigilance and tracking of moving objects (Culham et al., 1998). These areas play an important role in driving, especially with increased driving difficulty as in this study while driving through a construction site with reduced lane widths.

We proceeded to compare the brain areas predictive to driving difficulty to those areas predictive to WML independent of driving difficulty, previously shown by Unni et al. (2017) using the same data. The comparison of the anatomical locations of predictive maxima for WML predictions revealed only partial overlap between the brain resources predictive to the different task demands. Variations in WML levels was best predicted in bilateral inferior frontal gyrus (IFG, putative BA 45), which was further posterior to the lateral dorsal frontal areas (putative BA 46) predictive for driving difficulty. An interesting point to note was that the bilateral inferior and left superior parietal areas (putative BA 39 and BA 7, respectively), which showed increased predictivity to driving difficulty, showed negative correlations in the WML level predictions independent of driving difficulty. This could indicate that the two tasks interact at a common, task unspecific cognitive resource at the brain level. The changing pattern of driving difficulty related brain areas across WML levels could indicate potential changes in the multitasking strategy with level of WML demand.

The task interactions at brain level could be explained on the basis of the Multiple Resource Theory (Wickens, 2002) where an executive control system adjusts and allocates resources between the two tasks. The bilateral dorsal frontal areas could potentially represent the executive control system. From the predictivity patterns of the brain maps, we observed that these areas show increased predictivity to driving difficulty up to the 3-back WML level suggesting an increase in the difference in effort by the participants for driving difficulty. The increased cognitive resources allocated by the executive control to the WML task rather than for increased visuospatial attention may have reduced the predictivity pattern in the parietal areas representing visuomotor co-ordination. It has been shown in a driving simulator study that participants can strategically prioritize among subtasks and adapt effort and driving behavior accordingly (Cnossen et al., 2000).

In our study, prediction accuracies and *F1*-scores derived from fNIRS brain activation measurements decreased for 3- and 4-back WML levels. Participants might have reached their maximum capacity at 3-back or 4-back WML levels. According to multiple theories (Kahneman, 1973; Tombu and Jolicoeur, 2003; Wickens et al., 2015), once the maximum resource capacity is reached, limited resources are distributed across subtasks. This would suggest that there were only limited resources available for visuospatial attention needed for increased driving difficulty in the higher WML levels. This can explain the drop in task performance, the decrease in prediction accuracies and *F1*-scores, as well as the decreased predictivity of localized brain areas associated with increased driving difficulty for high WML levels.

The notion of a competition of cognitive resources available for the two tasks was further supported by the analysis of the behavioral data. Participants made more errors in the working memory task with increased driving difficulty and

had to make more steering adjustments (indicated by higher steering reversal rates) with increased WML levels. Hence, an increase in cognitive demands for one domain led to a decrease in performance associated with the other cognitive domain. These results are in line with Salvucci and Beltowska (2008) who observed that increasing working memory demand of a concurrent task substantially reduced driving performance with respect to lateral control and brake response. Further, this task interference became larger at high WML. Specifically, at high WML levels (3- and 4-back), increased driving difficulty led to a much larger drop in performance in the working memory task, as compared to low and intermediate WML levels (0-back to 2-back) at which the effect of increased driving difficulty on the working memory task performance was substantially smaller.

There are some limitations in this study that need to be addressed. First, our sample population was low. Second, the working memory task used is novel and other than traditional memory span task used in driving research, where digits are presented auditory (e.g., Mehler et al., 2012), the presentation of stimuli in this task was visually, at a lower frequency, which added an additional encoding and retention component to the task. Future studies using the same paradigm should also consider that a participant needs to pass *n*-number of speed signs to reach the corresponding *n*-back WML level and might therefore want to include more speed signs for higher *n*-back levels. Third, the construction condition is not well validated. For example, driving through a construction site is associated with increased workload (Shakouri et al., 2018), even if the lane width is not reduced (Vrieling et al., 2014). For example, the construction condition had different lane markings than the non-construction condition, which can influence driving behavior (Davidse et al., 2004; Charlton et al., 2018). Further, pylons marked the beginning of the construction sites in this experiment that could possibly affect the preferred driving speed in construction sites (Blackman et al., 2014; Steinbakk et al., 2017). In general, rich driving environments increase visual demands and uncertainty in the driver (Kujala et al., 2016), which might have made it more difficult for the driver to detect and encode speed signs in the construction condition necessary for the WML task. Thus, increased effort in scanning for speed signs in the construction condition could have altered lane-keeping. We also have to point out that participants received feedback for the working memory task only, possibly shifting the focus toward this task, whereas in real driving, lane keeping would have been prioritized over speed regulation. To assure participants had the effective WML level as intended, we have excluded time segments, in which participants didn't reach their target speed.

Our results could potentially have practical implications in the field of brain-based adaptive driver state assessment. Assessment of a driver's cognitive state has the goal to detect when the driver's workload is too high to keep up with the demands of operating a vehicle safely (Aghaei et al., 2016). In such situations, a driver assistance system could provide feedback to the driver (Feng and Donmez, 2013). For example, the use of a haptic steering wheel providing haptic feedback to the driver for ideal

steering movements helped to decrease driving difficulty (Steele and Gillespie, 2001). Alternatively, adaptive automation systems have the goal to detect the driver's current cognitive demands and to adjust the level of automation accordingly (Parasuraman et al., 2000). Our results illustrate the challenge to disentangle different types of workloads calling for new methods in workload assessment for an accurate assessment of cognitive demands in applied multiple task settings.

## CONCLUSION

Our study indicates brain level interactions between visuospatial attentional demands and WML while driving using fNIRS brain activation measurements. As an explanation for the dependency of those two different cognitive demands, we proposed that once maximum capacity is reached, the two tasks must compete for available resources. Further, there could be an interaction at a common, task unspecific cognitive resource at the brain level. The interaction of those different driving relevant tasks constitutes a challenge in brain-based driver state assessment for adaptive automation systems. Future studies should investigate how different subtasks in driving influence each other and how they could be assessed independently. This could eventually lead to more specific support for the driver in operating the car safely.

## ETHICS STATEMENT

The experiments of this study were carried out in accordance with the recommendations of the guidelines of the German Aerospace Center and approved by the ethics committee of the Carl von Ossietzky University, Oldenburg, Germany. All subjects gave

written informed consent in accordance with the Declaration of Helsinki.

## AUTHOR CONTRIBUTIONS

AU, KI, MJ, and JR planned the research. AU and KI done data collection. JS, AU, and JR carried out data analysis. JS, AU, KI, MJ, and JR prepared the manuscript.

## FUNDING

This work was supported by the funding initiative Niedersächsisches Vorab of the Volkswagen Foundation and the Ministry of Science and Culture of Lower Saxony as a part of the Interdisciplinary Research Centre on Critical Systems Engineering for Socio-Technical Systems and a DFG-grant RI1511/2-1 to JR.

## ACKNOWLEDGMENTS

We thank Andrew Koerner, Dirk Assmann, and Henrik Surm for their technical assistance and Christina Dömeland and Helena Schmidt for helping out with the data collection.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnhum.2018.00542/full#supplementary-material>

## REFERENCES

- Aghaei, A. S., Donmez, B., Liu, C. C., He, D., Liu, G., Plataniotis, K. N., et al. (2016). Smart driver monitoring: when signal processing meets human factors: in the driver's seat. *IEEE Signal Process. Mag.* 33, 35–48. doi: 10.1109/MSP.2016.2602379
- Ahn Son, L., Aoki, H., Murase, F., and Ishida, K. (2018). A novel method for classifying driver cognitive distraction under naturalistic conditions with information from near-infrared spectroscopy. *Front. Hum. Neurosci.* 12:431. doi: 10.3389/fnhum.2018.00431
- Andersen, R. A. (2011). "Inferior parietal lobule function in spatial perception and visuomotor integration," in *Comprehensive Physiology*, ed. R. Terjung. doi: 10.1002/cphy.cp010512
- Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., and Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *Neuroimage* 59, 36–47. doi: 10.1016/j.neuroimage.2011.06.023
- Baddeley, A. (2003). Working memory: looking back and looking forward. *Nat. Rev. Neurosci.* 4, 829–839. doi: 10.1038/nrn1201
- Balardin, J. B., Zimeo Morais, G. A., Furucho, R. A., Trambaiolli, L., Vanzella, P., Biazoli, Jr, et al. (2017). Imaging brain function with functional near-infrared spectroscopy in unconstrained environments. *Front. Hum. Neurosci.* 11:258. doi: 10.3389/fnhum.2017.00258
- Benedetto, S., Pedrotti, M., Bremond, R., and Baccino, T. (2013). Leftward attentional bias in a simulated driving task. *Transp. Res. Part F Traffic Psychol. Behav.* 20, 147–153. doi: 10.1016/j.trf.2013.07.006
- Blackman, R. A., Debnath, A. K., and Haworth, N. L. (2014). "Work zone items influencing driver speeds at roadworks: worker, driver and expert perspectives," in *Proceedings of the Australasian Road Safety Research, Policing and Education Conference (RSRPE 2014)*, Melbourne, VI, 12–14.
- Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., and Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev.* 44, 58–75. doi: 10.1016/j.neubiorev.2012.10.003
- Brigadoi, S., Ceccherini, L., Cutini, S., Scarpa, F., Scatturin, P., Selb, J., et al. (2014). Motion artifacts in functional near-infrared spectroscopy: a comparison of motion correction techniques applied to real cognitive data. *Neuroimage* 85, 181–191. doi: 10.1016/j.neuroimage.2013.04.082
- Brooks, J. R., Passaro, A. D., Kerick, S. E., Garcia, J. O., Franaszczuk, P. J., and Vettel, J. M. (2018). Overlapping brain network and alpha power changes suggest visuospatial attention effects on driving performance. *Behav. Neurosci.* 132, 23–33. doi: 10.1037/bne0000224
- Caplan, J. B., Luks, T. L., Simpson, G. V., Glaholt, M., and McIntosh, A. R. (2006). Parallel networks operating across attentional deployment and motion processing: a multi-seed partial least squares fMRI study. *Neuroimage* 29, 1192–1202. doi: 10.1016/j.neuroimage.2005.09.010
- Charlton, S. G., Starkey, N. J., and Malhotra, N. (2018). Using road markings as a continuous cue for speed choice. *Accid. Anal. Prev.* 117, 288–297. doi: 10.1016/j.aap.2018.04.029
- Choi, M.-H., Kim, H.-S., Yoon, H.-J., Lee, J.-C., Baek, J.-H., Choi, J.-S., et al. (2017). Increase in brain activation due to sub- tasks during driving: fMRI study using new MR-compatible driving simulator. *J. Physiol. Anthropol.* 36:11. doi: 10.1186/s40101-017-0128-8
- Cnossen, F., Rothengatter, T., and Meijman, T. (2000). Strategic changes in task performance in simulated car driving as an adaptive response to task demands.

- Transp. Res. Part F Traffic Psychol. Behav. 3, 123–140. doi: 10.1016/S1369-8478(00)00021-8
- Cooper, J. M., Medeiros-Ward, N., and Strayer, D. L. (2013). The impact of eye movements and cognitive workload on lateral position variability in driving. *Hum. Factors* 55, 1001–1014. doi: 10.1177/0018720813480177
- Cooper, R., Selb, J., Gagnon, L., Phillip, D., Schytz, H. W., Iversen, H. K., et al. (2012). A systematic comparison of motion artifact correction techniques for functional near-infrared spectroscopy. *Front. Neurosci.* 6:147. doi: 10.3389/fnins.2012.00147
- Cowan, N. (2010). The magical mystery four: how is working memory capacity limited, and why? *Curr. Dir. Psychol. Sci.* 19, 51–57. doi: 10.1177/0963721409359277
- Cui, X., Bray, S., Bryant, D. M., Glover, G. H., and Reiss, A. L. (2011). A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *Neuroimage* 54, 2808–2821. doi: 10.1016/j.neuroimage.2010.10.069
- Culham, J. C., Brandt, S. A., Cavanagh, P., Kanwisher, N. G., Dale, A. M., and Tootell, R. B. H. (1998). Cortical fMRI activation produced by attentive tracking of moving targets. *J. Neurophysiol.* 80, 2657–2670. doi: 10.1152/jn.1998.80.5.2657
- da Silva, F. P. (2014). Mental workload, task demand and driving performance: what relation? *Proc. Soc. Behav. Sci.* 162, 310–319. doi: 10.1016/j.sbspro.2014.12.212
- Davidse, R., van Driel, C., and Goldenbeld, C. (2004). *The Effect of Altered Road Markings on Speed and Lateral Position*. Leidschendam: SWOV Inst. Road Saf. Res.
- De Waard, D. (1996). *The Measurement of Drivers' Mental Workload*. Ph.D. thesis, Groningen University, Groningen.
- De Waard, D., Jessurun, M., Steyvers, F. J., Reggatt, P. T., and Brookhuis, K. A. (1995). Effect of road layout and road environment on driving performance, drivers' physiology and road appreciation. *Ergonomics* 38, 1395–1407. doi: 10.1080/00140139508925197
- Durant, G., Gagnon, J.-F., Tremblay, S., and Dehais, F. (2014). Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behav. Brain Res.* 259, 16–23. doi: 10.1016/j.bbr.2013.10.042
- Engström, J., Markkula, G., Victor, T., and Merat, N. (2017). Effects of cognitive load on driving performance: the cognitive control hypothesis. *Hum. Factors* 59, 734–764. doi: 10.1177/0018720817690639
- Essenpreis, M., Elwell, C. E., Cope, M., der Zee, P., Arridge, S. R., and Delpy, D. T. (1993). Spectral dependence of temporal point spread functions in human tissues. *Appl. Opt.* 32, 418–425. doi: 10.1364/AO.32.000418
- Feng, J., and Donmez, B. (2013). *Design of Effective Feedback: Understanding Driver, Feedback, and Their Interaction*. Available at: <https://ir.uiowa.edu/cgi/viewcontent.cgi?article=1519&context=drivingassessment>
- Fischer, M., Richter, A., Schindler, J., Plättner, J., Temme, G., Kelsch, J., et al. (2014). “Modular and scalable driving simulator hardware and software for the development of future driver assistance and automation systems,” in *Proceedings of the New Developments in Driving Simulation Design and Experiments: Driving Simulation Conference 2014*, Paris.
- Foy, H. J., Runham, P., and Chapman, P. (2016). Prefrontal cortex activation and young driver behaviour: a fNIRS study. *PLoS One* 11:e0156512. doi: 10.1371/journal.pone.0156512
- Gao, J., and Zhang, X. (2016). Driver lane keeping and vision behavior in fog based on dynamic clustering method. *Rev. Ibérica Sist. Tecnol. Informação* E12, 157–165.
- Guerrier, J. H., Manivannan, P., and Nair, S. N. (1999). The role of working memory, field dependence, visual search, and reaction time in the left turn performance of older female drivers. *Appl. Ergon.* 30, 109–119. doi: 10.1016/S0003-6870(98)00017-9
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). “Unsupervised learning,” in *The Elements of Statistical Learning*, eds T. Hastie, R. Tibshirani, and J. Friedman (New York, NY: Springer), 485–585. doi: 10.1007/978-0-387-84858-7\_14
- He, J., and McCarley, J. S. (2011). “Effects of cognitive distraction on lane-keeping: performance loss or improvement?” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, (Thousand Oaks, CA: SAGE Publications), 1894–1898. doi: 10.1177/1071181311551394
- He, J., McCarley, J. S., and Kramer, A. F. (2014). Lane keeping under cognitive load. *Hum. Factors* 56, 414–426. doi: 10.1177/0018720813485978
- Holdgraf, C. R., Rieger, J. W., Micheli, C., Martin, S., Knight, R. T., and Theunissen, F. E. (2017). Encoding and decoding models in cognitive electrophysiology. *Front. Syst. Neurosci.* 11:61. doi: 10.3389/fnsys.2017.00061
- Horrey, W. J., and Wickens, C. D. (2003). “Multiple resource modeling of task interference in vehicle control, Hazard awareness and in-vehicle task performance,” in *Proceedings of the 2nd International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design: Driving Assessment 2003* (Park City, UT: Public Policy Center of Iowa).
- Huppert, T. J., Franceschini, M. A., and Boas, D. A. (2009). *Noninvasive Imaging of Cerebral Activation with Diffuse Optical Tomography*, 2nd Edn. Boca Raton, FL: CRC Press. doi: 10.1201/9781420076851.ch14
- Huppert, T. J., Hoge, R. D., Diamond, S. G., Franceschini, M. A., and Boas, D. A. (2006). A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *Neuroimage* 29, 368–382. doi: 10.1016/j.neuroimage.2005.08.065
- Jolliffe, I. T. (1972). Discarding variables in a principal component analysis. I: artificial data. *Appl. Stat.* 21, 160–173. doi: 10.1177/0003702816652322
- Jovicich, J., Peters, R. J., Koch, C., Braun, J., Chang, L., and Ernst, T. (2001). Brain areas specific for attentional load in a motion-tracking task. *J. Cogn. Neurosci.* 13, 1048–1058. doi: 10.1162/089892901753294347
- Just, M. A., Keller, T. A., and Cynkar, J. (2008). A decrease in brain activation associated with driving when listening to someone speak. *Brain Res.* 1205, 70–80. doi: 10.1016/j.brainres.2007.12.075
- Kahneman, D. (1973). *Attention and Effort*. Cliffs, NJ: Prentice-Hall.
- Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika* 23, 187–200. doi: 10.1007/BF02289233
- Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *J. Exp. Psychol.* 55, 352–358. doi: 10.1037/h0043688
- Kübler, A., Dixon, V., and Garavan, H. (2006). Automaticity and reestablishment of executive control—An fMRI study. *J. Cogn. Neurosci.* 18, 1331–1342. doi: 10.1162/jocn.2006.18.8.1331
- Kujala, T., Mäkelä, J., Kotilainen, I., and Tokkonen, T. (2016). The attentional demand of automobile driving revisited: occlusion distance as a function of task-relevant event density in realistic driving scenarios. *Hum. Factors* 58, 163–180. doi: 10.1177/0018720815595901
- Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Curr. Dir. Psychol. Sci.* 19, 143–148. doi: 10.1177/0963721410370295
- Lewandowsky, S., Oberauer, K., Yang, L.-X., and Ecker, U. K. H. (2010). A working memory test battery for Matlab. *Behav. Res. Methods* 42, 571–585. doi: 10.3758/BRM.42.2.571
- Li, L.-P., Liu, Z.-G., Zhu, H.-Y., Zhu, L., and Huang, Y.-C. (2018). Functional near-infrared spectroscopy in the evaluation of urban rail transit drivers' mental workload under simulated driving conditions. *Ergonomics* 11, 1–21. doi: 10.1080/00140139.2018.1535093
- Liu, S., Wang, J., and Fu, T. (2016a). Effects of lane width, lane position and edge shoulder width on driving behavior in underground urban expressways: a driving simulator study. *Int. J. Environ. Res. Public Health* 13:E1010. doi: 10.3390/ijerph13101010
- Liu, T., Pelowski, M., Pang, C., Zhou, Y., and Cai, J. (2016b). Near-infrared spectroscopy as a tool for driving research. *Ergonomics* 59, 368–379. doi: 10.1080/00140139.2015.1076057
- Lust, J. M., Geuze, R. H., Groothuis, A. G. G., der Zwan, J. E., Brouwer, W. H., Van Wolfelaar, P. C., et al. (2011). Driving performance during word generation—testing the function of human brain lateralization using fTCD in an ecologically relevant context. *Neuropsychologia* 49, 2375–2383. doi: 10.1016/j.neuropsychologia.2011.04.011
- Macdonald, W. A., and Hoffmann, E. R. (1980). Review of relationships between steering wheel reversal rate and driving task demand. *Hum. Factors J. Hum. Factors Ergon. Soc.* 22, 733–739. doi: 10.1177/001872088002200609
- MacIntosh, B. J., Klassen, L. M., and Menon, R. S. (2003). Transient hemodynamics during a breath hold challenge in a two part functional imaging study with simultaneous near-infrared spectroscopy in adult humans. *Neuroimage* 20, 1246–1252. doi: 10.1016/S1053-8119(03)00417-8
- Masataka, N., Perlovsky, L., and Hiraki, K. (2015). Near-infrared spectroscopy (NIRS) in functional research of prefrontal cortex. *Front. Hum. Neurosci.* 9:274. doi: 10.3389/fnhum.2015.00274
- Matthews, G., Reinerman-Jones, L., Wohleber, R., Lin, J., Mercado, J., and Abich, J. (2015). “Workload is multidimensional, not unitary: what now?” in *Proceedings*



- of the *International Conference on Augmented Cognition*, Cham, 44–55. doi: 10.1007/978-3-319-20816-9\_5
- Medeiros-Ward, N., Cooper, J. M., and Strayer, D. L. (2014). Hierarchical control and driving. *J. Exp. Psychol. Gen.* 143, 953–958. doi: 10.1037/a0035097
- 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. *Hum. Factors* 54, 396–412. doi: 10.1177/0018720812442086
- Mehta, R. K., and Parasuraman, R. (2013). Neuroergonomics: a review of applications to physical and cognitive work. *Front. Hum. Neurosci.* 7:889. doi: 10.3389/fnhum.2013.00889
- Naseer, N., and Hong, K.-S. (2015). fNIRS-based brain-computer interfaces: a review. *Front. Hum. Neurosci.* 9:3. doi: 10.3389/fnhum.2015.00003
- Obrig, H., Neufang, M., Wenzel, R., Kohl, M., Steinbrink, J., Einhäupl, K., et al. (2000). Spontaneous low frequency oscillations of cerebral hemodynamics and metabolism in human adults. *Neuroimage* 12, 623–639. doi: 10.1006/nimg.2000.0657
- Oka, N., Yoshino, K., Yamamoto, K., Takahashi, H., Li, S., Sugimachi, T., et al. (2015). Greater activity in the frontal cortex on left curves: a vector-based fNIRS study of left and right curve driving. *PLoS One* 10:e0127594. doi: 10.1371/journal.pone.0127594
- Owsley, C., and McGwin, Jr (2010). Vision and driving. *Vis. Res.* 50, 2348–2361. doi: 10.1016/j.visres.2010.05.021
- Parasuraman, R., Sheridan, T. B., and Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* 30, 286–297. doi: 10.1109/3468.844354
- Patten, C. J. D., Kircher, A., Östlund, J., Nilsson, L., and Svenson, O. (2006). Driver experience and cognitive workload in different traffic environments. *Accid. Anal. Prev.* 38, 887–894. doi: 10.1016/j.aap.2006.02.014
- Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., et al. (2018). The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Ann. N. Y. Acad. Sci.* doi: 10.1111/nyas.13948 [Epub ahead of print].
- Prahl, S. (1999). *Optical Absorption of Hemoglobin*. Available at: <http://omlc.org/spectra/hemoglobin/index.html>
- Qian, J., Hastie, T., Friedman, J., Tibshirani, R., and Simon, N. (2013). *Glmnet for Matlab 2013*. Available at: [http://web.stanford.edu/~hastie/glmnet\\_matlab/](http://web.stanford.edu/~hastie/glmnet_matlab/)
- Reichert, C., Fendrich, R., Bernarding, J., Tempelmann, C., Hinrichs, H., and Rieger, J. W. (2014). Online tracking of the contents of conscious perception using real-time fMRI. *Front. Neurosci.* 8:116. doi: 10.3389/fnins.2014.00116
- Rieger, J. W., Reichert, C., Gegenfurtner, K. R., Noesselt, T., Braun, C., Heinze, H.-J., et al. (2008). Predicting the recognition of natural scenes from single trial MEG recordings of brain activity. *Neuroimage* 42, 1056–1068. doi: 10.1016/j.NEUROIMAGE.2008.06.014
- Rosey, F., and Auberlet, J.-M. (2012). Trajectory variability: Road geometry difficulty indicator. *Saf. Sci.* 50, 1818–1828. doi: 10.1016/j.ssci.2012.04.003
- Ross, V., Vossen, A. Y., Smulders, F. T. Y., Ruiter, R. A. C., Brijts, K., Brijts, K., et al. (2018). Measuring working memory load effects on electrophysiological markers of attention orienting during a simulated drive. *Ergonomics* 61, 429–443. doi: 10.1080/00140139.2017.1353708
- Salvucci, D. D., and Beltowska, J. (2008). Effects of memory rehearsal on driver performance: experiment and theoretical account. *Hum. Factors* 50, 834–844. doi: 10.1518/001872008X354200
- Sassaroli, A., and Fantini, S. (2004). Comment on the modified Beer-Lambert law for scattering media. *Phys. Med. Biol.* 49, N255–N257. doi: 10.1088/0031-9155/49/14/N07
- Schmitz, C. H., Klemer, D. P., Hardin, R., Katz, M. S., Pei, Y., Graber, H. L., et al. (2005). Design and implementation of dynamic near-infrared optical tomographic imaging instrumentation for simultaneous dual-breast measurements. *Appl. Opt.* 44:2140. doi: 10.1364/AO.44.002140
- Schneider, P., Piper, S., Schmitz, C., Schreier, N., Volkwein, N., Lüdemann, L., et al. (2011). Fast 3D near-infrared breast imaging using indocyanine green for detection and characterization of breast lesions. *Röfo* 183, 956–963. doi: 10.1055/s-0031-1281726
- Scholkman, F., Kleiser, S., Metz, A. J., Zimmermann, R., Mata Pavia, J., Wolf, U., et al. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *Neuroimage* 85, 6–27. doi: 10.1016/j.neuroimage.2013.05.004
- Schroeter, M. L., Kupka, T., Mildner, T., Uludağ, K., and von Cramon, D. Y. (2006). Investigating the post-stimulus undershoot of the BOLD signal—a simultaneous fMRI and fNIRS study. *Neuroimage* 30, 349–358. doi: 10.1016/j.neuroimage.2005.09.048
- Shakouri, M., Ikuma, L. H., Aghazadeh, F., and Nahmens, I. (2018). Analysis of the sensitivity of heart rate variability and subjective workload measures in a driving simulator: the case of highway work zones. *Int. J. Ind. Ergon.* 66, 136–145. doi: 10.1016/j.ergon.2018.02.015
- Shalev-Shwartz, S., and Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge: Cambridge university press. doi: 10.1017/CBO9781107298019
- Shimizu, T., Hirose, S., Obara, H., Yanagisawa, K., Tsunashima, H., Marumo, Y., et al. (2009). Measurement of frontal cortex brain activity attributable to the driving workload and increased attention. *SAE Int. J. Passeng. Cars Mech. Syst.* 2, 736–744. doi: 10.4271/2009-01-0545
- Sibi, S., Ayaz, H., Kuhns, D. P., Sirkin, D. M., and Ju, W. (2016). “Monitoring driver cognitive load using functional near infrared spectroscopy in partially autonomous cars,” in *Proceedings of the IEEE Intelligent Vehicles Symposium, Proceedings*, Gothenburg. doi: 10.1109/IVS.2016.7535420
- Steele, M., and Gillespie, R. B. (2001). Shared control between human and machine: using a haptic steering wheel to aid in land vehicle guidance. *Proc. Hum. Factors* 45, 1671–1675. doi: 10.1177/154193120104502323
- Steinbakk, R. T., Ulleberg, P., Sagberg, F., and Fostervold, K. I. (2017). Analysing the influence of visible roadwork activity on drivers' speed choice at work zones using a video-based experiment. *Transp. Res. Part F Traffic Psychol. Behav.* 44, 53–62. doi: 10.1016/j.trf.2016.10.003
- Suzuki, S. (2017). Optimized statistical parametric mapping procedure for NIRS data contaminated by motion artifacts?: neurometric analysis of body schema extension. *Brain Inform.* 4, 171–182. doi: 10.1007/s40708-017-0070-x
- Tjur, T. (2009). Coefficients of determination in logistic regression models—a new proposal: the coefficient of discrimination. *Am. Stat.* 63, 366–372. doi: 10.1198/tast.2009.08210
- Tombu, M., and Jolicoeur, P. (2003). A central capacity sharing model of dual-task performance. *J. Exp. Psychol. Percept. Perform.* 29, 3–18. doi: 10.1037/0096-1523.29.1.3
- Unni, A., Ihme, K., Jipp, M., and Rieger, J. W. (2017). Assessing the driver's current level of working memory load with high density functional near-infrared spectroscopy: a realistic driving simulator study. *Front. Hum. Neurosci.* 11:167. doi: 10.3389/FNHUM.2017.00167
- Victor, T. W., Harbluk, J. L., and Engström, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transp. Res. Part F Traffic Psychol. Behav.* 8, 167–190. doi: 10.1016/j.trf.2005.04.014
- Villringer, A., Planck, J., Hock, C., Schleinkofer, L., and Dirnagl, U. (1993). Near infrared spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults. *Neurosci. Lett.* 154, 101–104. doi: 10.1016/0304-3940(93)90181-J
- Vingerhoets, G., and Stroobant, N. (1999). Lateralization of cerebral blood flow velocity changes during cognitive tasks: a simultaneous bilateral transcranial doppler study. *Stroke* 30, 2152–2158. doi: 10.1161/01.STR.30.10.2152
- Virtanen, J., Noponen, T. E. J., and Meriläinen, P. (2009). Comparison of principal and independent component analysis in removing extracerebral interference from near-infrared spectroscopy signals. *J. Biomed. Opt.* 14:54032. doi: 10.1117/1.3253323
- Vossen, A. Y., Ross, V., Jongen, E. M. M., Ruiter, R. A. C., and Smulders, F. T. Y. (2016). Effect of working memory load on electrophysiological markers of visuospatial orienting in a spatial cueing task simulating a traffic situation. *Psychophysiology* 53, 237–251. doi: 10.1111/psyp.12572
- Vrieling, J., de Waard, D., and Brookhuis, K. (2014). Driving behaviour while driving through two types of road works. *Int. J. Traffic Transp. Eng.* 3, 141–148. doi: 10.5923/j.jitte.20140303.01
- Wang, Y.-K., Jung, T.-P., and Lin, C.-T. (2018). Theta and alpha oscillations in attentional interaction during distracted driving. *Front. Behav. Neurosci.* 12:3. doi: 10.3389/fnbeh.2018.00003
- Weichwald, S., Meyer, T., Özdenizci, O., Schölkopf, B., Ball, T., and Grosse-Wentrup, M. (2015). Causal interpretation rules for encoding and decoding models in neuroimaging. *Neuroimage* 110, 48–59. doi: 10.1016/j.neuroimage.2015.01.036

- Wickens, C. D. (2002). Multiple resources and performance prediction. *Theor. Issues Ergon. Sci.* 3, 159–177. doi: 10.1080/14639220210123806
- Wickens, C. D. (2008). Multiple resources and mental workload. *Hum. Factors* 50, 449–455. doi: 10.1518/001872008X288394
- Wickens, C. D., Hollands, J. G., Banbury, S., and Parasuraman, R. (2015). *Engineering psychology & Human Performance*. Hove: Psychology Press.
- Wickens, C. M., Toplak, M. E., and Wiesensthal, D. L. (2008). Cognitive failures as predictors of driving errors, lapses, and violations. *Accid. Anal. Prev.* 40, 1223–1233. doi: 10.1016/j.aap.2008.01.006
- Wood, G., Hartley, G., Furley, P. A., and Wilson, M. R. (2016). Working memory capacity, visual attention and hazard perception in driving. *J. Appl. Res. Mem. Cogn.* 5, 454–462. doi: 10.1016/j.aap.2011.10.006
- Xu, L., Wang, B., Xu, G., Wang, W., Liu, Z., and Li, Z. (2017). Functional connectivity analysis using fNIRS in healthy subjects during prolonged simulated driving. *Neurosci. Lett.* 640, 21–28. doi: 10.1016/j.neulet.2017.01.018
- Xu, Y., Graber, H. L., and Barbour, R. L. (2014). nirsLAB: a computing environment for fNIRS neuroimaging data analysis. *Biomed. Opt.* 2014:BM3A.1. doi: 10.1364/BIOE.2014.BM3A.1
- Yoshino, K., Oka, N., Yamamoto, K., Takahashi, H., and Kato, T. (2013a). Correlation of prefrontal cortical activation with changing vehicle speeds in actual driving: a vector-based functional near-infrared spectroscopy study. *Front. Hum. Neurosci.* 7:895. doi: 10.3389/fnhum.2013.00895
- Yoshino, K., Oka, N., Yamamoto, K., Takahashi, H., and Kato, T. (2013b). Functional brain imaging using near-infrared spectroscopy during actual driving on an expressway. *Front. Hum. Neurosci.* 7:882. doi: 10.3389/fnhum.2013.00882
- Zhang, Q., Strangman, G. E., and Ganis, G. (2009). Adaptive filtering to reduce global interference in non-invasive NIRS measures of brain activation: how well and when does it work? *Neuroimage* 45, 788–794. doi: 10.1016/j.neuroimage.2008.12.048
- Zhang, Y., Brooks, D. H., Franceschini, M. A., and Boas, D. A. (2005). Eigenvector-based spatial filtering for reduction of physiological interference in diffuse optical imaging. *J. Biomed. Opt.* 10:11014. doi: 10.1117/1.1852552

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Scheunemann, Unni, Ihme, Jipp and Rieger. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# A Review of Psychophysiological Measures to Assess Cognitive States in Real-World Driving

Monika Lohani<sup>1\*</sup>, Brennan R. Payne<sup>2</sup> and David L. Strayer<sup>2</sup>

<sup>1</sup> Department of Educational Psychology, University of Utah, Salt Lake City, UT, United States, <sup>2</sup> Department of Psychology, University of Utah, Salt Lake City, UT, United States

## OPEN ACCESS

### Edited by:

Bruce Mehler,  
Massachusetts Institute of  
Technology, United States

### Reviewed by:

Joost De Winter,  
Delft University of Technology,  
Netherlands  
Mickael Causse,  
National Higher School of Aeronautics  
and Space, France  
Dick De Waard,  
University of Groningen, Netherlands  
Edmund Wascher,  
Leibniz Research Centre for Working  
Environment and Human Factors  
(IfADO), Germany

### \*Correspondence:

Monika Lohani  
monika.lohani@utah.edu

**Received:** 01 May 2018

**Accepted:** 01 February 2019

**Published:** 19 March 2019

### Citation:

Lohani M, Payne BR and Strayer DL  
(2019) A Review of  
Psychophysiological Measures to  
Assess Cognitive States in Real-World  
Driving. *Front. Hum. Neurosci.* 13:57.  
doi: 10.3389/fnhum.2019.00057

As driving functions become increasingly automated, motorists run the risk of becoming cognitively removed from the driving process. Psychophysiological measures may provide added value not captured through behavioral or self-report measures alone. This paper provides a selective review of the psychophysiological measures that can be utilized to assess cognitive states in real-world driving environments. First, the importance of psychophysiological measures within the context of traffic safety is discussed. Next, the most commonly used physiology-based indices of cognitive states are considered as potential candidates relevant for driving research. These include: electroencephalography and event-related potentials, optical imaging, heart rate and heart rate variability, blood pressure, skin conductance, electromyography, thermal imaging, and pupillometry. For each of these measures, an overview is provided, followed by a discussion of the methods for measuring it in a driving context. Drawing from recent empirical driving and psychophysiology research, the relative strengths and limitations of each measure are discussed to highlight each measures' unique value. Challenges and recommendations for valid and reliable quantification from lab to (less predictable) real-world driving settings are considered. Finally, we discuss measures that may be better candidates for a near real-time assessment of motorists' cognitive states that can be utilized in applied settings outside the lab. This review synthesizes the literature on in-vehicle psychophysiological measures to advance the development of effective human-machine driving interfaces and driver support systems.

**Keywords:** psychophysiology, cognition, driving, traffic safety, real-world

## THE IMPORTANCE OF PSYCHOPHYSIOLOGICAL MEASURES IN TRAFFIC SAFETY

Suboptimal level of cognitive functioning (e.g., inattention, drowsiness) is a key cause of traffic accidents and poor driving performance. According to Traffic Safety Culture Index, 87.5% of drivers identify distracted driving to be a greater concern today than in past years and 87.9% perceive drowsiness as a threat to their safety (AAA Foundation for Traffic Safety, 2018). Traffic safety researchers are constantly working on methods to improve driving performance by assessing cognitive states, such as drivers' workload, inattention, and fatigue. One way to improve the assessment of covert cognitive states is to adopt a multi-method approach to measure changes in central and peripheral nervous system functioning in order to sense near-real time information about cognitive states of motorists. Such assessments of internal states can also promote the development

of Advanced Driver Assistance Systems (ADAS) that can predict and augment risky driving behavior.

## Why Adopt Psychophysiological Measures?

Cognitive states can be assessed using subjective, behavioral, and physiological measures (Mauss and Robinson, 2009; Strayer et al., 2015; Lohani et al., 2018). Subjective measures can be limiting if the assessment is disruptive to the real-time task (i.e., primary task intrusion, see O'Donnell and Eggemeier, 1986). More importantly, humans may not always be accurate in making judgements about their cognitive states (Schmidt et al., 2009). Motorists can be inaccurate in making judgments about their internal and cognitive states (such as their attention, workload, and drowsiness levels). For instance, motorists were inaccurate at self-assessments of vigilance (Schmidt et al., 2009); even though objective physiological indicators (e.g., heart rate, EEG, and ERPs) suggested poor vigilance levels at the end of a 3-h drive, participants self-reported improved vigilance instead (Schmidt et al., 2009). Such misjudgments in assessment of cognitive states suggest that objective measures are required to assess and augment human behavior in order to reduce risk for traffic safety. While behavioral measures (such as head movement detection to assess distraction) are also useful, given the intent of this review, we will focus on physiological measures. Accuracy in detecting cognitive workload has been found to significantly increase when physiological data was utilized (Lenneman and Backs, 2009, 2010; Solovey et al., 2014; Borghini et al., 2015; Yang et al., 2016). Some work has also found that physiological measures were sensitive to variations in cognitive load during secondary tasks while behavioral driving measures like steering wheel reversals and velocity (Belyusar et al., 2015) and lane-keeping measures (Lenneman and Backs, 2009) were not. Unlike behavioral measures (e.g., verbal and facial behavior), many physiological measures are not under voluntary control of motorists. Moreover, cognitive states such as mental workload are a multi-faceted and dynamic concept and self-report alone cannot be used to operationalize it, but multiple measures (e.g., performance and physiology) are warranted (de Waard and Lewis-Evans, 2014). Thus, inclusion of physiological data can complement and extend behavioral metrics and improve assessments of motorists' state-level changes in cognition (Brookhuis and de Waard, 1993; Mehler et al., 2012).

As automation is likely to become more prevalent over time, real-time monitoring behaviors required by motorists may decline as they are less involved in the driving process. This is a critical reason why non-behavior-based metrics will become more relevant to incorporate into our understanding of the motorists' cognitive states. Moreover, distracted motorists of a self-driving vehicle compared to manually driving motorists take longer to gain control of the driving task once automation deactivates (Vogelpohl et al., 2018). Intelligent driving assistance systems should be capable of reliably sensing and assessing distraction and drowsiness levels of motorists to be able to augment safe-driving conditions. Building reliable systems to be able to predict decreased levels of vigilance or dangerous levels of

fatigue, drowsiness, or workload could help augment them in a timely manner (Balters et al., 2018).

## Cognition in Dynamic Real-World Driving Contexts

In general, psychophysiological measures can be used to assess degree of arousal or activation (Mauss and Robinson, 2009). Importantly, multiple psychological constructs can influence variations in psychophysiological measures. For instance, heart rate, skin conductance, and electrical activity of the brain are sensitive to many psychological constructs experienced by motorists, such as workload, drowsiness, stress, etc. In the past years, important contributions have reviewed the literature on specific cognitive states, such as workload (Borghini et al., 2014; Costa et al., 2017), distraction (Matthews et al., 2019), drowsiness (Sahayadhas et al., 2012; Borghini et al., 2014), and stress (Rastgoo et al., 2018) in driving research. These reviews provide an understanding of physiological outcomes that can explain variations in specific constructs based on carefully manipulated and well-controlled designs. Unlike highly controlled lab-based settings, where a single construct (e.g., workload) can be successfully manipulated and its effect on psychophysiological measures examined, real-world settings are more dynamic and complex.

In a real-world setting, the net resulting cognitive state of a motorist is a combination of variation among several interrelated constructs (e.g., attention allocation, stress, workload, fatigue). Broadly speaking, the net cognitive state of a motorist, composed of variation among these many dimensions, can be classified along an arousal-spectrum ranging from lower-arousal and passive states, to a state of optimal performance, to a hyper-aroused or over-active state. Indeed, this concept is not new; Yerkes and Dodson (1908) established strong non-linear relationships between arousal-level and performance, and such relationships have since been well-established across many human performance domains (Hebb, 1955; Broadhurst, 1959; Wekselblatt and Niell, 2015). Although these ideas are not new, there has been a recent resurgence in a formal understanding of arousal-performance relationships, including an expanded understanding of the underlying neuromodulatory systems involved in regulating task engagement and optimal performance (e.g., the adaptive-gain control theory, Aston-Jones and Cohen, 2005). Given the recent increase in understanding of the mapping between physiological indices of arousal and human performance in the lab, such models serve as a clear starting point in delineating the predictive capacity of psychophysiological measures for understanding cognitive states and human performance in the vehicle.

For instance, low-arousal states relevant to the driving task can be driven by a combination of psychological constructs including low workload, reduced stress, and high drowsiness. On the other hand, an over-aroused state could be due to a combination of high workload and high stress in the presence of low drowsiness. Similarly, other combinations of constructs can also lead to changes in general arousal states as well. Given the likely dynamic interplay among these interrelated constructs in applied settings,



the current review focuses on psychophysiological measures that can be utilized to capture motorists' states in real-world driving settings. Indeed, one major applied goal of this work is to be able to accurately capture the dynamic and highly variable changes in arousal that occur in ecologically valid driving settings, a goal that is critical for building accurate predictive models (Yarkoni and Westfall, 2017) of individual motorist's states and future driving performance.

Specifically, there are two novel contributions of this review. First, instead of focusing on a selective construct and related measures of interest, the goal of this current review is to focus on psychophysiological measures that may have the potential to be adopted in real-world and applied settings to measure state level variations in motorists. The paper provides a broad but selective review of a number of psychophysiological measures that we believe show the greatest promise in their utilization to assess low-arousal vs. over-arousal (passive vs. over-active) states in real-world driving environments. The most commonly used physiology-based measures of cognitive states are considered as potential candidates relevant for driving research. The following physiological measures are reviewed (see section "Psychophysiological Measures to Assess Cognitive States" and **Tables 1, 2**) in assessing arousal state in real-world driving research: electroencephalography and event-related potentials, optical imaging, heart rate, and heart rate variability, blood pressure, skin conductance, electromyography, thermal imaging, and pupillometry. As reviewed in classical contributions by Cacioppo et al. (Cacioppo and Tassinari, 1990; Cacioppo et al., 2007), inference of unique psychological constructs based on physiological indices (one-to-one relation) is still unresolved and is not the aim of this review (see further discussion in section "Research Applicability in Real-World Settings"). However, we discuss how multiple measures (that are sensitive to several interrelated internal states) may be combined to delineate net resulting changes across multiple inter-related cognitive state-level variations. Second, for each measure, we make the distinction between useful research measures and practical measures for real-world application (see section Research Applicability in Real-World Settings and **Table 2**). Throughout, we have tried to highlight the practical relevance of measures in the driving context. Although this review focuses primarily on on-road and simulated driving contexts, when relevant, we have also drawn research from related contexts (traffic operators, pilots, or ship navigators) to more thoroughly characterize each measure.

## PSYCHOPHYSIOLOGICAL MEASURES TO ASSESS COGNITIVE STATES

### Electroencephalogram (EEG) and Event-Related Potentials (ERP)

#### EEG Quantification

The EEG is a record of both oscillatory and aperiodic brain electrical activity. Neural activity (largely post-synaptic potentials) from multiple simultaneous generators propagate throughout the brain and skull and summate at a distance,

where voltages can be measured relatively non-invasively via electrodes placed on the scalp. The dominant sources of scalp-recorded EEG come from cortical pyramidal cells arranged in the columnar organization of the cortex (Nunez and Srinivasan, 2006). Pyramidal cells are the most numerous cortical excitatory cell type and play a critical role in advanced cognitive functions (Spruston, 2008). The laminar organization of the cortex results in cortical pyramidal cells following an *open-field* alignment with a consistent orientation that is perpendicular to the skull, such that their post-synaptic potentials can summate at a distance. Importantly, EEG allows for a high temporal resolution (millisecond) and direct record of neural activity. This detailed temporal resolution also allows for a decomposition of the time-domain EEG signal into spectral information via Fourier analysis, allowing for an examination of oscillatory activity in canonical frequency bands (e.g., alpha, ~8–12 Hz; theta, ~4–7 Hz), which have been related to specific neurocognitive functions. For instance, mental workload increases theta power and reduce alpha power activity (Mun et al., 2017), whereas fatigue increases alpha power (Käthner et al., 2014). Moreover, the development of novel computational techniques for analyzing spectral activity has promoted a wide range of new tools for probing ongoing neural dynamics during human cognition via EEG; such as cross-frequency coupling, phase coupling (Cohen, 2011), independent component analysis (Dasari et al., 2017), and neighborhood component analysis (Lim et al., 2018). In addition, more traditional analyses of transient neural activity that is tied to specific perceptual, motor, or cognitive events can be gleaned from continuous EEG, via the calculation of event-related brain potentials.

#### ERP Quantification

ERPs are electrophysiological responses that are consistently linked in time with specific sensory, cognitive, or motor events. They are derived from the continuously recorded EEG by time-aligning epochs of EEG relative to an event of interest, such as a stimulus onset or a participant's response and averaging many of these similar EEG segments to reveal activity that is time and phase locked to the event. Such discrete events can be added in the experimental design, e.g., every time a participant responded to a secondary task while driving. The logic of this approach is that systematic activity that is locked in time and space to some specific activity will remain in the averaged ERP waveform, whereas activity that is not time- and phase-locked will average to zero with a large enough number of trials (Luck and Kappenman, 2012). The resulting ERP waveform is plotted as voltage over time at a given set of electrodes. ERP topography can also be examined, showing the distribution of activity over the entire space within a particular time-window. A major benefit of ERPs is that the waveform has characteristic *components*, stereotyped features of the ERP with specific eliciting conditions. ERP components are defined empirically by a combination of their polarity, timing, scalp distribution, and sensitivity to task manipulations.

Extensive work has characterized and validated specific ERP components with respect to their associations with specific cognitive and neural processes (e.g., Fabiani et al., 2007; Luck and Kappenman, 2012; Mun et al., 2017). Cognitive

**TABLE 1** | Overview of relationships between arousal state and physiological indices in real-world driving.

| Measure  | Under-arousal state  | Over-arousal state   |
|--|--|--|
| Electroencephalogram                               | <ul style="list-style-type: none"> <li>Increased alpha due to increase in drowsiness and attentional withdrawal<sup>b</sup></li> <li>Changes in theta and delta activity related to transition to fatigue</li> </ul>                       | <ul style="list-style-type: none"> <li>Increase in theta activity due to mental workload</li> <li>Alpha activity suppression due to workload</li> </ul>  |
| Event related potential                            | <ul style="list-style-type: none"> <li>Reduced ERP amplitudes with fatigue, time on task, and lower vigilance over time while driving</li> </ul>   | <ul style="list-style-type: none"> <li>Also, reduced ERP amplitude to driving relevant stimuli under high workload</li> </ul>  |
| Optical imaging for cerebral flow                  | <ul style="list-style-type: none"> <li>A decrease in cerebral oxygenation with fatigue and drowsiness</li> </ul>   | <ul style="list-style-type: none"> <li>An increased concentration of oxygenated hemoglobin and a decreased concentration of deoxygenation with mental workload and stress</li> </ul>                                 |
| Heart rate and Heart rate variability <sup>b</sup> | <ul style="list-style-type: none"> <li>Decrease in heart rate with drowsiness, decrease in vigilance, use of self-driving technology</li> <li>Increase in HRV indices (e.g., RMSSD) with drowsiness, fatigue, and disengagement</li> </ul> | <ul style="list-style-type: none"> <li>Increase in heart rate with mental workload and stress</li> <li>Decrease in HRV with workload, stress, and vigilance</li> </ul>   |
| Blood Pressure <sup>b</sup>                        | <ul style="list-style-type: none"> <li>Decrease in blood pressure relative to baseline with fatigue and drowsiness</li> </ul>  | <ul style="list-style-type: none"> <li>Increase in systolic BP with workload and stress</li> </ul>   |
| Electrodermal activity                             | <ul style="list-style-type: none"> <li>Lower EDA relative to baseline activity range with drowsiness</li> </ul>  | <ul style="list-style-type: none"> <li>EDA increase with workload, stress, lower trust in automation, and anxiety</li> </ul>   |
| Electromyography                                   | <ul style="list-style-type: none"> <li>Decreases in mean and median power frequency of EMG due to decline in muscle activities and fatigue<sup>a</sup></li> </ul>  | <ul style="list-style-type: none"> <li>High muscle activity relative to baseline with stress</li> </ul>  |
| Thermal imaging                                    | <ul style="list-style-type: none"> <li>Temperature around baseline levels<sup>a</sup></li> </ul>   | <ul style="list-style-type: none"> <li>Higher task difficulty increases forehead temperature and decreases nose temperature and thus an increase in the difference between forehead and nose temperatures</li> </ul> |
| Pupillometry                                       | <ul style="list-style-type: none"> <li>Decreases in average pupil diameter with drowsiness</li> <li>Increases in standard deviation of pupil diameter</li> </ul>   | <ul style="list-style-type: none"> <li>Increases in pupil diameter with cognitive load</li> <li>Decreases in standard deviation of pupil diameter</li> </ul>   |

<sup>a</sup>Limited findings available.<sup>b</sup>Mixed findings reported.

demands can modulate several ERP components, such as P3 (discussed below; Käthner et al., 2014), mismatch negativity (MMN is a negative ERP component sensitive to pre-attentive information processing; Wanyan et al., 2018), and late positive potentials amplitude (a later ERP component like P6 that is related to attentional allocation similar to P3; Mun et al., 2017). The P3 component is associated with attentional and memory processes required to detect any changes in incoming stimuli-related information (Polich, 2007). The canonical P3 has two distinct but related components – the P3a and P3b (see Polich, 2007 for a review). The P3a, with an anterior distribution, is associated with novel stimulus-driven attentional processing or orienting responses. The P3b, with a centro-posterior distribution, is associated with task-relevant stimulus-driven attentional, decision making, and subsequent memory processing (Polich, 2007). Both components have been used in driving research. Recent work has also examined how neural indices (as measured by both P3 ERP components) are associated with subjective workload (as measured by NASA-TLX) and how this covariation is influenced by cognitive effort (Yakobi, 2018). Novel techniques (such as intra-block averaging of ERP amplitudes; Horat et al., 2016) can enable robust electrophysiological measurement of cognitive demands over time. Thus, ERPs are an attractive measure for studying cognitive states and performance in driving contexts.

### EEG/ERP in Driving Context

EEG and ERPs have a long history in the study of the neural indices of cognitive effort and attention allocation in both laboratory and applied settings. EEG is perhaps one of the

most widely used neurophysiological methods to study driving behavior. Several frequencies (e.g., power in alpha frequency band) and time (e.g., P3) domain indices can reliably measure changes in cognitive demands (Käthner et al., 2014). This makes EEG is viable measure for applied driving settings.

### Over-arousal in driving context

Over-aroused states, such as increased workload while driving can be indexed by decreases in alpha power and increases in theta power (Borghini et al., 2014; Käthner et al., 2014). A recent study found alpha band power to be higher during the relaxed condition compared with the engaged condition in an autonomous driving setting (Zander et al., 2017). This highlights the sensitivity of alpha power to internal factors such as attentional engagement. In addition to internal factors, external factors (such as task load and time on task) can also influence alpha and theta power bands in opposite directions (Wascher et al., 2018). For instance, a decrease in task load and time on task led to an increase in relative alpha power, but a decrease in theta power (Getzmann et al., 2018; Wascher et al., 2018). To account for both power bands, past work has also used a ratio of frontal theta and parietal alpha power spectral density to operationalize workload in pilots (Borghini et al., 2015). This ratio approach may be relevant for driving research as well, however this has been a point of debate, as discussed shortly.

The application of known ERP indicators of attentional workload (and their eliciting tasks) can be successfully translated into the driving domain as well. One of the most commonly adopted components in driving research is the P3b (Brookhuis and de Waard, 2010; Solís-Marcos and Kircher, 2018). Mental

**TABLE 2 |** Tentative framework for considering the research applicability of these measures in lab and real-world settings.

| Measure                           | Lab-setting | Real-world                 | Advantages   | Disadvantages  |
|-----------------------------------|-------------|----------------------------|--|--|
| Electroencephalogram              | High        | High to medium             | <ul style="list-style-type: none"> <li>• High temporal resolution</li> <li>• Direct measure of neural activity</li> </ul>  | <ul style="list-style-type: none"> <li>• Contact sensors</li> <li>• Longer setup time</li> </ul>   |
| Event related potential           | High        | Medium <sup>a</sup>        | <ul style="list-style-type: none"> <li>• Same benefits as EEG</li> <li>• Well-characterized components (e.g., P3)</li> <li>• High temporal resolution</li> </ul> | <ul style="list-style-type: none"> <li>• Same disadvantages as EEG</li> <li>• Generally needs higher number of trials and post-processing</li> <li>• Needs time-locking event</li> <li>• Interpretation of amplitude is very context specific</li> </ul> |
| Optical Imaging for Cerebral Flow | High        | Low to medium <sup>a</sup> | <ul style="list-style-type: none"> <li>• Higher spatial resolution</li> <li>• Feasible in naturalistic settings with technical advancements</li> </ul>           | <ul style="list-style-type: none"> <li>• Lower temporal resolution (e.g., fNIRS)</li> <li>• Need systematic investigation/replication</li> </ul>   |
| Heart Rate/Heart Rate Variability | High        | High                       | <ul style="list-style-type: none"> <li>• Higher signal-to-noise ratio (SNR)</li> <li>• Easy to collect</li> </ul>  | <ul style="list-style-type: none"> <li>• Very sensitive to artifacts</li> <li>• Sensitive to variation in respiration</li> </ul>   |
| Blood Pressure                    | High        | Medium <sup>a</sup>        | <ul style="list-style-type: none"> <li>• Reliable</li> <li>• Higher SNR</li> </ul>   | <ul style="list-style-type: none"> <li>• Limitations of equipment; can disrupt driving task</li> </ul>   |
| Electrodermal activity            | High        | High                       | <ul style="list-style-type: none"> <li>• Sympathetic activity</li> <li>• Easy to collect</li> </ul>  | <ul style="list-style-type: none"> <li>• Lagged response</li> <li>• Not all individuals show EDA response</li> </ul>   |
| Electromyography                  | High        | Low <sup>a</sup>           | <ul style="list-style-type: none"> <li>• Reliable</li> <li>• High temporal resolution</li> </ul>   | <ul style="list-style-type: none"> <li>• Slightly longer setup time</li> <li>• Sensitive to movement</li> <li>• Lower SNR</li> </ul>   |
| Thermal Imaging                   | High        | Medium <sup>a</sup>        | <ul style="list-style-type: none"> <li>• Low setup time</li> <li>• Promising technology</li> <li>• Non-contact</li> </ul>  | <ul style="list-style-type: none"> <li>• Need systematic investigation/replication</li> </ul>  |
| Pupillometry                      | High        | Low <sup>a</sup>           | <ul style="list-style-type: none"> <li>• Non-contact</li> <li>• Quick setup time</li> </ul>  | <ul style="list-style-type: none"> <li>• Signal strongly sensitive to variable lighting conditions (pupillary light reflex)</li> </ul>   |

<sup>a</sup>Limited findings available.

workload can be indexed by increases in P3b latencies (Ying et al., 2011) and amplitude (Strayer and Drews, 2007). For example, Strayer and Drews (2007) examined the amplitude of the P3b time-locked to the onset of a pace break light under single-task driving conditions or dual-tasking via cell-phone-induced distraction. Drawing on basic experimental work that has shown that the P3b is sensitive to the degree of attention allocated to a task (e.g., Sirevaag et al., 1989), they also showed that cell-phone induced distraction resulted in reduced P3b amplitudes to brake lights. Similar effects have been observed in comparing the workload of “single-task” driving in laboratory simulator vs. real-life driving contexts, where for example, the diversion of attention to other concurrent activities in the vehicle result in additional attentional demands in real-world driving (Strayer et al., 2015).

A recent study compared mental workload due to increased information processing demands consumed by in-vehicle information systems (Solís-Marcos and Kircher, 2018). They found both P3b and N1 latencies and amplitudes to be sensitive to cognitive demands of processing additional in-vehicle information systems. For instance, P3b amplitudes decreased with additional information processing related tasks (Solís-Marcos and Kircher, 2018). P3a amplitude was also found to decrease with additional task-related load (Getzmann et al., 2018). High mental workload has been associated with increased latencies in MMN during driving (Ying et al., 2011) and also increased frontal MMN in flight simulation tasks (Wanyan et al., 2018), however a recent study did not find workload to influence

MMN amplitudes (Getzmann et al., 2018). Future work will help clarify sensitivity of MMN in driving research.

### *Under-arousal in driving context*

Extensive work has focused on electrophysiological indicators of under-arousal via EEG. A substantial number of papers have implicated changes in alpha amplitude during fatigued driving (e.g., Schier, 2000; Jensen and Mazaheri, 2010; Simon et al., 2011; Zhao et al., 2012; Borghini et al., 2014; Jagannath and Balasubramanian, 2014; Arnau et al., 2017; Brouwer et al., 2017), such that fatigued driving is associated with increased alpha activity. However, other work has challenged these alpha power links with fatigue and claim that alpha power changes may be due to the decreases in task-demands and visual input during monotonous driving tasks and not due to decline in cognitive processing abilities (Wascher et al., 2014). Increases in relative alpha band power with increased time on task, easier driving route, and lower control of driving situations, which suggested that relative alpha power increases imply attentional withdrawal and not fatigue (Wascher et al., 2014, 2018). Wascher et al. (2014, 2018) have argued that mid-frontal theta activity may be a more appropriate neural marker of cognitive-control related processes in driving than occipital alpha activity. Low task load is associated with relatively reduced theta activity, which suggests that theta activity is sensitive to declines in cognitive processing ability. Instead of alpha activity, Wascher et al. (2014, 2018) recommend that indices of oscillatory synchronization (e.g., inter-trial phase clustering) and ERPs (such as P3a) are more reliable and valid

indices of changes in cognitive state associated with mental fatigue. For instance, time on task (Wascher et al., 2014), fatigue (Massar et al., 2010), and decreases in vigilance over time (Schmidt et al., 2009) were found to reduce P3a amplitude while driving. Similarly, mind-wandering during driving is associated with a reduction in P3a amplitude (Baldwin et al., 2017). One other study found both P3a and P3b components' amplitudes were reduced due to driving-related fatigue (Guoping and Zhang, 2009). These findings show that ERP components could be utilized to detect variations in neurophysiological arousal due to interrelated cognitive constructs in driving contexts.

Some researchers have argued that LF/HF ratios (e.g., frontal theta/beta) are potential biomarkers for attentional control, and have established some evidence that such measures have good psychometric properties, for e.g., test-retest reliability (Putman et al., 2014; Angelidis et al., 2016). Decreases in beta power (e.g., Zhao et al., 2012; Jagannath and Balasubramanian, 2014) have been found, along with changes in theta and delta activity as markers related to transition to fatigue. This has led some researchers to propose spectral ratio indices (e.g., alpha/beta; Eoh et al., 2005; Wang et al., 2018), as biomarkers of alertness. However, ratio indices have also been criticized for being an inadequate method because it combines frequency bands with distinct topographic specificity that change differently over time (Wascher et al., 2014). There is existing criticism of this ratio approach, especially in driving research (Wascher et al., 2018), and more broadly, researchers in cognitive electrophysiology have been moving away from such highly constrained "band-based" approaches given their lack of replicability across studies. Alternatively, researchers have increasingly endorsed methods that allow for broad-band assessment of spectral dynamics (e.g., 1/f scaling, Voytek and Knight, 2015) and methods that can address narrow-band dynamics without a priori selection of frequency (e.g., cluster-based permutation testing in time-frequency data; Maris and Oostenveld, 2007). Other recent work has used EEG-based detection algorithms to detect fatigue and drowsiness (Li et al., 2017; Morales et al., 2017; Belakhdar et al., 2018; Gao et al., 2018; Wei et al., 2018). However, other work reported no additional benefit of utilizing EEG measures in drowsiness and fatigue detection in sleep deprivation contexts (Perrier et al., 2016; Liang et al., 2017). Another line of work has aimed to apply machine-learning techniques to brain computing interfaces in order to classify states of drowsiness and fatigue in real-time (e.g., Lin et al., 2005; Correa et al., 2014). Recent work has also shown data filtering and processing techniques such as artifact subspace reconstruction and independent component analysis could be utilized for "online" processing of EEG data collected while driving in order to attenuate movement- and noise-related artifacts (Krol et al., 2017). Together, these findings suggest that EEG and ERPs can be utilized as objective techniques to assess state-level variations in cognitive demands.

## Practical Considerations

There are a number of important considerations when applying EEG indices to real-world driving environments. Typical EEG

artifacts arising from muscle- and eye movements (de Waard, 1996; Zander et al., 2017), impedance shifts, environmental line (60 Hz) noise, and other complications are potentially amplified in real-world environments. As such, real-time monitoring of good quality EEG signals is critical for effective data collection. The commercial introduction of high-impedance systems with active electrodes and small electrically shielded mobile EEG amplifiers has spawned a large increase in real-world EEG applications. Many of these systems are capable of high density (<128 channel) recording, but it is critical for the researcher to decide whether and to what degree an increase in the number of channels may result in a decrease in the quality of the recorded EEG (Luck and Kappenman, 2012). Importantly, the well-understood limitations of the spatial resolution of EEG limit the utility of high-density recording in ecologically valid environments (e.g., where measurement of EEG sensors co-localized in 3D space on a single-subject basis may be unfeasible). Moreover, with increasing channel density comes increases in the likelihood for poorly recorded or poorly monitored channels during recording. As such, if source-localization of underlying EEG/ERP generators is not a primary aim of the methodology (and we expect, in most applied cases it would not be), researchers may wish to record from a smaller density (e.g., 32 channels or fewer), at the benefit of better monitoring of data quality throughout the experiment.

On the theoretical side—researchers in human factors automotive research should carefully consider the linking hypotheses between specific electrophysiological indicators (e.g., P3b ERP amplitude, alpha power increases) and their purported cognitive interpretations. The ERP literature has a massive basic literature in which specific components have been very well-characterized relative to their eliciting conditions and underlying cognitive interpretations (Luck and Kappenman, 2012). One such example was reviewed earlier on characterizing the P3b under different states of distraction during driving. Limited work (e.g., Strayer et al., 2015) has attempted to examine ERP components in naturalistic settings. In future work, inventive approaches can be validated to use task-related responses or behaviors (such as eye-blink potentials or frequent vs. infrequent vehicle cues) as discrete events that can be recorded to estimate ERP components in real-world settings. At the same time, such characterizations in the spectral domain are not as clearly developed to date. However, this is changing, as basic research in cognitive electrophysiology shifts toward a more complete understanding of oscillatory mechanisms underlying human perception and cognition (e.g., Kahana, 2006), involving development in standardized analysis methods (Cohen, 2011), careful experimental characterization of specific oscillatory markers (e.g., alpha phase and perception, Mathewson et al., 2009; midline frontal theta and conflict resolution; Cavanagh and Frank, 2014), and the development of neurophysiologically guided models (Jensen and Mazaheri, 2010; Voytek and Knight, 2015). We expect that such development of basic research findings in cognitive electrophysiology will be a great asset in future applied research in contexts such as driving.



## Optical Imaging for Cerebral Blood Flow Optical Imaging Quantification

*Optical imaging* methods allow for the visualization of the interaction of photons with tissues (Villringer et al., 1993). In recent years, there has been a rapid advancement in the application of non-invasive optical imaging methods such as functional near infrared spectroscopy (fNIRS) to study human brain and cognitive functioning. fNIRS is a neuroimaging method based on the principles of near-infrared spectroscopy, which was originally developed in humans for investigating clinical features of brain functioning (e.g., cerebral oxygenation; Jobsis, 1977). These principles have been extended to measure local changes in cerebral hemodynamic activity that can be used to infer information on the underlying neural activity due to neurovascular coupling, following similar logic to the Blood Oxygen Level Dependent (BOLD) signal in functional magnetic resonance imaging. NIR (700–1,000 nm) light is able to penetrate several centimeters through the skull and into brain tissue, allowing for non-invasive measurement of certain optical properties of cortical tissue. For example, changes in the concentration of oxy- and deoxy-hemoglobin can be measured via NIRS because oxy- and deoxy-hemoglobin have distinct absorption spectra that correspond to the different coloration of arterial and venous blood (Grinvald et al., 1986). These absorption characteristics make it possible to use a spectroscopic approach to measure changes in the concentration of oxy- and deoxy- hemoglobin as a function of neural activity, for example during cognitive task performance. In typical optical imaging systems, optical fibers, called optodes or sources, carry NIR light to the scalp while other optical fibers, called detectors, collect the photons as they emerge from the scalp. Each source–detector pair is a single channel. Multi-channel and wearable fNIRS systems have become commercially available with diverse montages capable of measuring brain activity across the entire scalp.

## Optical Imaging for Cerebral Blood Flow in Driving Context

The application of fNIRS in driving research is in its infancy. Nevertheless, a number of interesting demonstrations of the utility of fNIRS for studying over-arousal states such as driver workload have emerged (e.g., Tsunashima and Yanagisawa, 2009; Liu et al., 2012, 2016; Sibi et al., 2016). For example, increases in oxygenated hemoglobin have been reported during simulated driving tasks under cognitive load compared to control conditions (Liu et al., 2012). A recent study (Unni et al., 2017) utilized fNIRS in a naturalistic driving simulator while doing a secondary task (modified version of 0–4 back). They found systematic increases in bilateral inferior frontal and temporo-occipital brain regions with increments in workload. Another study reported that fNIRS could be used to differentiate between low vs. high workload (n-back task) related hemodynamic activity in the prefrontal cortex while motorists drove in a realistic driving simulator (Herff et al., 2017). Furthermore, fNIRS have been used to monitor pilot's task engagement and working memory load in real-time (Gateau et al., 2015). On a related note, fNIRS have been found sensitive to increase in task difficulty in flight simulators (Causse et al., 2017) as indicated

by an increased concentration of oxygenated hemoglobin and a decreased deoxygenated hemoglobin.

Other work has investigated effects of under-arousal related states with fNIRS. Research has related decreases in hemodynamic measures of cerebral oxygenation with fatigue in simulated driving (Li et al., 2009), and findings have been extended into actual highway driving (Yoshino et al., 2013). An increase in fatigue can be indexed by a decrease in cerebral oxygenation and mental stress can be indexed by an increase in cerebral oxygenation. Tsunashima and Yanagisawa (2009) examined changes in prefrontal activity via multi-channel frontal fNIRS systems in driving with and without adaptive cruise control. Their findings revealed substantial decreases in prefrontal activity when participants drove with adaptive cruise control relative to without, which was correlated with perceived workload (via the NASA-TLX). Similar decreases in activation of prefrontal cortex (lower cognitive load associated with drowsiness) were reported while participants monitored a simulated autonomous car driving task relative to higher prefrontal cortex activation during manual driving task (Sibi et al., 2016). Such findings indicate that optical imaging for cerebral blood flow is a valuable tool for assessing performance and neural efficiency in well-controlled realistic driving contexts.

## Practical Considerations

One important limitation of fNIRS is that, because it relies on the measurement of absorption properties of light as a function of vascular changes in the brain, its temporal resolution is limited by the time-course of hemodynamic activity (on the order of seconds). In contrast, the development of recent 'fast' optical imaging methods, such as the event-related optical signal (EROS; Gratton and Fabiani, 2001, 2003), which measures scattering properties of light as a function of changes in neural activity, have a much higher temporal resolution (on the order of milliseconds). Although applications of this method in human factors research is sparse, fast optical imaging methods have growing promise. While the spatial resolution of optical imaging methods is higher than EEG, such spatial inference is constrained by the penetration depth of NIR light, which reaches only a few cm from the scalp surface. Therefore, imaging of activity from deep cortical and subcortical sources (beyond the outer cortical mantle) is limited. Recent work has also employed wearable fNIRS systems (Piper et al., 2014; McKendrick et al., 2016; Le et al., 2018) and simultaneous collection of fNIRS and EEG (Kassab et al., 2018), which can enable real-world monitoring in ecologically valid settings.

## Heart Rate (HR) and Heart Rate Variability (HRV)

### Heart Activity Quantification

*Heart rate* (in beats per minute or bpm) is the number of heartbeats in 1 min (Jennings et al., 1981). Electrocardiography (ECG) is a well-established method to record the electrical activity of the heart. In psychophysiology, a lead II configuration (i.e., placing the negative electrode in the region of right collar bone, the ground near the left collar bone, and the positive lead over the lower left ribcage, or functionally similar variant) is

commonly used to be able to record electrical activity of the heart via research grade equipment. A single heart beat wave in an ECG signal shows changes in electrical potentials (referred to as the P, Q, R, S, & T components and together they are referred to as the *QRS complex*, for review please see Berntson et al., 2007). The R component (one for each heart beat) is due to ventricular depolarization and for a lead II configuration, it has a larger magnitude and a sharper inflection than the rest of the components making it easily detectable. While heart rate is a count of beat per minute, *heart period* (also called inter-beat-interval) is the time in milliseconds between successive R spikes (Berntson et al., 2007). Heart rate is generally derived by converting mean heart period (in milliseconds) to heart rate (in beats per minute), see Berntson et al. (2007).

Heart data can also be collected via other technique including photoelectric plethysmography (PPG) and photoplethysmography imaging (PPGI). PPG technique includes use of a photocell (such as an infrared light-emitting diode) placed over an area of tissue with blood capillaries that is easily accessible (e.g., finger or ear lobe). Energy emitted from an infrared source passes through the tissue and reflects off the tissue. Changes in blood volume (due to heart beats) in an area can thus be assessed by the amount of light that was reflected back to the photodetector, and thus forms the basis of estimating heart beats (Berntson et al., 2007; Laborde et al., 2017). A similar concept is used in “wearables” which have photo-emitters and detectors placed on a convenient location (e.g., wrists and earlobes) making them easy to wear and collect data from them (Byrom et al., 2018; van Gent et al., 2018). This idea is used in vehicles with photo-emitters and detectors placed on the steering wheels, which allow collecting heart data (heart rate, HRV, and blood volume pulse) while driving. Another advancement in PPG is a contactless measurement technique called PPGI that detects color changes (e.g., the forehead area) in a video due to blood perfusions (Blöcher et al., 2017). Instead of photodiodes used in PPG, PPGI uses detector arrays in cameras to collect image sequences that contain information about bio-signals (e.g., blood volume pulse and respiration). Image and signal processing methods are utilized for beat-to-beat heart rate estimation (Blöcher et al., 2017; Madan et al., 2018).

On a related note, established guidelines for heart beat detection processing, with recommended parameters to derive heart rate and heart rate variability are provided in Jennings et al. (1981), Berntson et al. (2007), and Shaffer and Ginsberg (2017). Custom and open-source software has also been developed to automatically detect R peaks to calculate heart beats. As is true for most physiological measures, data should be visually checked to inspect the ECG data for artifacts and irregularities. Artifacts can be introduced in these data due to numerous reasons (such as motorists’ excessive motion, sneezing and coughing, and irregular heartbeats) any of which can disrupt the ECG measurement or directly impact normal heart-beat patterns. Visual inspection helps insure that the heart beats are correctly marked by the detection software and physiologically improbable values are detected and then corrected.

*HRV* is variability in the time intervals of adjacent heartbeats (Berntson et al., 2007; Shaffer and Ginsberg, 2017). HRV can

be derived from ECG data over a period of time ranging from short intervals (~1–5 min) up to longer intervals (~24 h). HRV metrics can be roughly categorized as falling under time-domain, frequency-domain, or non-linear measures of HRV (for a review see Shaffer and Ginsberg, 2017). Time domain-based parameters calculate the variations in heart beat intervals, such as standard deviation of R-R intervals (SDRR), percentage of successive R-R intervals that differ by more than 50 ms (pNN50), and root mean square of successive R-R intervals (RMSSD). A few time-domain parameters also represent geometric shape of R-R interval distributions, such as the HRV triangular index (i.e., plotting the integral of the ratio of RR interval density histogram by its height) and the baseline width of the RR intervals histogram (TINN), for details see Shaffer and Ginsberg (2017). Frequency-domain based measures transform the beat-to-beat variations in heart beat (R-R intervals) into frequency power bands via Fourier analysis (Task Force of the European Society of Cardiology, 1996). The most commonly used frequency-domain methods are low- and high-frequency power. A low-frequency (LF) power is the energy of heart rate oscillations in a lower-frequency (0.04–0.15 Hz) band. Similarly, high-frequency (HF) power is the energy of heart rate oscillations in a higher-frequency (0.15–0.4 Hz) band (Task Force of the European Society of Cardiology, 1996; Shaffer and Ginsberg, 2017). A peak in these frequency bands can also be calculated, which is an estimate of the peak frequency in the specific frequency band. Non-linear measures of HRV are useful in capturing the unpredictability and dynamic nature of heart rate time-series data (Shaffer and Ginsberg, 2017). Common measures include fitting an elliptical-shape to represent non-linear HRV and calculating approximate entropy (ApEn) and sample entropy (SmpEn), which characterize the complex pattern of time-series heart data (Shaffer and Ginsberg, 2017). Detailed discussions can be found elsewhere (Task Force of the European Society of Cardiology, 1996; Berntson et al., 2007; Laborde et al., 2017; Shaffer and Ginsberg, 2017).

## HR/HRV in Driving Context

### *Over-arousal in driving context*

Heart rate is a commonly measured index of physiological arousal in response to changes in driving demands. One of the most studied over-aroused cognitive states is workload. Numerous studies have examined changes in heart rate as a function of workload (Lenneman and Backs, 2009, 2010; Mehler et al., 2012; Heine et al., 2017). Heart rate was also found to increase while performing visual and auditory dual-tasks relative to single-task of driving in a simulator (Lenneman and Backs, 2009). Similarly, heart rate has been shown to be incrementally higher for systematically more difficult auditory dual-tasks while driving in a simulator (Mehler et al., 2009) as well as while driving on-road (Reimer et al., 2009). These findings of an incremental change in heart have been replicated in younger-aged (20–29 years old), middle-aged (40–49 years old), and older-aged (60–69 years old) adults (Mehler et al., 2012). Thus, heart rate increases with workload due to cognitive demand (Lenneman and Backs, 2009; Mehler et al., 2012; Ruscio et al., 2017; Hidalgo-Muñoz et al., 2018; c.f., Engström et al., 2005). Other efforts have also been made to utilize rhythmic and morphological parameters

of a heart activity to explore mental workload. A recent study examined the influence of mental workload (due to a secondary task) on morphological parameters from ECG while completing a lane change task (Heine et al., 2017). They found that a combination of derived HR and HRV features (such as mean HR, RMSSD, pNN50, etc.) could be extracted from ECG data that could distinguish between workload levels and suggest that a combination of ECG features can be used to detect mental workload (for details see Heine et al., 2017).

Relative to HR, a fewer number of studies have examined HRV, especially in a systematic manner. HRV decreases with increasing task demands (Luque-Casado et al., 2016). HRV has been found to be sensitive to variations in attention levels while driving that may not be necessarily evident in driving performance (Lenneman and Backs, 2009) and thus HRV can have more sensitivity than behavioral measures. LF- and HF-HRV power bands are influenced by driving task (Zhao et al., 2012; Tozman et al., 2015; Wang et al., 2018). A study (Tozman et al., 2015) compared effect of demand levels (boredom, average demand, and high demand) on HRV in a driving simulator. Both LF- and HF-HRV varied for all the three conditions. High task demands reduced both LF-HRV and HF-HRV (Tozman et al., 2015). Some work has indicated that stress-inducing real-world driving tasks lead to increased heart rate and decreased SDNN, RMSSD, pNN50 (Lee et al., 2007). HRV also varies with workload experienced by drivers during simulated driving (Zhao et al., 2012; Heine et al., 2017; Hidalgo-Muñoz et al., 2018) and on-road driving (Lee et al., 2007). In addition, HRV variations due to cognitive workload have also been found in city traffic operators (Fallahi et al., 2016) and unmanned aerial vehicles operators (Jasper et al., 2016). HRV is sensitive to workload increases due to vigilance and situational awareness demands of the task (Saus et al., 2001; Stuiver et al., 2014; Jasper et al., 2016). However, at least one study (Shakouri et al., 2018) found no variation in heart rate variability metrics (RMSSD, LF, HF, and LF/HF ratio) as a function of higher traffic density while driving in a simulator, even though variations in subjective workload were found.

### *Under-arousal in driving context*

HR and HRV are also sensitive to low-arousal states, such as vigilance and drowsiness. Decreases in vigilance over the course of a 3-h continuous driving task were indexed by a significant drop in heart rate over time (Schmidt et al., 2009). Drowsiness experienced in car drivers and aircraft pilots can also be associated with decreases in HR (Borghini et al., 2014). A recent on-road study (Biondi et al., 2018) found that driving a Tesla in semi-automated mode (e.g., autopilot) led to a lower heart rate relative to manual driving on a freeway. Another study found heart rate was sensitive to activity of the Adaptive Cruise Control (ACC) technology (Brouwer et al., 2017). Heart rate increased when ACC decelerated more suddenly compared to instances when the car decelerated more gradually (Brouwer et al., 2017). These findings suggest that heart rate is a sensitive measure that can assess cognitive processing pertaining to advanced technology in semi-autonomous vehicles.

Other studies have found that LF-HRV and HF-HRV vary with fatigue (Liang et al., 2009; Sugie et al., 2016). A recent

study (Wang et al., 2018) found that changes in fatigue levels while driving can be represented by non-linear measures of HRV (e.g., sample entropy). Variations in drowsiness levels can also impact HRV (Noda et al., 2015; Piotrowski and Szypulska, 2017). Another recent study found that variations in HRV (TINN and RMSSD) was higher when participants drove a vehicle in automated mode relative to the manual mode (Biondi et al., 2018). Perhaps, drowsiness and a lack of engagement in the driving task during automated mode may have led to a higher HRV. HRV and blink rates have also been shown to assess sleep onset (Noda et al., 2015). HRV-based assessment algorithms can be used for early detection of fatigue and drowsiness to augment attention and performance (Patel et al., 2011; Zhao et al., 2012; Abe et al., 2016; Vicente et al., 2016).

### **Practical Considerations**

Heart rate and its variability are inexpensive and reliable measures that are relatively easy to record with research-quality equipment that meets recommended guidelines (Task Force of the European Society of Cardiology, 1996). It has good signal to noise ratio as well (R-R peaks can be detected even in very noisy environments). Consequently, it is also not difficult to collect in lab as well as in unpredictable field studies, especially with the availability of mobile data recording systems. However, these advantages can also lead to misuse of this methodology. Great attention to the data collection and processing are required to have meaningful data. Skin preparation (e.g., cleaning with alcohol wipes) before electrode placement and signal monitoring to collect good quality data can drastically reduce post-processing (e.g., Berntson et al., 2007). Participants should be comfortably positioned to avoid physiologically induced changes in heart rate such as altered breathing rate due to postural adjustments. Body movements should be minimized and accounted for as such movements can add noise and also add movement-related heart rate changes. Effective data cleaning to remove artifacts and noise are a must, otherwise heart data will be uninterpretable.

Some recording devices do not utilize the traditional QRS complex from an ECG to calculate HR and HRV. For example, PPG uses a photoelectric sensor that estimates changes in blood volume to calculate HR. There are a few methodological challenges that should be considered before adopting such PPG-based systems. PPG records a lagged cardiac response further away from the heart (e.g., from fingers and earlobes). Unlike ECG based estimates that have a sharp spike for the R component, PPG-based methods instead show a less pronounced curved peak of the blood volume pulse signal, which makes accurate and automatic detection of heart period relatively more difficult (Laborde et al., 2017). Moreover, ECG-based estimates of HR and HRV are recommended for more reliable results because it allows visual inspection and artifact correction of heart data. Such methodological differences between PPG and ECG can explain why PPG and ECG findings are comparable during rest, but are not comparable during stress, for example (Schäfer and Vagedes, 2013).

On a related note, commercialized equipment meant for exercise and fitness tracking fail to meet established guidelines for heart data collection and processing (e.g., minimum



sampling rate and access to raw data for necessary artifact correction methods), which are necessary to make meaningful interpretations (see Berntson et al., 2007; Quintana et al., 2016; Shaffer and Ginsberg, 2017). Similarly, smartphone camera-based assessments have methodological challenges, including very poor sampling rate, illumination variation (due to confounds like weather and time of day), poor signal-to-noise ratio, and motion-related artifacts that can lead to inaccurate interpretations (Laborde et al., 2017; cf., Nowara et al., 2018; van Gent et al., 2018). Ensuring the validity and inter-device variability of wearables (which utilize a PPG-based or camera-based HR system) with an established ECG-based equipment is a necessary step to be able to validate data collected from wearables. However, most commercialized equipment has not been validated in such a manner (Quintana et al., 2016). Without this critical validation step, data collected from commercialized non-research grade equipment does not have convergent validity and should be discouraged by the scientific community until such standards are met. While innovation is critical to be able to collect psychophysiological data in real-world settings, careful adoption and cross-checks with existing gold standards are necessary to make meaningful progress in the adoption of these technologies in real-world driving research.

Moreover, HF-HRV has been found to be impacted by parasympathetic nervous system, however, LF-HRV is influenced by both sympathetic and parasympathetic nervous systems (Berntson et al., 2007; Laborde et al., 2017). Thus, LF-HRV should not be described as a metric of sympathetic activity, but instead be interpreted as a mixture of sympathetic and parasympathetic influences. On a related note, the LF/HF ratio has been a controversial metric as it assumes that LF is due to sympathetic activity while HF is due to parasympathetic (Billman, 2013). The LF/HF ratio was originally based on 24 h recordings, while shorter duration recordings (even 5 min long) have also been calculated. The duration of recording (e.g., 5 min vs. 24 h) can also lead to uncorrelated findings and some metrics are better for short term recordings than others (Shaffer and Ginsberg, 2017).

Another metric we would like to highlight is heart period. Heart rate and heart period have been used interchangeably, however in some instances heart period may be a better choice. Even though, heart rate is more commonly used metric, use of heart period instead of heart rate is recommended measure of autonomic activity because heart period changes more linearly over time (Quigley and Berntson, 1996; Berntson et al., 2007). Heart period should specially be used when comparing changes in heart activity due to experimental manipulation or due to between group differences for short time periods. Further information on heart activity related metrics can be found in detailed reviews (Jennings et al., 1981; Task Force of the European Society of Cardiology, 1996; Berntson et al., 2007; Laborde et al., 2017; Shaffer and Ginsberg, 2017).

Not all heart-based metrics may be sensitive to the variations in cognitive state during driving task. For instance, a study compared several commonly used metrics for HR and HRV cognitive workload during highway driving (Mehler et al., 2011). While HR was robust in differentiating between cognitive

workload in single vs. dual tasks, HRV indices were less robust (e.g., smaller effect sizes). A few HRV indices varied with workload (RMSSD, SDSD, and LF power), however others (SDNN, NN50, pNN50, HF power, and LF/HF) did not significantly differ with workload (Mehler et al., 2011). These findings suggest that depending upon the task, certain indices may be more sensitive to variation in cognitive state than other indices that may be less robust.

In addition, researchers should consider other contextual factors that may vary across participants and may confound study interpretations. A confounding factor that can potentially bias HF-HRV comparisons between conditions of interest is differences in respiration (Grossman, 1992; Berntson et al., 2007; Laborde et al., 2017). Respiration related-parameters should be accounted for by using them as covariates with such HRV indices (for a detailed discussion, see Berntson et al., 2007; Laborde et al., 2017). Similarly, other factors may impact HR/HRV, including task characteristics and motorists' state (relaxation, engagement, and motivation) and activities (smoking and posture). For instance, HRV may increase over time if the task becomes less difficult over time, which may put motorists in a more relaxed state (Jasper et al., 2016). Similarly, HRV may also increase over time with disengagement or demotivation to perform a difficult task (Jasper et al., 2016). Careful consideration of contextual factors will afford accurate and reliable measurement of HR/HRV indices in applied driving settings.

## Blood Pressure (BP)

### BP Quantification

BP (in millimeters of mercury, also written as mmHg) is the force exerted against the walls of the blood vessels (Shapiro et al., 1996; Berntson et al., 2007). Depending upon the stage of the dynamic cardiac cycle, BP differs from lowest to highest levels. During a single cardiac cycle, *diastolic BP* is the lowest level of arterial pressure when the heart is filled with blood and *systolic BP* is relatively the highest level of arterial pressure (Shapiro et al., 1996; Berntson et al., 2007). As invasive methods to record BP require additional safeguards and equipment, most psychophysiology research studies focus on non-invasive approaches to record blood pressure. Three relatively non-invasive methods are auscultatory or oscillometric methods, arterial tonometry, or the volume-clamp methods (see for details, Berntson et al., 2007). The most common method is auscultatory measurement, which records the sounds of blood flow by placing a cuff on the upper arm and a stethoscope placed over the brachial artery to identify the systolic and diastolic blood pressure (Shapiro et al., 1996; Berntson et al., 2007). Physiological arousal during mentally effortful situations leads to greater vasoconstriction and cardiovascular reactivity evidenced by increased heart rate and blood pressure and decreased heart rate variability (Lundberg et al., 1994; Ottaviani et al., 2016). BP increases with psychological stress (Ottaviani et al., 2016) and is correlated with self-reported stress (Lundberg et al., 1994). However, cognitive workload may not reliably influence BP (Elkomy et al., 2017).



## BP in Driving Context

Limited research has examined over- and under-arousal via BP in driving contexts. Systolic BP and BP variability have been found to increase while driving in simulated high traffic conditions that had high workload demands (Stuiver et al., 2014). Fatigue was also associated with a decrease in systolic BP and HR (Liang et al., 2009). However, other studies have not found a reliable effect of stress on BP (Simonson et al., 1968; Littler et al., 1973; Lee et al., 2007). One study found no significant change in BP from beginning to end of the drive with a short period of arterial pressure changes during events such as overtaking that returned to baseline (Littler et al., 1973). BP was also not found to vary in an on-road stressful driving task speed in a simulator even though HRV parameters were significantly impacted (Lee et al., 2007).

Nevertheless, BP is a very useful measure to understand the factors that impact driving performance. One clear example of this comes from a simulator-based study investigating aggressive driving behavior in irregular traffic flow and under time pressure (Drews et al., 2012). Irregular traffic patterns were not found to impact BP. However, male drivers who were under time pressure to drive faster in order to receive a monetary incentive, had elevated systolic BP compared to females under time pressure or compared to male drivers who were not under time pressure. In fact, females did not show any elevated blood pressure under time pressure (Drews et al., 2012). These findings suggest that individual difference factors such as sex differences and motivation to drive aggressively may impact driving behavior and associated physiological signals. Other studies have shown that trait-level variation in BP (such as a history of high BP i.e., hypertension) is an important measure to capture health and age-related impact on driving performance in vulnerable older populations (Lyman et al., 2001; Siren et al., 2004). A 5-year longitudinal study that examined the effect of urban bus driving on BP found that the number of hours driven per week predicted higher diastolic BP (Johansson et al., 2012), suggesting that there are cumulative effects of cognitive demands and stress of continuous driving.

## Practical Considerations

While heart-rate was reported to rapidly change in response to car racing, BP was “less responsive” (Simonson et al., 1968). Other studies have found that BP does not change significantly during on-road driving (Littler et al., 1973; Lee et al., 2007). A few BP recording-related reasons could play a role. BP can rapidly change over time so multiple readings are recommended for a more accurate estimate. However, a limiting factor is the BP equipment. The pressure from a cuff worn by the responder can become uncomfortable and disruptive within a few minutes. Continuous reliable BP measurement (especially via volume-clamp) is uncomfortable, distracting, and potentially disruptive to driving. This limits the frequency of samples that could be collected, which are about 1 reading per minute. Also, the BP recordings are sensitive to movement so in an on-road study, it is less feasible to accurately record multiple BP reading from participants while drivers are actively involved in the driving process. While some alternative methods to

record blood pressure (e.g., plethysmography) may be available, methodological issues similar to those discussed in recording heart activity apply to BP as well and it is crucial to evade poor quality unreliable equipment. In sum, BP provides valuable insights about vulnerable states of the drivers, however, in a real-world driving context, methodological concerns can limit reliable data collection. Much future work is required to be able to measure reliable and non-invasive BP activity.

## Electrodermal Activity (EDA)

### EDA Quantification

EDA, previously known as galvanic skin response, is a change in electrical potentials of the skin that can be used to make interpretations about the psychological phenomena of the responder (Boucsein et al., 2012). EDA can be measured via exosomatic or endosomatic techniques. Exosomatic techniques—a more commonly used method used in applied research—apply a small current through a pair of electrodes and then measure electrical resistance (or its reciprocal, i.e., conductance) from the skin. Because the current is kept constant, it is possible to measure changes in the voltage between the electrodes that will vary directly with changes in skin resistance, following Ohm's law (see Dawson et al., 2007 for a technical review). Endosomatic techniques measure passive changes in intrinsic electrical activity without application of an external current. For details on EDA recording techniques, see Fowles (1986), Dawson et al. (2007), and Boucsein et al. (2012). Higher EDA is indicative of physiological arousal due to increased sympathetic autonomic nervous activity (Dawson et al., 2007; Lohani and Isaacowitz, 2014). EDA is sensitive to physiological reactivity and many other factors, such as respiration and mental effort (Dawson et al., 2007). Commonly derived EDA metrics (Dawson et al., 2007; Boucsein et al., 2012) include slowly varying tonic level of electrical conductivity (skin conductance level; SCL) and phasic increase in magnitude electrical conductance in response to an unexpected or relevant event (skin conductance response; SCR). Non-linear EDA metrics that can differentiate between increased cognitive load vs. recovery phases of stressors have been identified as well (Visnovcova et al., 2016).

### EDA in Driving Context

In driving research, systematic variation in several arousal-related constructs can impact EDA. Most commonly investigated is cognitive workload. SCL is higher during increased workload in dual-task relative to single-task driving (Mehler et al., 2012). A systematic investigation of workload increments in one on-road driving study (Mehler et al., 2012) found a systematic increase in SCL as a function of three levels of auditory workload secondary tasks relative to single driving task for young, middle, and older age groups. These findings suggest that SCL can be used to index workload levels in driving context. High SCR has also been found to increase with workload experienced by motorists while driving on difficult road types that required avoiding more traffic and making more decisions (Schneegass et al., 2013). A recent study reported SCR amplitude increased with cognitive load due to dual-task driving (Ruscio et al., 2017). Additional workload experienced due to texting and navigation (Seo et al., 2017) and

speeding (Kajiwar, 2014) while simulated driving was also found to increase EDA.

EDA also varies with other physiological arousal-related constructs. EDA based indices can be used to detect stressful events during driving (Affanni et al., 2018). A recent study utilized feature extraction and discrimination processing techniques to classify EDA data into low, medium, vs. high stress levels with about 82% recognition rate (Liu and Du, 2018). Another recent study found higher SCLs when participants drove a simulated vehicle in autonomous mode compared to manual mode (Morris et al., 2017). Higher skin conductance levels could be indicative of lower levels of trust in the autonomous mode than manual mode. State anxiety during simulated driving was also found to be associated with SCL (Barnard and Chapman, 2018). Another recent study found that relative to sleepiness, higher skin conductance levels are found during wakefulness, effects which are indicative of comparatively higher sympathetic activity (Schmidt et al., 2017).

### Practical Considerations

In driving contexts, EDA is shown to vary due to many cognitive states, such as workload, stress, anxiety, sleepiness, all of which are influenced by sympathetic nervous system activity. This allows the use of EDA in assessment of various psychological phenomena (Dawson et al., 2007). Therefore, caution should be exercised while interpreting changes in EDA in an applied and less-controlled setting as it is sensitive to not one, but many psychological variables. In the driving context, careful choice of filters to remove artifacts (Affanni et al., 2018) and identification of cognition-related features (Chen et al., 2017; Liu and Du, 2018) that have been successfully implemented could be utilized to improve accuracy and detection. One disadvantage of EDA is that it has a slower response (lag of 1–3 s) after the stimulus has occurred (Dawson et al., 2007). In instances when near-real time physiological responses need to be detected, EDA may be relatively slower (than cardiovascular measures). Another point to consider is that, similar to other physiological measures, not all individuals have the expected skin conductance response (Dawson et al., 2007). This is another reason to avoid reliance on a single measure, but multiple channels, to capture the psychological phenomena of interest.

## Electromyography (EMG)

### EMG Quantification

EMG is used to measure the electrical activity generated by muscle fibers (Fridlund and Cacioppo, 1986; van Boxtel, 2001). Surface EMG is captured by placing small surface electrodes on specific muscles of interest, which is then digitized and amplified to record muscle activity (Fridlund and Cacioppo, 1986). Numerous features can be extracted from the EMG signals. Root mean square of the signal (in microvolts) is a recommended and commonly reported EMG signal amplitude (Fridlund and Cacioppo, 1986). Other commonly assessed statistical features are peak spectral density, peak amplitude, and peak frequency. A specific muscle's activity can provide insights into the psychological processes underplay. For instance, the smile muscle (or zygomaticus major) and the frown muscle (or corrugator

supercilii) have been used a lot in emotion research to identify positive and negative behavioral expressions. For example, more frown muscle activation can be an index of negative behavioral expressions (Lohani and Isaacowitz, 2014; Lohani et al., 2018). Psychological processes (e.g., stress) can lead to sympathetic nervous system activity (Lundberg et al., 1994), which can elicit muscular tension. Researchers have studied muscular activations under controlled conditions to index mental processes (Lundberg et al., 1994; Wijsman et al., 2013; Luijckx et al., 2014). Applied driving research has successfully assessed psychological processes by assessing EMG (Healey et al., 1999; Fu et al., 2016; cf., Morris et al., 2017; Ma et al., 2018).

### EMG in Driving Context

In driving contexts, surface EMG has been utilized to study psychological and physiological stress (Jonsson and Jonsson, 1975; Wikström, 1993; Balasubramanian and Adalarasu, 2007; Ahlström et al., 2018). Stress and fatigue have been studied by recording electrical activity from relevant muscles. For instance, variations in the trapezius muscle (a major back muscle that extends from the neck to shoulder blades and lower spine) and deltoid (triangular muscle located on uppermost part of an arm and the top of shoulder) are influenced by mental stress (Wikström, 1993; Balasubramanian and Adalarasu, 2007; Hirao et al., 2007; Wijsman et al., 2013; Luijckx et al., 2014; cf., Morris et al., 2017). A recent study (Lee et al., 2017a) recorded trapezius muscle activity to detect stress in a driving simulator under relaxed and stressed conditions. A continuous increase over time in muscular tension was associated with greater stress experienced due to driving task (Lee et al., 2017a). Muscular tension can thus be a useful metric of stress level that can be utilized in driving research.

It is worth noting that muscular fatigue and discomfort are not isolated issues (Leinonen et al., 2005) and they cause psychological distress and disrupt cognitive performance while driving. Muscle fatigue while driving has been studied by examining changes in muscular tension in shoulder and neck muscles (Sheridan et al., 1991; Wikström, 1993; Balasubramanian and Adalarasu, 2007; Hirao et al., 2007). Compared to the beginning of the drive, continuous driving can lead to reduced back muscles (e.g., trapezius and deltoid) activity and fatigue. Muscular fatigue (measured by EMG of back muscles) is associated with decreases in power of EMG activity-related frequency band (Hostens and Ramon, 2005; Balasubramanian and Adalarasu, 2007; Hirao et al., 2007). Surface EMG is a helpful way of identifying discomfort in fatigued and weak muscles and targeting rehabilitation for skeletomuscular problems specially in professional or long-distance drivers (Balasubramanian and Adalarasu, 2007). A recent study (Artanto et al., 2017) has also used a low-cost EMG system to detect drowsiness. An EMG sensor attached to muscles around eyelid region captured the duration of eyelid closure as an indicator of drowsiness (Artanto et al., 2017). Another recent study has proposed a system that can detect real-time changes in EMG (Mazzetta et al., 2018). Further research is needed to validate EMG's applicability in real-world settings.

## Practical Considerations

EMG measurement enable recording continuous data from the specific muscle of interest without obstructing the driving task. Such objective information can be helpful in learning about muscular activity (and relevant cognitive states) that may not be necessarily visible to the researchers or under the awareness of the responder. However, it is essential to pay attention to any outliers or irrelevant events that may add noise to the EMG signal and impact signal interpretation. Irrelevant events can include muscular activity due to driving-unrelated (e.g., continuous posture change, scratching skin, or touching the electrodes) and driving-related (e.g., functional steering activity) movement and yet unrelated to the cognitive state (e.g., mental workload) of the driver (Mehler et al., 2009). In real-world settings, it can be tedious to tease apart muscular activity due to other confounding reasons from activity relevant to changes in cognitive states. Furthermore, the task under investigation is also of importance. For instance, a study that compared muscular tension while driving car autonomously vs. manually found no differences in EMG signals, but significant differences were found for SCL (Morris et al., 2017). This suggests that for some tasks the muscular activity may not significantly differ, but may still be psychologically different in other modalities. This also highlights the importance of multiple measures.

## Thermal Imaging

### Thermal Imaging Quantification

The measurement of changes in skin temperature is a useful technique to detect and track attributes of a responder, such as body posture and emotional expression (Gade and Moeslund, 2014; Rai et al., 2017). A special merit of this technology is that it enables sensing the real-time state of motorists non-invasively without disrupting driving related tasks. In addition, unlike RGB cameras, thermal cameras do not depend on an external illumination (Gade and Moeslund, 2014; Rai et al., 2017). Objects that emit radiations in the mid-to-long wavelength infrared spectrum (3–14  $\mu\text{m}$ ), such as the human body (but not inanimate objects) can be detected via thermal imaging (Gade and Moeslund, 2014; Rai et al., 2017). Changes in temperature distribution, as captured by the thermal cameras, are utilized to make meaningful interpretations. For instance, facial thermography can be used to capture the heat distribution in facial locations known to vary with sympathetic activity as a metric of the varying psychological phenomena. Most commonly investigated facial locations include the forehead and nasal temperature changes.

Sympathetic autonomous nervous system activation may lead to constrictions of blood vessels, thereby decreasing temperature in extremities, such as the nose (Or and Duffy, 2007; Gade and Moeslund, 2014). For example, mental workload changes lead to temperature variations in the forehead, nose, cheeks, and chin regions (Stemberger et al., 2010; Marinescu et al., 2018). A recent study examined the validity and sensitivity of thermal imaging in assessing variation in cognitive load (Abdelrahman et al., 2017). Increased cognitive task difficulty led to significant increases in the forehead temperature and decreases in nose temperature (Abdelrahman et al., 2017). The largest effect sizes were found

when the difference in forehead and nose temperature was estimated. Higher task difficulty led to an increase in forehead and nose temperature differences (Abdelrahman et al., 2017). Additional work has also examined real-time sensitivity of thermal imaging and found that specialized thermal cameras can detect changes in cognitive load with a latency of 0.7 s post eliciting event (Abdelrahman et al., 2017). This finding suggests that this methodology has a high relevance for real-time assessments of cognitive load in applied settings like driving.

### Thermography in Driving Context

In driving contexts, facial thermography was found to be useful in assessing over-arousal constructs such as mental workload (Or and Duffy, 2007; Murai et al., 2008). Performing a secondary workload task (mental arithmetic) while driving in a simulator as well as an on-road car led to a decrease in nasal temperature with stable forehead temperatures (Or and Duffy, 2007). Drop in nasal temperature also correlated with self-reported workload (Or and Duffy, 2007). Another study found increases in the difference between nose and forehead temperature increased with mental workload (Kajiwar, 2014). Participants' nasal temperature varied as a function of mental workload in simulated driving (Kajiwar, 2014). Workload variation indexed by changes in nasal temperature were also reported during ship navigation using a simulator (Murai et al., 2008), highlighting its utility in applied settings.

Furthermore, facial thermography can be useful to examine and infer heat distribution in faces during emotional states. This method could be promising and may provide a non-invasive approach to capture emotional states because current methods of emotion recognition using facial features detection software have limitations. One study used an infrared thermal camera to non-invasively detect face regions and recognize emotional states of motorists (Kolli et al., 2011). This study suggests that thermography can improve face detection algorithm for in-vehicle settings thereby facilitating ADAS.

In another line of work (Cheng et al., 2007), a combination of thermal infrared and color cameras have shown to be effective in sensing body movements in real-time on-road driving. Similarly, infrared streaming has been used to develop posture and occupancy sensory systems (Kato et al., 2004; Trivedi et al., 2004). Another recent study reported successful use of near-infrared light and thermal camera sensors to identify aggressive driving behavior (Lee et al., 2018) and were able to categorize aggressive driving from relaxed driving. The above studies suggest that thermography has the potential to be a useful non-invasive technique that can be validated to capture cognition-relevant states and improve traffic safety.

### Practical Considerations

Thermal cameras are used in numerous industrial, agricultural, and military settings (Gade and Moeslund, 2014). They can be extremely useful in vehicular technology because they are non-contact sensors and can work regardless of external illumination. Nevertheless, further testing is needed to better understand how this technology would improve our



understanding of cognitive states in traffic safety. Further systematic investigation and replication of thermography as a function of cognitive workload, stress, and drowsiness after controlling for confounding factors, such as environmental factors (e.g., weather conditions and air conditioning), are needed to be able to make confident assessments of cognitive states. The results so far look promising.

## Pupillometry

### Pupil Quantification

Pupillometry is the measurement of pupil size and reactivity. Modern pupillometry is measured via optical eye-trackers that use some combination of monitoring infrared light reflections from the cornea, the back of the lens, and the pupil, as well as absorption of light by the pupil (e.g., dark-pupil tracking). Most modern eye-tracking devices can monitor pupil location (and eye-fixation location) with very high resolution ( $>1,000$  Hz) non-invasively and at a substantial distance from a participant. Thus, measurement can occur in highly ecologically valid environments, without participants having to make any overt responses. Since the 1960's it has been shown that pupil dilation changes as a result of mental activity—for example, increases in arousal and cognitive workload (e.g., Hess and Polt, 1964). In a classic study demonstrating the sensitivity of pupillometry to cognitive demands, Kahneman and Beatty (1966) showed that pupil dilation increases parametrically with an increasing number of words to recall in a simple word list memory task. Moreover, they showed that this increase in workload persists over a maintenance interval, and reduces parametrically as each word is retrieved (and released) from memory. These findings, along with a number of other demonstrations of pupillary sensitivity to cognitive workload, for example in math problem solving (Sirois and Brisson, 2014), working memory and individual differences in intelligence (Tsukahara et al., 2016), aging and verbal memory load (Piquado et al., 2010), has led to wide interest in this measure as a physiological marker of arousal and cognitive effort.

Janisse (1977) remarked that the eye is the only “visible part of the brain.” Indeed, detailed models of the neurophysiology of pupillomotor functioning are developed and growing, including an understanding of the innervation of the sphincter and dilator muscles by the autonomic nervous system (Miller et al., 2005), as well as the neuromodulatory relationship between pupil dilation, activity in the locus-coeruleus (LC; a neuromodulatory nucleus in the dorsal pons of the brainstem strongly linked to phasic and tonic arousal, cognitive control, and monitoring functions), and norepinephrine (Gilzenrat, 2006). For instance, a high correlation (0.6) between spike frequency and pupil diameter has been found, whereby large pupil diameter equates to high LC activity (Rajkowski et al., 1994). Demberg (2013) have also recently reported changes in pupillometry due to linguistically induced cognitive load (e.g., comprehending syntactically demanding sentences). Other recent work has also examined user state related changes in pupil diameter in lab-settings such as variations in valence and arousal (Kassem et al., 2017) and interest in real-time (Jacob et al., 2018).

## Pupillometry in Driving Context

Eye-tracking has been used extensively in studying visual perception and attention in driving contexts, however the unique use of pupillometry as an index of real-time physiological indicator of cognitive workload is only lately growing in popularity (Schwalm et al., 2008). For example, Cegovnik et al. (2018) recently validated a low-cost eye-tracker and showed that pupil dilation increases with increments in cognitive load due to a secondary memory task (n-back) (see also Recarte and Nunes, 2000 for similar results). Pupillometry has also been adopted in driving research while motorists drove in a simulated driving context. Pupil diameter was found to reliably increase with increases in cognitive load (Palinko et al., 2010; Faure et al., 2016). Other work has use machine learning algorithms to detect cognitive load while driving from pupillometry data (Yoshida et al., 2014). A recent study found that during simulated driving, pupil dilation could detect increases in cognitive load imposed by a secondary task within a lag of 1 s (Prabhakar et al., 2018). This suggests that pupillometry could be used as a near-real time index of cognitive load.

Pupillometry has also been used to differentiate between alertness and drowsiness (Soares et al., 2013). Alertness is associated with increased mean pupil diameter and decreases in standard deviation (i.e., stable), whereas drowsiness is associated with decreases in diameter, but increases in standard deviation (i.e., fluctuations) in pupil diameter (Morad et al., 2000; Wilhelm et al., 2009). Fluctuations in pupil size have been proposed to be a reliable index of drowsiness-related impairment while driving (Maccora et al., 2018). Pupil dilation was also found sensitive to fatigue levels while driving with a decrease in fatigue being associated with an increase in pupil diameter (Schmidt et al., 2017). Although early, these findings, along with others (for a recent review see Marquart et al., 2015; Maccora et al., 2018) suggest that pupillometry is an efficient, ecologically valid, and low-cost physiological reporter variable for indexing cognitive states in driving in highly-controlled environments like realistic driving simulators.

## Practical Considerations

In lab settings, pupil diameter was found to be a reliable, non-invasive, and real-time measure of workload (Marinescu et al., 2018). However, in on-road settings, it is quite challenging to capture interpretable pupil information due to large variations in luminance that are hard to control across conditions and participants. Indeed, photopupillary reflex is massive in magnitude relative to changes in pupil size related to cognitive and attentional factors. As such, if there are considerable changes in lighting conditions (e.g., sunny vs. cloudy days), this can create considerable noise in the pupillary signal. Moreover, if specific conditions of interest are confounded with respect to overall luminance (e.g., driving during the day vs. driving at night), this overall pupillary light reflex-related shift should be taken into consideration. Furthermore, if investigating event-related pupillary responses in driving, one should be careful to determine that differences in pupil dilation are not only due to differences in visual stimulation (e.g., presenting a luminant STOP sign). Modeling techniques have also developed methods



to infer cognitive workload after accounting for some variations in lighting conditions (Pfleger et al., 2016; Reilly et al., 2018).

Marshall (2002) have developed a signal processing method for extracting high-frequency changes in pupil dilation that they argue is uniquely related to cognitive components (*Index of Cognitive Activity* or ICA). However, this method is a commercially available “black box” system, and should be interpreted with caution given that the exact algorithm used to calculate ICA from raw pupillometry is not open source. Other work has estimated an *Index of Pupillary Activity* (IPA) inspired by ICA, that uses wavelet-based algorithms to decompose pupil data (Duchowski et al., 2018). IPA was found to differentiate between low vs. high mental workload (Duchowski et al., 2018). Another important feature to consider is that measurement of pupil dilation is affected by eye-movements and relative gaze position (e.g., Gagli et al., 2011). When gaze position changes from central to peripheral locations, the recorded pupil shifts from a circular to an elliptical shape from the point of view of fixed camera location. This change in the recorded geometry of the pupil is accompanied by changes in overall pupil size, irrespective of actual changes in dilation or constriction. Gagli et al. (2011) have developed methods for the measurement and removal of such systematic influences. Nevertheless, researchers should be careful to measure gaze position and to design studies such that likely visual target locations are not confounded across conditions of interest.

## CHALLENGES AND RECOMMENDATIONS

Psychophysiological research has made tremendous progress in developing methods to quantify cognitive processes. Most of this research has been conducted in carefully controlled environments to be able to interpret with certainty what changes in a physiological signal may imply about the psychological phenomena under investigation. Physiological signals are valuable to understand how people interact in real-world contexts. Driving research is an excellent application of psychophysiological methods to understand and interpret how people interact with automation in natural settings, which in turn can inform intelligent systems to improve driving performance and safety. As evidenced by much of the growing research base discussed above, psychophysiological measures can be successfully adopted to meet these goals. At the same time, lack of adherence to research protocols and guidelines can seriously jeopardize meaningful use of these methodologies. Here we highlight a few general challenges and recommendations that cut across all psychophysiological measures in driving research when collecting data from real-world driving settings—which are less predictable than lab settings—to improve data-quality and aid in effective interpretation.

### Valid and Reliable Quantification of Construct

Depending upon the task and setting (lab-based simulator or field study), some physiological measures will be more suitable and feasible than others. For example, in a simulator with very

controlled body movement, continuous blood pressure using the volume clamp method can be collected. However, while on-road, this equipment may compromise drivers' safety and thus is not feasible. Other measures like ECG and thermal cameras are highly mobile and feasible. Careful observations can allow interpretation of cognitive processes while driving. One important concern is the possibility of misinterpreting the relationship between physiological signals and cognitive processes (Cacioppo and Tassinari, 1990; Cacioppo et al., 2007). Often, physiological measures (such as HR, EDA, EMG) are impacted by multiple processes, such as drowsiness, stress, and workload, which can lead to interpretive caveats. Systematic variations in different experimental conditions can help tease apart the underlying mechanism causing autonomic activations to be able to draw clear inferences. However, in an applied setting like driving a car in unpredictable traffic, control over the experimental task is largely out of the control of the researcher. Confirmatory independent measures are important to validate the construct of interest in the study. Similarly, it is helpful to ensure that the construct of interest reliably varies across conditions and that the experimental manipulation was effective.

### Individual Differences

A combination of factors may influence physiological signals, including trait-level variables such as demographic factors (age, gender), task experience (professional, experienced, inexperienced), anxiety, and certain health conditions and medications (e.g., cardiovascular health). State-level variations such as stress-levels unrelated to task, caffeine intake (which may change autonomic activity), and engagement/motivation and frustration during the task can also interact with individual differences in ways that may not be readily apparent. Combining data from participants after considering such trait- and state-level variables can help in proper interpretation of study findings.

On a related note, a critical challenge in multi-modal recordings is that individuals may be highly reactive as assessed by one measure but not necessarily, according to another. There is considerable variability across individuals in how closely physiological, behavioral, and subjective measures covary over time with one another (Lohani et al., 2018). Furthermore, it is possible that only some individuals may be sensitive to the experimental manipulation (Drews et al., 2012). Such individual differences may lead to variations in psychophysiological assessments and may also explain to some extent lack of significant differences across experimental conditions. Many, if not all, of these measures are currently utilized within paradigms where we are studying relative changes in the outcome across conditions (e.g., P3b amplitude is a difference wave, HRV% change, %signal change in BOLD response, etc.), for which these measures do not have currently well-understood absolute thresholds for making strong absolute judgements. While there isn't a fixed threshold for physiological measures that can be used across individuals to define high and low arousal levels, relative changes from baseline can be a useful way of assessing variations in arousal levels from optimal levels for the individual. If the system can be calibrated on what is a “normal” range for an

individual, then significant variations from this calibrated range can be a way to detect sub-optimal arousal levels.

## Baseline Assessments

Baseline assessments provide insights about the physiological state of the responder when the experimental condition was absent. It also allows to control for physiological activity due to any prior conditions, so that the change in the experimental condition of interest is interpreted relative to the state right before the condition started. A single baseline is generally not enough, especially when there are multiple conditions. It is a good practice to capture as many baseline assessments and as close to the experimental condition as possible. Another alternate design to consider (for measures with high temporal resolution) is an event-related design, where activity is time-locked to specific events of interest. In this design, pre-event activity in the measure is subtracted from the overall physiological time series, resulting in a strong baseline control for each trial (e.g., ERPs).

## Sampling Rate, Filtering, and Signal Quality

Nearly all physiological signals discussed above are analog signals, which have to be digitized for further processing. Choice of optimal sampling rate and filtering helps avoid signal distortions (Jennings and Allen, 2016), and as such, knowledge of signal processing characteristics of the target physiological measures is necessary for researchers to effectively use these tools. Optimal sampling rate differs by the physiological signal's frequency characteristics, and poor sampling rate can distort waveform characteristics, and induce artificial oscillatory characteristics that are not part of the true analog signal (i.e., aliasing). For example, for HRV analysis, the recommended sampling rate is at least 250 Hz (Task Force of the European Society of Cardiology, 1996). Some commercial wearables (e.g., fitness-related wrist watch sensors) have sampling rate as low as 60 Hz, which will lead to signal aliasing (Jennings and Allen, 2016) and inaccurate and uninterpretable HRV values. The sampling rate needs to be at least above the Nyquist frequency (2x the sampling rate of the highest frequency), and current standards suggest a sample rate 3–4 times the highest frequency component of physiological signal. Advancements in modern computing allow for research-grade equipment to sample far above Nyquist for most of the measures discussed (>2,000 Hz) during data acquisition. Of course, data can always be down-sampled post data collection. As discussed in sections “Heart Activity Quantification” and “Practical Considerations” on heart activity, quantification using wearables can lead to inaccurate assessments (Laborde et al., 2017) due to poor sampling rates, lagged responses, and noisier signals to name a few, which would lead to inaccurate interpretations.

Filters are helpful in getting rid of artifacts and noise not relevant for the physiological signal being processed. For instance, muscle and electrical noise (around 60 Hz) are not meaningful while interpreting EEG and ERP data, and thus data outside the range of interest (typically not higher than 40–50 Hz) can be bandpass filtered. However, if EMG activity, which has a much higher frequency content, is of interest, then bandpass filtering with allow low-pass cutoff at 500 Hz

and high-pass cutoff at 20 Hz, is often suitable (van Boxtel, 2001). Visual inspection pre- and post-filtering process can help determine how filtering is affecting a signal. Note that all filters distort the waveform and spectral characteristics, so unnecessary filtering should be avoided and researchers should take care to understand exactly how filters are impacting their data in time and frequency domains.

For each psychophysiological measure discussed, researchers have a growing number of indices that can be examined (for example, for HRV, time-based, frequency-based, and non-linear measures can be derived). Choice of metrics should be carefully evaluated, as some metrics may be more suitable to meet the goals of the study, while others may not be suitable. For instance, some metrics require minimum duration of data and falling short of such requirements will lead to misrepresentative findings (e.g., standard deviation of R-R heart beats or SDRR is considered more accurate when calculated over 24 h vs. 5 min or shorter intervals; Shaffer and Ginsberg, 2017). Such choices should be made a priori, based on the research question of interest and links between a measure and its purported psychological interpretation based on prior research. Such flexibility in multi-modal recording comes at the cost of an increasing number of “experimenter degrees of freedom,” that can lead to inflated Type-I error rates, if a consistent analysis pipeline is not followed. It is also important to use comparable durations of physiological signals across conditions and participants for appropriate interpretation. Finally, great attention to accurate event markers is critical for valid interpretation within and across participants in event-related designs. This can be an issue when using commercial products that are not designed for research purposes.

## Innovation

A limitation of most current psychophysiological research-grade measures is the need for using contact sensors (placed on skin). Non-contact sensors are beginning to be tested in applied settings, which can make physiological data collection even less invasive. For instance, ECG data can be derived from high-quality RGB cameras, or sensors could be placed on the steering wheel and driving seats (but should meet the recommended requirements). While these can potentially be a great approach to counter the limitations of contact sensors, caution is advised while considering them because new limitations or inaccuracies in assessment are possible and further research and testing is required to adopt them in research. Commercial products may not meet the requirements recommended by the scientific community, which can lead to poor data quality and invalid interpretations. For example, smartphone camera-based PPG sensing estimates have poor sampling rate and can lead to inaccurate assessments (Laborde et al., 2017). It is essential to ensure that the guidelines for measures are met before investing time and resources to avoid technical issues in data collection and interpretation. For instance, as discussed earlier, it is critical to collect physiological data with recommended frequency sampling to avoid aliasing (Jennings and Allen, 2016). Only equipment that have been or can be validated against research-grade devices should be adopted for research purposes.

## Classification

Reliable and valid assessment of cognitive states is the groundwork to develop inputs to advance state detection-workload managers and “aware” systems. For instance, a recent study reported a reliable method to elicit stress in naturalistic driving scenarios (Baltodano et al., 2018). Given that one measure may not be enough to reliably measure subtle changes in cognitive state, a multi-method approach is critical to capture state-level variations that may not be apparent through a single measure alone. Research has shown that multi-modal approaches provide a reliable (Schmidt et al., 2011; Borghini et al., 2014; Chen et al., 2017) way to sense and assess cognitive states of motorists in real-world settings. Notably, due to the dynamic nature of the physiological signals, conventional linear approaches are not always appropriate in modeling and predicting cognitive state (Chen et al., 2015). The discussed physiological signals are often non-stationary overall but for the briefest periods of time. As such, innovative methods of combining temporal and spectral resolution (time-frequency analysis) have been developed in some domains (e.g., EEG), but their application to other physiological signals is only in its infancy.

Once data have been processed to remove artifacts or irrelevant noise, machine learning techniques could be trained on these data to identify “risky” sub-optimal levels of cognitive states, such as low-arousal states of drowsiness and fatigue associated with unsafe driving performance. During the training phase, multimodal features extracted from physiological training data could be used to train models to classify observations into high-arousal states (e.g., due to high stress and workload), optimal-arousal state, vs. low-arousal state (e.g., due to drowsiness and fatigue). During the test phase, the fully-specified machine learning algorithm can be tested in terms of its capacity to accurately classify observations into respective arousal states. Indeed, cognitive state detection based on multimodal feature analysis and classifiers have been also used to detect stress (Yang et al., 2016; Chen et al., 2017; Lee et al., 2017b), alertness and drowsiness (Forsman et al., 2013; Correa et al., 2014; Chen et al., 2015; Wang and Chuan, 2016), fatigue (Fu and Wang, 2014; Wang, 2015; Fu et al., 2016; Li et al., 2017; Wang et al., 2017), and workload (Borghini et al., 2014; Yang et al., 2016) in real-time. Such studies have integrated data from more than one measure by conducting multi-modal analysis to extract the relevant features to capture the psychological phenomena at hand. A comparison of multiple classifiers to train & optimize machine learning algorithms can help determine the best fitting model to represent changes in cognitive states that can explain driving performance (Nadeau and Bengio, 2000; Fairclough et al., 2015; Balters and Steinert, 2017; Tran et al., 2017). Thus, utilizing multi-modal physiological signals, models could be trained to learn and predict motorists’ sub-optimal cognitive states associated with unsafe-driving behavior.

The optimized machine learning algorithms could accordingly inform advanced state detection managers to trigger warnings or otherwise intervene when sub-optimal cognitive states associated with risky driving behavior are detected (Aidman et al., 2015). The ability to predict unsafe

levels of physiological arousal will enable targeted augmentation to modify motorists’ cognitive state to promote safer driving behavior (Schmidt and Bullinger, 2017; Schmidt et al., 2017; Aricò et al., 2018). For instance, countermeasures to augment cognitive states, such as thermal stimulation (Schmidt and Bullinger, 2017; Schmidt et al., 2017) and warning signs or verbal communication (Schmidt et al., 2011; Aidman et al., 2015) can be used by an automated system to modify drivers’ cognitive state. This may especially benefit vulnerable groups such as inexperienced drivers (Noordzij et al., 2017; Yan et al., 2017) and older (Costa et al., 2017) drivers who may be more susceptible to cognitive overload. Furthermore, a person-centered approach can account for individual differences, such as the role of age, driving profile, trust, and reliance on automation. For instance, a recent study used discriminant analysis to account for motorists’ driving-styles and individual difference factors (e.g., gender, age, anxiety, anger) and also identify motorists’ EEG and EDA response features to classify motorists’ safe vs. risky driving tendencies (Liang and Lin, 2018). This study shows that individual differences can explain variations in driving performance and a customized approach may also help improve model prediction over time by accounting for motorists’ characteristics and preferences. For example, the low, normal, and high physiological arousal ranges will vary depending on attributes such as anxious, risky, and distress reduction driving styles of an individual (Liang and Lin, 2018) and prediction of cognitive state-level variations may be more accurate when predictions account for such individual-level variations. Thus, a person-centered approach will improve reliable predictions of cognitive states in real-world contexts by intelligent driving systems.

## RESEARCH APPLICABILITY IN REAL-WORLD SETTINGS

As the reviewed literature in section, “Psychophysiological Measures to Assess Cognitive States” suggests, many interrelated states could lead to a similar pattern of findings on a physiological measure (e.g., mental fatigue, drowsiness, lower vigilance, and mind wandering are all sensitive to similar EEG/ERP indices). After considering the overlap across findings from interrelated constructs, in **Table 1** we have summarized the expected pattern that each physiological measure will have during a low vs. high arousal state in an applied driving context. There are a few points to consider. First, changes in several related cognitive states can lead to similar changes in arousal. For example, increases in driver workload, stress, or vigilance may occur under different contexts, but may similarly lead to heightened arousal. Second, even though arousal is continuous, we chose to classify driver states into categories of low and high arousal because both extremes are sub-optimal for driving performance. Third, cognitive states are complex and change across time. For instance, in the current review, we have placed mind wandering in a low-arousal state based on similar patterns of findings as drowsiness. However, mind wandering is a convenient short-hand for a



more complex constellation of non-externally directed cognitive states (see Smallwood and Schooler, 2006 for a review) and depending on the context, such mind-wandering states can yield states of heightened-arousal as well. Similarly, fatigue can be categorized as high-arousal due to prolonged cognitive overload or it can be passive, because of underload due to monotonous driving conditions, for example (Saxby et al., 2008; Matthews et al., 2019). With further empirical evidence in naturalistic environments, a better characterization of complex cognitive states could be developed.

It is still an open question if interrelated cognitive states could be successfully differentiated from other similar states in naturalistic environments (see Cacioppo et al., 2007 for challenges with psychological inference). However, physiological measures could be used to assess sub-optimal levels of general arousal in real-world settings and intelligent systems can use this information to trigger augmentation strategies even if we cannot fully differentiate between specific cognitive states besides along their arousal axis. We have reviewed how physiological responses across multiple measures can provide a rich array of response data relevant to domains that are of interest to driving researchers (e.g., attention, fatigue, workload, etc.). These measures provide unique information and unique sensitivity to experimental manipulations beyond behavioral responses alone. Thus, their current and future utility in real-world driving research is important. This does not mean that measuring one or even a large number of these measures alone will provide us with a *direct* interpretation of a covert state (e.g., becoming increasingly frustrated about an aggressive driver behind you). Before the state of the research matures to be able to address such a lofty goal as predicting specific cognitive states (Yarkoni and Westfall, 2017), we first need careful on-road experimental work to understand the sensitivity and specificity of these measures to specific changes in driver-relevant states in observational and experimental research in real-world settings. Thus, the focus of the current review is not to claim that measurement of multiple physiological measures in real-world driving could accurately predict motorists' specific cognitive state. Rather, our goal is to summarize the feasibility of each of these measures for integrating high-quality psychophysiological methodology into real-world driving research. **Table 1** presents the current working predictions that are expected based on the available literature, but more work is needed to be able to use physiological signals to infer psychological processes. The current review represents a summary of initial steps in that direction.

In **Table 2**, we have summarized the research applicability of the reviewed psychophysiological measures. Although all of these measures can provide valuable insights in the controlled settings of a lab, some measures are more feasible to use and interpret than others in real-world driving contexts. A few factors that may play a role in determining the practical use of physiological measures in applied settings are: the degree of coupling between the measure and subtle changes in cognitive states, temporal resolution, psychometric reliability, ease of data collection (e.g., setup time), sensitivity to artifacts, and the degree of invasiveness and disruption to normal driving. After considering the available evidence, we have categorized each measure's real-world research

applicability into low, medium, or high levels. Moreover, certain measures may be better candidates than others for a near real-time assessment in applied settings. We review the real-world applicability and feasibility of each of the measures in **Table 2**.

Some promising work suggests that cardiovascular measures may be robust in detecting near real-time changes across multiple domains. Studies have shown that cardiovascular data can reliably detect changes in workload (Mehler et al., 2009, 2012; Lenneman and Backs, 2010; Stuiver et al., 2014), fatigue (Patel et al., 2011; Matthews et al., 2019), and drowsiness (Vicente et al., 2016; Kurosawa et al., 2017). Like any physiological signal, cardiovascular data is susceptible to artifacts that could otherwise lead to inaccurate estimations. However, recent analytical advances have led to an improved use in real-world settings even in the presence of substantial recording artifact. For instance, an analysis approach using short segments of cardiovascular data (e.g., a moving window of 30 s; Stuiver et al., 2012) can be used to detect workload demands during driving (Stuiver et al., 2014). Use of smaller temporal windows of data allow for an investigation of the short-term effects of cognitive state without being overly susceptible to artifacts. Recent work has shown that frequency analysis techniques on ECG data can also be utilized to detect early onset of fatigue (Matthews et al., 2019). While the limitations of PPG discussed earlier still apply, recent preliminary work using near-infrared illumination PPG (which overcomes confounds of illumination and motion-related inaccuracies) while driving seems a promising direction for future practical applications (Nowara et al., 2018). Another recent work has developed a noise-resistant algorithm specifically designed to analyze PPG waveforms (van Gent et al., 2018), which can provide researchers an open-source and validated heart rate analysis software to overcome some existing limitations of PPG data processing, making it more feasible for applied driving research.

EDA has been found to be a robust measure of sympathetic arousal in driving contexts in real-world settings (Mehler et al., 2012; Schneegass et al., 2013; Ruscio et al., 2017). EDA is also easy to set up and collect from a motorist without obstructing the driving process. Even though it has a slower response time and provides only a broad sense of arousal (a combination of workload, stress, fatigue, etc.), EDA in an applied uncontrolled environment can estimate relative changes and periods of stability in sympathetic activity of a motorist with an upper temporal resolution of approximately 3–5 s. For example, recent work found EDA to be suitable in capturing stress-level variations in a real-time unconstrained setting (ElKomy et al., 2017). Feature extraction and pattern recognition algorithms have also shown reasonable success recently in detecting changes in cognitive states (Chen et al., 2017; Liu and Du, 2018). Moreover, adaptive filters have been successfully used to remove motion-related artifacts for automatic and accurate detection (up to 95% sensitivity) of state-level variations in cognition (Affanni et al., 2018). Such recent processing and analytic advances with EDA data has shown its high relevance in applied intelligent automation. For example, a development approach proposed for monitoring driver's fatigue levels and functional state utilizes automated analysis of EDA indices in their detection module



to improve intelligent vehicular systems (Liu and Du, 2018; Savchenko and Poddubko, 2018).

EEG, a direct measure of brain's electrical activity, can provide robust measures of cognitive state variations while driving, including levels of drowsiness (Liang et al., 2006; Wei et al., 2018), fatigue (Liu et al., 2015; Fu et al., 2016; Hung et al., 2017), and workload (Dasari et al., 2017; Zander et al., 2017). EEG has high temporal resolution and is a direct measure of brain activity. However, data collection (e.g., longer setup time) and processing in real-world setting (e.g., movement artifacts) can be quite challenging to implement into a real-world driving research protocol (Popescu et al., 2008). At the same time, there have been innovative technological and analytical developments in EEG acquisition. For instance, efforts in brain computer interface applications have utilized a single electrode to classify relaxed vs. cognitive workload phases (Shirazi et al., 2014) and monitor fatigue levels (Morales et al., 2017). Recent work extracted features from a 6-channel EEG dataset to classify mental tasks with up to 83% accuracy rate (Neshov et al., 2018). Other recent work has reported detection algorithms that can be used to accurately classify fatigue (Li et al., 2017; Gao et al., 2018). In other work, a novel approach to detect drowsiness has been proposed which reduces calibration time for a new user by 90% using a hierarchical clustering method, which accounts for inter- and intra-subject variability (Wei et al., 2018). Automatic drowsiness detection algorithms based on only a single target channel can allow real-time neural assessments of cognitive states (Belakhdar et al., 2018). With increasing advancements in sensor development and data processing, we hold an optimistic view of adopting EEG-based measures in driving research, albeit after considerable validation (Kosiachenko and Si, 2017; Krol et al., 2017; Zander et al., 2017; Byrom et al., 2018). Recent work has also shown the applicability of specific ERP components (such as the P300), some of which show good psychometric properties (e.g., Cassidy et al., 2012), and can be adopted to brain-computer interfaces (Piña-Ramírez et al., 2018). Future work and reliable replication of studies are required to ensure EEG and ERPs could be assimilated in human-machine automation interface.

Traditional fNIRS has lower temporal resolution and may additionally be difficult to collect in applied settings. However, recently, mobile-friendly systems have been developed and used in applied domains (von Lüthmann et al., 2015) including exercise physiology (Byun et al., 2014), clinical monitoring (Kassab et al., 2018), and infant developmental research (Quaresima et al., 2012). Importantly, these advancements mean that fNIRS measurements can be performed in naturalistic environments without considerable restraint. As the development of ultra-portable systems grows (e.g., battery powered mobile systems, McKendrick et al., 2016), fNIRS will likely form a novel complement to the many other physiological measures discussed here, in part because of its unique capability to image neural hemodynamics and reveal changes in brain activity with improved spatial resolution compared to other portable and non-invasive neurophysiological methods (e.g., EEG; Ahn and Jun, 2017). For instance, a recent study adopted a wearable fNIRS system (with sensors placed on a baseball cap making it less intrusive) to measure cognitive distraction while driving

(Le et al., 2018). Thus, while these methods are still in their infancy compared to many of the other methods discussed here, the ability to reveal neural mechanisms of cognitive states in real-world domains such as driving is promising.

Similar to fNIRS, thermal imaging also shows some early promise. It is a non-contact technology that has high relevance in applied settings, including driving (Lee et al., 2018). For example, recent work has shown the validity of thermal imaging in indexing cognitive load. In these studies, changes in nasal and forehead temperatures were observed as a function of task difficulty in a non-driving context (Abdelrahman et al., 2017; Marinescu et al., 2018). However, research in real-world settings is currently limited. Existing preliminary work has focused primarily on understanding the sensitivity of this measure in well-controlled environments. Future work will help qualify the utility and validity of thermal imaging in real-world conditions.

On the other hand, several measures, despite clear utility in a lab environment, may be currently of less use in real-world settings. For example, pupillometry in well-controlled lab settings can provide helpful information in interpreting user state (e.g., Pflieger et al., 2016; Cegovnik et al., 2018). Moreover, with the development of desktop-mounted eye trackers, pupil dilation and constriction can be measured non-invasively and remotely with high spatial and temporal resolution. In lab settings, where features such as luminance can be controlled and measured, recent work has shown success in using pupillometry to examine mental workload in an unconstrained setting (e.g., Lego construction; Bækgaard et al., 2019). In driving, some researchers have suggested that pupil-based measurements are highly relevant for assessment of drowsiness (Maccora et al., 2018). However, detection of pupil diameter in real-world settings with rapidly changing and uncontrollable variations in luminance is a critical confounding factor in the utility of pupillometry in driving (Kassem et al., 2017).

Similarly, EMG can be utilized in lab settings to understand psychological processes. For example, EMG in combination with other psychophysiological measures was recently utilized in detecting fatigue in drivers (Fu et al., 2016; Ma et al., 2018). Preliminary research has also proposed the use of EMG to detect drowsiness (Artanto et al., 2017) and real-time monitoring of muscle activity (Mazzetta et al., 2018). However, in applied settings such as driving, EMG may have only low utility, in part because the necessary motor activity needed to engage in the task (e.g., turning the steering wheel and actuation of break) can cause uncontrolled changes in muscle activity that can be confounded with the psychological variance in EMG, which is an order of magnitude smaller than these artifacts.

At the same time, ongoing methodological developments are resulting in more efficient systems, improved signal-to-noise ratio, and improved signal-processing methods, all of which culminate in rapidly improving the reliability and validity of acquisition across these multiple methodologies. Some attempts to assess cognitive states using multiple methods have been integrated in non-driving domains (ElKomy et al., 2017; Ko et al., 2017; Moghaddam and Lowe, 2019) and multi-method work in real-world driving contexts are already underway (Fu et al., 2016; Brouwer et al., 2017; Zander et al., 2017; Aricò et al., 2018;

Belakhdar et al., 2018; Haouij et al., 2018; Paredes et al., 2018; Rastgoo et al., 2018).

Taken together, we have reviewed a growing body of empirical evidence suggesting that physiological measures can be used to sense and assess changes in the cognitive states of motorists during real-world driving. Through this selective review, we believe that the strengths and limitations of adopting physiological measures in driving can clearly extend to other domains such as the use of aircraft, trains, and ships. Furthermore, we see growing promise for the application of covert monitoring methods like those reviewed above with the increasing rise in semi-automated technology, where motorists will become less directly involved in the driving process. As such, the development of intelligent driving assistance systems will need to utilize non-behavior-based measures to index covert cognitive states of a motorist in the absence of any overt behavior. The physiological measures reviewed above have the potential to detect sub-optimal arousal levels associated with risky driving behavior and inform state detection-workload managers and “aware” systems to trigger warnings

or intervene, resulting in a closed-loop system in the absence of any overt-driving behaviors. Before we reach such a future however, the field needs to adopt rigorous standards for the use of psychophysiological measurement in real-world settings. We hope to see a future of increased collaboration and integration of basic psychophysiology, human factors, and traffic safety research. Such integration is necessary to advance the development of effective human-machine driving interfaces and driver support systems, with the ultimate goal of improving traffic safety.

## AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

## FUNDING

Support for this paper was provided by a grant from AAA Foundation for Traffic Safety.

## REFERENCES

- AAA Foundation for Traffic Safety (2018). *2017 Traffic Safety Culture Index*. Washington, DC: AAA Foundation for Traffic Safety.
- Abdelrahman, Y., Velloso, E., Dingler, T., Schmidt, A., and Vetere, F. (2017). “Cognitive heat: exploring the usage of thermal imaging to unobtrusively estimate cognitive load,” in *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, (New York, NY). doi: 10.1145/3130898
- Abe, E., Fujiwara, K., Hiraoka, T., Yamakawa, T., and Kano, M. (2016). Development of drowsiness detection method by integrating heart rate variability analysis and multivariate statistical process control. *SICE J. Control Meas. Syst. Integr.* 9, 10–17. doi: 10.9746/jcmsi.9.10
- Affanni, A., Bernardini, R., Piras, A., Rinaldo, R., and Zontone, P. (2018). Driver's stress detection using skin potential response signals. *Measurement* 122, 264–274. doi: 10.1016/j.measurement.2018.03.040
- Ahlström, C., Gink Lövgren, M., Nilsson, M., Dukic Willstrand, T., and Anund, A. (2018). The effect of an active steering system on city bus drivers muscle activity. *Int. J. Occup. Saf. Ergonom.* 2018, 1–9. doi: 10.1080/10803548.2018.1445465
- Ahn, S., and Jun, S. C. (2017). Multi-modal integration of EEG-fNIRS for brain-computer interfaces—current limitations and future directions. *Front. Hum. Neurosci.* 11:503. doi: 10.3389/fnhum.2017.00503
- Aidman, E., Chadunow, C., Johnson, K., and Reece, J. (2015). Real-time driver drowsiness feedback improves driver alertness and self-reported driving performance. *Accid. Anal. Prev.* 81, 8–13. doi: 10.1016/j.aap.2015.03.041
- Angelidis, A., van der Does, W., Schakel, L., and Putman, P. (2016). Frontal EEG theta/beta ratio as an electrophysiological marker for attentional control and its test-retest reliability. *Biol. Psychol.* 121, 49–52. doi: 10.1016/j.biopsycho.2016.09.008
- Aricò, P., Borghini, G., Di Flumeri, G., Sciaraffa, N., and Babiloni, F. (2018). Passive BCI beyond the lab: current trends and future directions. *Physiol. Meas.* 39:08TR02. doi: 10.1088/1361-6579/aad57e
- Arnau, S., Möckel, T., Rinkenauer, G., and Wascher, E. (2017). The interconnection of mental fatigue and aging: an EEG study. *Int. J. Psychophysiol.* 117, 17–25. doi: 10.1016/j.ijpsycho.2017.04.003
- Artanto, D., Sulistyanto, M. P., Pranowo, I. D., and Pramesta, E. E. (2017). “Drowsiness detection system based on eye-closure using a low-cost EMG and ESP8266,” in *Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2017 2nd International Conferences* (Yogyakarta), 235–238. doi: 10.1109/ICITISEE.2017.8285502
- Aston-Jones, G., and Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* 28, 403–450. doi: 10.1146/annurev.neuro.28.061604.135709
- Bækgaard, P., Jalaliniya, S., and Hansen, J. P. (2019). Pupillary measurement during an assembly task. *Appl. Ergon.* 75, 99–107. doi: 10.1016/j.apergo.2018.09.004
- Balasubramanian, V., and Adalarasu, K. (2007). EMG-based analysis of change in muscle activity during simulated driving. *J. Bodywork Mov. Ther.* 11, 151–158. doi: 10.1016/j.jbmt.2006.12.005
- Baldwin, C. L., Roberts, D. M., Barragan, D., Lee, J. D., Lerner, N., and Higgins, J. S. (2017). Detecting and quantifying mind wandering during simulated driving. *Front. Human Neurosci.* 11:406. doi: 10.3389/fnhum.2017.00406
- Balters, S., Murnane, E. L., Landay, J. A., and Paredes, P. E. (2018). “Breath booster!: exploring in-car, fast-paced breathing interventions to enhance driver arousal state,” in *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare* (New York, NY). doi: 10.1145/3240925.3240939
- Balters, S., and Steinert, M. (2017). Capturing emotion reactivity through physiology measurement as a foundation for affective engineering in engineering design science and engineering practices. *J. Intell. Manuf.* 28, 1585–1607. doi: 10.1007/s10845-015-1145-2
- Baltodano, S., Garcia-Mancilla, J., and Ju, W. (2018). “Eliciting driver stress using naturalistic driving scenarios on real roads,” in *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Toronto, ON). doi: 10.1145/3239060.3239090
- Barnard, M. P., and Chapman, P. (2018). The effects of instruction and environmental demand on state anxiety, driving performance and autonomic activity: are ego-threatening manipulations effective? *Transp. Res. Part F Traffic Psychol. Behav.* 55, 123–135. doi: 10.1016/j.trf.2018.02.040
- Belakhdar, I., Kaaniche, W., Djemal, R., and Ouni, B. (2018). Single-channel-based automatic drowsiness detection architecture with a reduced number of EEG features. *Microprocess. Microsyst.* 58, 13–23. doi: 10.1016/j.micpro.2018.02.004
- Belyusar, D., Mehler, B., Solovey, E., and Reimer, B. (2015). Impact of repeated exposure to a multilevel working memory task on physiological arousal and driving performance. *Transp. Res. Rec. J. Transp. Res. Board* 2518, 46–53. doi: 10.3141/2518-06
- Berntson, G. G., Quigley, K. S., and Lozano, D. (2007). Cardiovascular psychophysiology. *Handb. Psychophysiol.* 3, 182–210. doi: 10.1017/CBO9780511546396.008
- Billman, G. E. (2013). The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance. *Front. Physiol.* 4:26. doi: 10.3389/fphys.2013.00026

- Biondi, F. N., Lohani, M., Hopman, R., Mills, S., Cooper, J. M., and Strayer, D. L. (2018). "80 MPH and out-of-the-loop: effects of real-world semi-automated driving on driver workload and arousal," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Sage, CA; Los Angeles, CA: SAGE Publications).
- Blöcher, T., Schneider, J., Schinle, M., and Stork, W. (2017). "An online PPGI approach for camera based heart rate monitoring using beat-to-beat detection," in *Sensors Applications Symposium (SAS), 2017 IEEE*. doi: 10.1109/SAS.2017.7894052
- Borghini, G., Aricò, P., Di Flumeri, G., Salinari, S., Colosimo, A., Bonelli, S., et al. (2015). "Avionic technology testing by using a cognitive neurometric index: a study with professional helicopter pilots," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE* (Milan). doi: 10.1109/EMBC.2015.7319804
- Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., and Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev.* 44, 58–75. doi: 10.1016/j.neubiorev.2012.10.003
- Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., et al. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology* 49, 1017–1034. doi: 10.1111/j.1469-8986.2012.01384.x
- Broadhurst, P. L. (1959). The interaction of task difficulty and motivation: the Yerkes Dodson Law revived. *Acta Psychol.* 16, 321–338. doi: 10.1016/0001-6918(59)90105-2
- Brookhuis, K. A., and de Waard, D. (1993). The use of psychophysiology to assess driver status. *Ergonomics* 36, 1099–1110. doi: 10.1080/00140139308967981
- Brookhuis, K. A., and de Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. *Accid. Anal. Prev.* 42, 898–903. doi: 10.1016/j.aap.2009.06.001
- Brouwer, A., Snelting, A., Jaswa, M., Flascher, O., Krol, L., and Zander, T. (2017). "Physiological effects of adaptive cruise control behaviour in real driving," in *Proceedings of the 2017 ACM Workshop on An Application-oriented Approach to BCI out of the Laboratory* (Limassol). doi: 10.1145/3038439.3038441
- Byrom, B., McCarthy, M., Schueler, P., and Muehlhausen, W. (2018). Brain monitoring devices in neuroscience clinical research: the potential of remote monitoring using sensors, wearables, and mobile devices. *Clin. Pharmacol. Ther.* 104, 59–71. doi: 10.1002/cpt.1077
- Byun, K., Hyodo, K., Suwabe, K., Ochi, G., Sakairi, Y., Kato, M., et al. (2014). Positive effect of acute mild exercise on executive function via arousal-related prefrontal activations: an fNIRS study. *Neuroimage* 98, 336–345. doi: 10.1016/j.neuroimage.2014.04.067
- Cacioppo, J. T., and Tassinari, L. G. (1990). Inferring psychological significance from physiological signals. *Am. Psychol.* 45:16. doi: 10.1037/0003-066X.45.1.16
- Cacioppo, J. T., Tassinari, L. G., and Berntson, G. G. (2007). "Psychophysiological science: interdisciplinary approaches to classic questions about the mind," in *Handbook of Psychophysiology*, eds J. T. Cacioppo, L. G. Tassinari, and G. G. Berntson (New York, NY: Cambridge University Press), 1–16. doi: 10.1017/CBO9780511546396.001
- Cassidy, S. M., Robertson, I. H., and O'Connell, R. G. (2012). Retest reliability of event-related potentials: evidence from a variety of paradigms. *Psychophysiology* 49, 659–664. doi: 10.1111/j.1469-8986.2011.01349.x
- Causse, M., Chua, Z., Peysakhovich, V., Campo, N., and Matton, N. (2017). Mental workload and neural efficiency quantified in the prefrontal cortex using fNIRS. *Sci. Rep.* 7:5222. doi: 10.1038/s41598-017-05378-x
- Cavanagh, J. F., and Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends Cogn. Sci.* 18, 414–421. doi: 10.1016/j.tics.2014.04.012
- Cegovnik, T., Stojmenova, K., Jakus, G., and Sodnik, J. (2018). An analysis of the suitability of a low-cost eye tracker for assessing the cognitive load of drivers. *Appl. Ergon.* 68, 1–11. doi: 10.1016/j.apergo.2017.10.011
- Chen, L. L., Zhao, Y., Ye, P. F., Zhang, J., and Zou, J. Z. (2017). Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers. *Expert Syst. Appl.* 85, 279–291. doi: 10.1016/j.eswa.2017.01.040
- Chen, L. L., Zhao, Y., Zhang, J., and Zou, J. Z. (2015). Automatic detection of alertness/drowsiness from physiological signals using wavelet-based nonlinear features and machine learning. *Expert Systems Appl.* 42, 7344–7355. doi: 10.1016/j.eswa.2015.05.028
- Cheng, S. Y., Park, S., and Trivedi, M. M. (2007). Multi-spectral and multi-perspective video arrays for driver body tracking and activity analysis. *Comput. Vis. Image Underst.* 106, 245–257. doi: 10.1016/j.cviu.2006.08.010
- Cohen, M. X. (2011). It's about time. *Front. Human Neurosci.* 5:2. doi: 10.3389/fnhum.2011.00002
- Correa, A. G., Orosco, L., and Laciari, E. (2014). Automatic detection of drowsiness in EEG records based on multimodal analysis. *Med. Eng. Phys.* 36, 244–249. doi: 10.1016/j.medengphy.2013.07.011
- Costa, N., Simões, P., Costa, S., and Arezes, P. (2017). "Driving workload indicators: the case of senior drivers," in *International Conference on Applied Human Factors and Ergonomics* (Cham: Springer), 604–615.
- Dasari, D., Shou, G., and Ding, L. (2017). ICA-Derived EEG correlates to mental fatigue, effort, and workload in a realistically simulated air traffic control task. *Front. Neurosci.* 11:297. doi: 10.3389/fnins.2017.00297
- Dawson, M. E., Schell, A. M., and Filion, D. L. (2007). The electrodermal system. *Handb. Psychophysiol.* 2, 200–223. doi: 10.1017/CBO9780511546396.007
- de Waard, D. (1996). *The Measurement of Drivers' Mental Workload*. Netherlands: Groningen University, Traffic Research Center.
- de Waard, D., and Lewis-Evans, B. (2014). Self-report scales alone cannot capture mental workload. *Cogn. Technol. Work* 16, 303–305. doi: 10.1007/s10111-014-0277-z
- Demberg, V. (2013). "Pupillometry: the index of cognitive activity in a dual-task study," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 35. Available online at: <https://escholarship.org/uc/item/4vfv5w6bn#author>
- Drews, F. A., Strayer, D. L., Uchino, B., and Smith, T. W. (2012). "The road to road rage," in *Vision in Vehicles X*, eds A. G. Gale, S. P. Taylor, and C. Castro (Elsevier), 346–352.
- Duchowski, A. T., Krejtz, K., Krejtz, I., Biele, C., Niedzielska, A., Kiefer, P., et al. (2018). "The index of pupillary activity: measuring cognitive load vis-à-vis task difficulty with pupil oscillation," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal, QC). doi: 10.1145/3173574.3173856
- ElKomy, M., Abdelrahman, Y., Funk, M., Dingler, T., Schmidt, A., and Abdennadher, S. (2017). "ABBAS: an adaptive bio-sensors based assistive system," *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '17* (Denver, CO; New York, NY: ACM), 2543–2550. doi: 10.1145/3027063.3053179
- Engström, J., Johansson, E., and Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transp. Res. Part F Traffic Psychol. Behav.* 8, 97–120. doi: 10.1016/j.trf.2005.04.012
- Eoh, H. J., Chung, M. K., and Kim, S. H. (2005). Electroencephalographic study of drowsiness in simulated driving with sleep deprivation. *Int. J. Indus. Ergon.* 35, 307–320. doi: 10.1016/j.ergon.2004.09.006
- Fabiani, M., Gratton, G., and Federmeier, K. D. (2007). "Event-related brain potentials: methods, theory, and applications," in *Handbook of Psychophysiology*, eds J. T. Cacioppo, L. G. Tassinari, and G. G. Berntson (New York, NY: Cambridge University Press), 85–119. doi: 10.1017/CBO9780511546396.004
- Fairclough, S. H., Karan, A. J., and Gilleade, K. (2015). "Classification accuracy from the perspective of the user: real-time interaction with physiological computing," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul), 3029–3038. doi: 10.1145/2702123.2702454
- Fallahi, M., Motamedzade, M., Heidarimoghdam, R., Soltanian, A. R., and Miyake, S. (2016). Effects of mental workload on physiological and subjective responses during traffic density monitoring: a field study. *Appl. Ergon.* 52, 95–103. doi: 10.1016/j.apergo.2015.07.009
- Faure, V., Lobjois, R., and Benguigui, N. (2016). The effects of driving environment complexity and dual tasking on drivers' mental workload and eye blink behavior. *Transp. Res. Part F Traffic Psychol. Behav.* 40, 78–90. doi: 10.1016/j.trf.2016.04.007
- Forsman, P. M., Vila, B. J., Short, R. A., Mott, C. G., and van Dongen, H. P. A. (2013). Efficient driver drowsiness detection at moderate levels of drowsiness. *Accid. Anal. Prev.* 50, 341–350. doi: 10.1016/j.aap.2012.05.005
- Fowles, D. C. (1986). The eccrine system and electrodermal activity. *Psychophysiology* 1, 51–96. doi: 10.4236/ojpsych.2013.31007



- Fridlund, A. J., and Cacioppo, J. T. (1986). Guidelines for human electromyographic research. *Psychophysiology* 23, 567–589. doi: 10.1111/j.1469-8986.1986.tb00676.x
- Fu, R., and Wang, H. (2014). Detection of driving fatigue by using noncontact EMG and ECG signals measurement system. *Int. J. Neural Syst.* 24:1450006. doi: 10.1142/S0129065714500063
- Fu, R., Wang, H., and Zhao, W. (2016). Dynamic driver fatigue detection using hidden Markov model in real driving condition. *Expert Syst. Appl.* 63, 397–411. doi: 10.1016/j.eswa.2016.06.042
- Gade, R., and Moeslund, T. B. (2014). Thermal cameras and applications: a survey. *Mach. Vis. Appl.* 25, 245–262. doi: 10.1007/s00138-013-0570-5
- Gagl, B., Hawelka, S., and Hutzler, F. (2011). Systematic influence of gaze position on pupil size measurement: analysis and correction. *Behav. Res. Methods* 43, 1171–1181. doi: 10.3758/s13428-011-0109-5
- Gao, Z., Li, S., Cai, Q., Dang, W., Yang, Y., Mu, C., et al. (2018). Relative wavelet entropy complex network for improving EEG-based fatigue driving classification. *IEEE Trans. Instrum. Meas.* 99, 1–7. doi: 10.1109/TIM.2018.2865842
- Gateau, T., Durantin, G., Lancelot, F., Scannella, S., and Dehais, F. (2015). Real-time state estimation in a flight simulator using fNIRS. *PLoS ONE* 10:e0121279. doi: 10.1371/journal.pone.0121279
- Getzmann, S., Arnau, S., Karthaus, M., Reiser, J. E., and Wascher, E. (2018). Age-related differences in pro-active driving behaviour revealed by EEG measures. *Front. Human Neurosci.* 12:321. doi: 10.3389/fnhum.2018.00321
- Gilzenrat, M. S. (2006). *The Role of the Locus Coeruleus in Cognitive Control: Pupillometric Measures and Computational Mechanisms*. 0724–0724. Ph.D. thesis, Princeton University.
- Gratton, G., and Fabiani, M. (2001). Shedding light on brain function: the event-related optical signal. *Trends Cogn. Sci.* 5, 357–363. doi: 10.1016/S1364-6613(00)01701-0
- Gratton, G., and Fabiani, M. (2003). The event-related optical signal (EROS) in visual cortex: replicability, consistency, localization, and resolution. *Psychophysiology* 40, 561–571. doi: 10.1111/1469-8986.00058
- Grinvald, A., Lieke, E., Frostig, R. D., Gilbert, C. D., and Wiesel, T. N. (1986). Functional architecture of cortex revealed by optical imaging of intrinsic signals. *Nature* 324, 361–364. doi: 10.1038/324361a0
- Grossman, P. (1992). Respiratory and cardiac rhythms as windows to central and autonomic biobehavioral regulation: selection of window frames, keeping the panes clean and viewing the neural topography. *Biol. Psychol.* 34, 131–161. doi: 10.1016/0301-0511(92)90013-K
- Guoping, S., and Zhang, K. (2009). An ERP study of effects of driving fatigue on auditory attention. *Psychol. Sci.* 32, 517–520. Available online at: <https://psycnet.apa.org/record/2009-12822-001>
- Haouij, N. E., Poggi, J.-M., Sevestre-Ghalila, S., Ghozi, R., and Jaïdane, M. (2018). “AffectiveROAD system and database to assess driver’s attention,” in *Proceedings of the 33rd Annual ACM Symposium on Applied Computing (Pau)*. doi: 10.1145/3167132.3167395
- Healey, J., Seger, J., and Picard, R. (1999). Quantifying driver stress: developing a system for collecting and processing bio-metric signals in natural situations. *Biomed. Sci. Instrum.* 35, 193–198.
- Hebb, D. O. (1955). Drives and the CNS (conceptual nervous system). *Psychol. Rev.* 62:243. doi: 10.1037/h0041823
- Heine, T., Lenis, G., Reichensperger, P., Beran, T., Doessel, O., and Deml, B. (2017). Electrocardiographic features for the measurement of drivers’ mental workload. *Appl. Ergon.* 61, 31–43. doi: 10.1016/j.apergo.2016.12.015
- Herff, C., Putze, F., and Schultz, T. (2017). “Evaluating fNIR-based workload discrimination in a realistic driving scenario,” in *The First Biannual Neuroadaptive Technology, Conference (Berlin)*.
- Hess, E. H., and Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science* 143, 1190–1192. doi: 10.1126/science.143.3611.1190
- Hidalgo-Muñoz, A. R., Béquet, A. J., Astier-Juvenon, M., Pépin, G., Fort, A., Jallais, C., et al. (2018). Respiration and heart rate modulation due to competing cognitive tasks while driving. *Front. Human Neurosci.* 12:525. doi: 10.3389/fnhum.2018.00525
- Hirao, A., Kato, K., Kitazaki, S., and Yamazaki, N. (2007). *Evaluations of Physical Fatigue During Long-Term Driving with a New Driving Posture*. SAE Technical Paper, SAE World Congress & Exhibition (No. 2007-01-0348). doi: 10.4271/2007-01-0348
- Horat, S. K., Herrmann, F. R., Favre, G., Terzis, J., Debatisse, D., Merlo, M. C., et al. (2016). Assessment of mental workload: a new electrophysiological method based on intra-block averaging of ERP amplitudes. *Neuropsychologia* 82, 11–17. doi: 10.1016/j.neuropsychologia.2015.12.013
- Hostens, I., and Ramon, H. (2005). Assessment of muscle fatigue in low level monotonous task performance during car driving. *J. Electromyogr. Kinesiol.* 15, 266–274. doi: 10.1016/j.jelekin.2004.08.002
- Hung, Y. C., Wang, Y. K., Prasad, M., and Lin, C. T. (2017). “Brain dynamic states analysis based on 3D convolutional neural network,” in *Systems, Man, and Cybernetics (SMC), 2017 IEEE International Conference (Banff, AB)*, 222–227. doi: 10.1109/SMC.2017.8122606
- Jacob, S., Bukhari, S. S., Ishimaru, S., and Dengel, A. (2018). “Gaze-based interest detection on newspaper articles,” in *Proceedings of the 7th Workshop on Pervasive Eye Tracking and Mobile Eye-Based Interaction (Warsaw)*, 4. doi: 10.1145/3208031.3208034
- Jagannath, M., and Balasubramanian, V. (2014). Assessment of early onset of driver fatigue using multimodal fatigue measures in a static simulator. *Appl. Ergon.* 45, 1140–1147. doi: 10.1016/j.apergo.2014.02.001
- Janisse, M. P. (1977). *Pupillometry: The Psychology of the Pupillary Response*. Halsted Press.
- Jasper, P., Sibley, C., and Coyne, J. (2016). “Using heart rate variability to assess operator mental workload in a command and control simulation of multiple unmanned aerial vehicles,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 60 (Sage CA: Los Angeles, CA: SAGE Publications), 1125–1129. doi: 10.1177/1541931213601264
- Jennings, J. R., and Allen, B. (2016). “Methodology,” Chapter. in *Handbook of Psychophysiology, 4th Edn.* eds J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson, 583–611. *Cambridge Handbooks in Psychology*. Cambridge: Cambridge University Press. doi: 10.1017/9781107415782.027
- Jennings, J. R., Bberg, W. K., Hutcheson, J. S., Obrist, P., Porges, S., and Turpin, G. (1981). Publication guidelines for heart rate studies in man. *Psychophysiology* 18, 226–231. doi: 10.1111/j.1469-8986.1981.tb03023.x
- Jensen, O., and Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front. Human Neurosci.* 4:186. doi: 10.3389/fnhum.2010.00186
- Jobis, F. F. (1977). Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. *Science* 198, 1264–1267. doi: 10.1126/science.929199
- Johansson, G., Evans, G. W., Cederström, C., Rydstedt, L. W., Fuller-Rowell, T., and Ong, A. D. (2012). The effects of urban bus driving on blood pressure and musculoskeletal problems: a quasi-experimental study. *Psychosom. Med.* 74, 89–92. doi: 10.1097/PSY.0b013e31823ba88f
- Jonsson, S., and Jonsson, B. (1975). Function of the muscles of the upper limb in car driving. *Ergonomics* 18, 375–388. doi: 10.1080/00140137508931471
- Kahana, M. J. (2006). The cognitive correlates of human brain oscillations. *J. Neurosci.* 26, 1669–1672. doi: 10.1523/JNEUROSCI.7373-05c.2006
- Kahneman, D., and Beatty, J. (1966). Pupil diameter and load on memory. *Science* 154, 1583–1585. doi: 10.1126/science.154.3756.1583
- Kajiwar, S. (2014). Evaluation of driver’s mental workload by facial temperature and electrodermal activity under simulated driving conditions. *Int. J. Automot. Technol.* 15, 65–70.
- Kassab, A., Le Lan, J., Tremblay, J., Vannasing, P., Dehbozorgi, M., Pouliot, P., et al. (2018). Multichannel wearable fNIRS-EEG system for long-term clinical monitoring. *Human Brain Mapp.* 39, 7–23. doi: 10.1002/hbm.23849
- Kassem, K., Salah, J., Abdrabou, Y., Morsy, M., El-Gendy, R., Abdelrahman, Y., et al. (2017). “DiVA: exploring the usage of pupil diameter to elicit valence and arousal,” in *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia (Stuttgart)*. doi: 10.1145/3152832.3152836
- Käthner, I., Wriessnegger, S. C., Müller-Putz, G. R., Kübler, A., and Halder, S. (2014). Effects of mental workload and fatigue on the P300, alpha and theta band power during operation of an ERP (P300) brain–computer interface. *Biol. Psychol.* 102, 118–129. doi: 10.1016/j.biopsycho.2014.07.014
- Kato, T., Fujii, T., and Tanimoto, M. (2004). “Detection of driver’s posture in the car by using far infrared camera,” in *Intelligent Vehicles Symposium, 2004 (Parma)*. doi: 10.1109/IVS.2004.1336406



- Ko, L. W., Komarov, O., Hairston, W. D., Jung, T. P., and Lin, C. T. (2017). Sustained attention in real classroom settings: an eeg study. *Front. Human Neurosci.* 11:388. doi: 10.3389/fnhum.2017.00388
- Kolli, A., Fasih, A., Al Machot, F., and Kyamakya, K. (2011). "Non-intrusive car driver's emotion recognition using thermal camera," in *Nonlinear Dynamics and Synchronization (INDS) & 16th Int'l Symposium on Theoretical Electrical Engineering* (Klagenfurt). doi: 10.1109/INDS.2011.6024802
- Kosiachenko, E., and Si, D. (2017). Classification of brain activity using brainwaves recorded by consumer-grade EEG devices. *Adv. Mass Data*, 14.
- Krol, L. R., Zander, T. O., Jaswa, M., Flascher, O., Snelting, A., and Brouwer, A.-M. (2017). "Online-capable cleaning of highly artefactual EEG data recorded during real driving," in *Proceedings of the 7th Graz Brain-Computer Interface Conference*. doi: 10.3217/978-3-85125-533-1-47
- Kurosawa, K., Takekawa, N., Sano, T., Miyamoto, S., Yasushi, M., and Hashimoto, H. (2017). "Drowsiness prediction system for vehicle using capacity coupled electrode type non-invasive ECG measurement," in *System Integration (SII), 2017 IEEE/SICE International Symposium*. doi: 10.1109/SII.2017.8279230
- Laborde, S., Mosley, E., and Thayer, J. F. (2017). Heart rate variability and cardiac vagal tone in psychophysiological research—recommendations for experiment planning, data analysis, and data reporting. *Front. Psychol.* 8:213. doi: 10.3389/fpsyg.2017.00213
- Le, A. S., Aoki, H., Murase, F., and Ishida, K. (2018). A novel method for classifying driver cognitive distraction under naturalistic conditions with information from near-infrared spectroscopy. *Front. Human Neurosci.* 12:431. doi: 10.3389/fnhum.2018.00431
- Lee, B. G., Chong, T. W., Lee, B. L., Park, H. J., Kim, Y. N., and Kim, B. (2017a). Wearable mobile-based emotional response-monitoring system for drivers. *IEEE Trans. Human Mach. Syst.* 47, 636–649. doi: 10.1109/THMS.2017.2658442
- Lee, D. S., Chong, T. W., and Lee, B. G. (2017b). Stress events detection of driver by wearable glove system. *IEEE Sens. J.* 17, 194–204. doi: 10.1109/JSEN.2016.2625323
- Lee, H. B., Kim, J. S., Kim, Y. S., Baek, H. J., Ryu, M. S., and Park, K. S. (2007). "The relationship between HRV parameters and stressful driving situation in the real road," in *6th International Special Topic Conference on Information Technology Applications in Biomedicine, 2007* (Tokyo). doi: 10.1109/ITAB.2007.4407380
- Lee, K. W., Yoon, H. S., Song, J. M., and Park, K. R. (2018). Convolutional neural network-based classification of driver's emotion during aggressive and smooth driving using multi-modal camera sensors. *Sensors* 18:957. doi: 10.3390/s1804095
- Leinonen, V., Kankaanpää, M., Vanharanta, H., Airaksinen, O., and Hänninen, O. (2005). Back and neck extensor loading and back pain provocation in urban bus drivers with and without low back pain. *Pathophysiology* 12, 249–255. doi: 10.1016/j.pathophys.2005.09.004
- Lenneman, J. K., and Backs, R. W. (2009). Cardiac autonomic control during simulated driving with a concurrent verbal working memory task. *Human Fact.* 51, 404–418. doi: 10.1177/0018720809337716
- Lenneman, J. K., and Backs, R. W. (2010). "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*. doi: 10.1145/1969773.1969777
- Li, Z., Chen, L., Peng, J., and Wu, Y. (2017). Automatic detection of driver fatigue using driving operation information for transportation safety. *Sensors* 17:1212. doi: 10.3390/s17061212
- Li, Z., Zhang, M., Zhang, X., Dai, S., Yu, X., and Wang, Y. (2009). Assessment of cerebral oxygenation during prolonged simulated driving using near infrared spectroscopy: its implications for fatigue development. *Eur. J. Appl. Physiol.* 107, 281–287. doi: 10.1007/s00421-009-1122-6
- Liang, B., and Lin, Y. (2018). Using physiological and behavioral measurements in a picture-based road hazard perception experiment to classify risky and safe drivers. *Transp. Res. Part F Traffic Psychol. Behav.* 58, 93–105. doi: 10.1016/j.trf.2018.05.024
- Liang, S. F., Lin, C. T., Wu, R. C., Chen, Y. C., Huang, T. Y., and Jung, T. P. (2006). "Monitoring driver's alertness based on the driving performance estimation and the EEG power spectrum analysis," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference* (Shanghai). doi: 10.1109/IEMBS.2005.1615791
- Liang, W. C., Yuan, J., Sun, D. C., and Lin, M. H. (2009). Changes in physiological parameters induced by indoor simulated driving: effect of lower body exercise at mid-term break. *Sensors* 9, 6913–6933. doi: 10.3390/s90906913
- Liang, Y., Horrey, W. J., Howard, M. E., Lee, M. L., Anderson, C., Shreeve, M. S., et al. (2017). Prediction of drowsiness events in night shift workers during morning driving. *Accid. Anal. Prev.* doi: 10.1016/j.aap.2017.11.004. [Epub ahead of print].
- Lim, W. L., Sourina, O., and Wang, L. P. (2018). STEW: Simultaneous task EEG workload dataset. *IEEE Trans. Neural Syst. Rehabil. Eng.* 26, 2106–2114. doi: 10.1109/TNSRE.2018.2872924
- Lin, C. T., Wu, R. C., Liang, S. F., Chao, W. H., Chen, Y. J., and Jung, T. P. (2005). EEG-based drowsiness estimation for safety driving using independent component analysis. *IEEE Trans. Circuits Syst. I Regular Papers* 52, 2726–2738. doi: 10.1109/TCSI.2005.857555
- Littler, W. A., Honour, A. J., and Sleight, P. (1973). Direct arterial pressure and electrocardiogram during motor car driving. *Br. Med. J.* 2, 273–277. doi: 10.1136/bmj.2.5861.273
- Liu, T., Pelowski, M., Pang, C., Zhou, Y., and Cai, J. (2016). Near-infrared spectroscopy as a tool for driving research. *Ergonomics* 59, 368–379. doi: 10.1080/00140139.2015.1076057
- Liu, T., Saito, H., and Oi, M. (2012). Distinctive activation patterns under intrinsically versus extrinsically driven cognitive loads in prefrontal cortex: a near-infrared spectroscopy study using a driving video game. *Neurosci. Lett.* 506, 220–224. doi: 10.1016/j.neulet.2011.11.009
- Liu, Y., and Du, S. (2018). Psychological stress level detection based on electrodermal activity. *Behav. Brain Res.* 341, 50–53. doi: 10.1016/j.bbr.2017.12.021
- Liu, Y.-T., Lin, Y.-T., Wu, S.-L., Hsieh, T.-Y., and Lin, C.-T. (2015). "Assessment of mental fatigue: an EEG-based forecasting system for driving safety," in *2015 IEEE International Conference on Systems, Man, and Cybernetics* (Kowloon). doi: 10.1109/SMC.2015.561
- Lohani, M., and Isaacowitz, D. M. (2014). Age differences in managing response to sadness elicitors using attentional deployment, positive reappraisal and suppression. *Cogn. Emot.* 28, 678–697. doi: 10.1080/02699931.2013.853648
- Lohani, M., Payne, B. R., and Isaacowitz, D. M. (2018). Emotional coherence in early and later adulthood during sadness reactivity and regulation. *Emotion* 18, 789–804. doi: 10.1037/emo0000345
- Luck, S. J., and Kappenman, E. S. (2012). *Oxford Handbook of ERP Components*. New York, NY: Oxford University Press.
- Luijckx, R., Hermens, H. J., Bodar, L., Vossen, C. J., van Os, J., and Lousberg, R. (2014). Experimentally induced stress validated by EMG activity. *PLoS ONE* 9:e95215. doi: 10.1371/journal.pone.0095215
- Lundberg, U., Kadefors, R., Melin, B., Palmerud, G., Hassmén, P., Engström, M., et al. (1994). Psychophysiological stress and EMG activity of the trapezius muscle. *Int. J. Behav. Med.* 354–370. doi: 10.1207/s15327558ijbm0104\_5
- Luque-Casado, A., Perales, J. C., Cárdenas, D., and Sanabria, D. (2016). Heart rate variability and cognitive processing: the autonomic response to task demands. *Biol. Psychol.* 113, 83–90. doi: 10.1016/j.biopsycho.2015.11.013
- Lyman, J. M., McGwin, G. Jr., and Sims, R. V. (2001). Factors related to driving difficulty and habits in older drivers. *Accid. Anal. Prev.* 33, 413–421. doi: 10.1016/S0001-4575(00)00055-5
- Ma, Z., Yao, S., and Li, J. (2018). "A study of integrated driver fatigue judging algorithm based on principal component analysis," in *MATEC Web of Conferences, (EDP Sciences)* (Nanjing). doi: 10.1051/mateconf/201817302011
- Maccora, J., Manousakis, J. E., and Anderson, C. (2018). Pupillary instability as an accurate, objective marker of alertness failure and performance impairment. *J. Sleep Res.* e12739. doi: 10.1111/jsr.12739
- Madan, C. R., Harrison, T., and Mathewson, K. E. (2018). Noncontact measurement of emotional and physiological changes in heart rate from a webcam. *Psychophysiology* 55:e13005. doi: 10.1111/psyp.13005
- Marinescu, A. C., Sharples, S., Ritchie, A. C., Sánchez López, T., McDowell, M., and Morvan, H. P. (2018). Physiological parameter response to variation of mental workload. *Hum. Factors* 60, 31–56. doi: 10.1177/0018720817733101
- Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024

- Marquart, G., Cabrall, C., and de Winter, J. (2015). Review of eye-related measures of drivers' mental workload. *Procedia Manuf.* 3, 2854–2861. doi: 10.1016/j.promfg.2015.07.783
- Marshall, S. P. (2002). "The index of cognitive activity: measuring cognitive workload," in *Human Factors and Power Plants, 2002. Proceedings of the 2002 IEEE 7th Conference* (Scottsdale, AZ). doi: 10.1109/HFPP.2002.1042860
- Massar, S. A. A., Wester, A. E., Volkerts, E. R., and Kenemans, J. L. (2010). Manipulation specific effects of mental fatigue: evidence from novelty processing and simulated driving. *Psychophysiology* 47, 1119–1126. doi: 10.1111/j.1469-8986.2010.01028.x
- Mathewson, K. E., Gratton, G., Fabiani, M., Beck, D. M., and Ro, T. (2009). To see or not to see: prestimulus  $\alpha$  phase predicts visual awareness. *J. Neurosci.* 29, 2725–2732. doi: 10.1523/JNEUROSCI.3963-08.2009
- Mathews, G., Wohleber, R., Lin, J., Funke, G., and Neubauer, C. (2019). "Monitoring task fatigue in contemporary and future vehicles: a review," in *Advances in Human Factors in Simulation and Modeling. AHFE 2018. Advances in Intelligent Systems and Computing*, Vol. 780, ed D. Cassenti (Cham: Springer) doi: 10.1007/978-3-319-94223-0\_10
- Mauss, I. B., and Robinson, M. D. (2009). Measures of emotion: a review. *Cogn. Emot.* 23, 209–237. doi: 10.1080/02699930802204677
- Mazzetta, I., Gentile, P., Pessione, M., Suppa, A., Zampogna, A., Bianchini, E., et al. (2018). Stand-alone wearable system for ubiquitous real-time monitoring of muscle activation potentials. *Sensors* 18:6. doi: 10.3390/s18061748
- McKendrick, R., Parasuraman, R., Murtza, R., Formwalt, A., Baccus, W., Paczynski, M., et al. (2016). Into the wild: neuroergonomic differentiation of hand-held and augmented reality wearable displays during outdoor navigation with functional near infrared spectroscopy. *Front. Hum. Neurosci.* 10:216. doi: 10.3389/fnhum.2016.00216
- 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. *Hum. Factors* 54, 396–412. doi: 10.1177/0018720812442086
- Mehler, B., Reimer, B., Coughlin, J. F., and Dusek, J. A. (2009). The impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transport. Res. Rec.* 2138, 6–12. doi: 10.3141/2138-02
- Mehler, B., Reimer, B., and Wang, Y. (2011). "A comparison of heart rate and heart rate variability indices in distinguishing single-task driving and driving under secondary cognitive workload," in *Proceedings of the 6th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design: Driving Assessment, 2011* (Lake Tahoe, CA), 590–597. doi: 10.17077/drivingassessment.1451
- Miller, N. R., Newman, N. J., Biousse, V., and Kerrison, J. B. (2005). *Walsh and Hoyt's Clinical-Neuroophthalmology, 6th Edn.* Philadelphia, PA: Lippincott Williams & Wilkins
- Moghaddam, G. K., and Lowe, C. R. (eds.). (2019). "Ex vivo biosignatures," in *Health and Wellness Measurement Approaches for Mobile Healthcare*. (Cham: Springer), 51–104. doi: 10.1007/978-3-030-01557-2\_3
- Morad, Y., Lemberg, H., Yofe, N., and Dagan, Y. (2000). Pupillography as an objective indicator of fatigue. *Curr. Eye Res.* 21, 535–542. doi: 10.1076/0271-3683(200007)2111-ZFT535
- Morales, J. M., Diaz-Piedra, C., Rieiro, H., Roca-González, J., Romero, S., Catena, A., et al. (2017). Monitoring driver fatigue using a single-channel electroencephalographic device: a validation study by gaze-based, driving performance, and subjective data. *Accid. Anal. Prev.* 109, 62–69. doi: 10.1016/j.aap.2017.09.025
- Morris, D. M., Erno, J. M., and Pilcher, J. J. (2017). "Electrodermal response and automation trust during simulated self-driving car use," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, (Los Angeles, CA: SAGE Publications), 61, 1759–1762. doi: 10.1177/1541931213601921
- Mun, S., Whang, M., Park, S., and Park, M. (2017). Effects of mental workload on involuntary attention: a somatosensory ERP study. *Neuropsychologia* 106, 7–20. doi: 10.1016/j.neuropsychologia.2017.08.021
- Murai, K., Hayashi, Y., Okazaki, T., Stone, L. C., and Mitomo, N. (2008). "Evaluation of ship navigator's mental workload using nasal temperature and heart rate variability," in *Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference* (Singapore). doi: 10.1109/ICSMC.2008.4811503
- Nadeau, C., and Bengio, Y. (2000). "Inference for the generalization error," in *Advances in Neural Information Processing Systems*, 307–313. doi: 10.1023/A:1024068626366
- Neshov, N. N., Manolova, A. H., Draganov, I. R., Tonschev, K. T., and Boumbarov, O. L. (2018). Classification of Mental Tasks from EEG Signals Using Spectral Analysis, PCA and SVM. *Cybern. Inf. Technol.* 18, 81–92. doi: 10.2478/cait-2018-0007
- Noda, A., Miyaji, M., Wakuda, Y., Hara, Y., and Yasuma, F. (2015). Simultaneous measurement of heart rate variability and blinking duration to predict sleep onset and drowsiness in drivers. *J. Sleep Disord Ther* 4, 2167–2277. doi: 10.4172/2167-0277.1000213
- Noordzij, M. L., Dorrestijn, S. M., and van den Berg, I. A. (2017). "An idiographic study into the physiology and self-reported mental workload of learning to drive a car," in *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2016 Annual Conference* (Prague). Available online at: <https://research.utwente.nl/en/publications/an-idiographic-study-into-the-physiology-and-selfreported-mental->
- Nowara, E., Marks, T. K., Mansour, H., and Veeraraghavan, A. (2018). "SparsePPG: towards driver monitoring using camera-based vital signs estimation in near-infrared," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (Salt Lake City, UT). doi: 10.1109/CVPRW.2018.00174
- Nunez, P. L., and Srinivasan, R. (2006). *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press. doi: 10.1093/acprof:oso/9780195050387.001.0001
- O'Donnell, R. D., and Eggemeier, F. T. (1986). "Workload assessment methodology," in *Handbook of perception and human performance. Vol. 2, Cognitive Processes and Performance*, eds K. R. Boff, L. Kaufman, and J. P. Thomas (New York, NY: Wiley), 1–49.
- Or, C. K. L., and Duffy, V. G. (2007). Development of a facial skin temperature-based methodology for non-intrusive mental workload measurement. *Occupational Ergonomics* 7, 83–94. Available online at: <https://content.iospress.com/articles/occupational-ergonomics/oer00140>
- Ottaviani, C., Brosschot, J. F., Lonigro, A., Medea, B., Van Diest, I., and Thayer, J. F. (2016). Hemodynamic profiles of functional and dysfunctional forms of repetitive thinking. *Ann. Behav. Med.* 51, 261–271. doi: 10.1007/s12160-016-9851-3
- Palinko, O., Kun, A. L., Shyrokov, A., and Heeman, P. (2010). "Estimating cognitive load using remote eye tracking in a driving simulator," in *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (Austin, TX). doi: 10.1145/1743666.1743701
- Paredes, P. E., Ordonez, F., Ju, W., and Landay, J. A. (2018). "Fast & furious: detecting stress with a car steering wheel," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal, QC). doi: 10.1145/3173574.3174239
- Patel, M., Lal, S. K., Kavanagh, D., and Rossiter, P. (2011). Applying neural network analysis on heart rate variability data to assess driver fatigue. *Expert Syst. Appl.* 38, 7235–7242. doi: 10.1016/j.eswa.2010.12.028
- Perrier, J., Jongen, S., Vuurman, E., Bocca, M. L., Ramaekers, J. G., and Vermeeren, A. (2016). Driving performance and EEG fluctuations during on-the-road driving following sleep deprivation. *Biol. Psychol.* 121, 1–11. doi: 10.1016/j.biopsycho.2016.09.010
- Pfleging, B., Fekety, D. K., Schmidt, A., and Kun, A. L. (2016). "A model relating pupil diameter to mental workload and lighting conditions," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, CA). doi: 10.1145/2858036.2858117
- Piña-Ramírez, O., Valdés-Cristerna, R., Medina-Bañuelos, V., and Yañez-Suárez, O. (2018). "Chapter 7 - P300-based brain-computer interfaces," in *Smart Wheelchairs and Brain-Computer Interfaces*, ed P. Diez (Academic Press), 131–170. Available online at: <http://www.sciencedirect.com/science/article/pii/B9780128128923000078>
- Piotrowski, Z., and Szypulska, M. (2017). Classification of falling asleep states using HRV analysis. *Biocybern. Biomed. Eng.* 37, 290–301. doi: 10.1016/j.bbe.2017.02.003
- Piper, S. K., Krueger, A., Koch, S. P., Mehnert, J., Habermehl, C., Steinbrink, J., et al. (2014). A wearable multi-channel fNIRS system for brain imaging in freely moving subjects. *Neuroimage* 85, 64–71. doi: 10.1016/j.neuroimage.2013.06.062

- Piquado, T., Isaacowitz, D. M., and Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology* 47, 560–569. doi: 10.1111/j.1469-8986.2009.00947.x
- Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148. doi: 10.1016/j.clinph.2007.04.019
- Popescu, F., Blankertz, B., and Müller, K. R. (2008). Computational challenges for noninvasive brain computer interfaces. *IEEE Intell. Syst.* 23, 78–79. Available online at: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.370.8282>
- Prabhakar, G., Madhu, N., and Biswas, P. (2018). “Comparing pupil dilation, head movement, and eeg for distraction detection of drivers,” in *Proceedings of the 32nd International BCS Human Computer Interaction Conference* (Belfast, UK). doi: 10.14236/ewic/HCI2018.69
- Putman, P., Verkuil, B., Arias-Garcia, E., Pantazi, I., and van Schie, C. (2014). EEG theta/beta ratio as a potential biomarker for attentional control and resilience against deleterious effects of stress on attention. *Cogn. Affect. Behav. Neurosci.* 14, 782–791. doi: 10.3758/s13415-013-0238-7
- Quaresima, V., Bisconti, S., and Ferrari, M. (2012). A brief review on the use of functional near-infrared spectroscopy (fNIRS) for language imaging studies in human newborns and adults. *Brain Lang.* 121, 79–89. doi: 10.1016/j.bandl.2011.03.009
- Quigley, K. S., and Berntson, G. G. (1996). Autonomic interactions and chronotropic control of the heart: heart period versus heart rate. *Psychophysiology* 33, 605–611. doi: 10.1111/j.1469-8986.1996.tb02438.x
- Quintana, D. S., Elstad, M., Kaufmann, T., Brandt, C. L., Haatveit, B., Haram, M., et al. (2016). Resting-state high-frequency heart rate variability is related to respiratory frequency in individuals with severe mental illness but not healthy controls. *Sci. Rep.* 6:37212 doi: 10.1038/srep37212
- Rai, M., Maity, T., and Yadav, R. K. (2017). Thermal imaging system and its real time applications: a survey. *J. Eng. Technol.* 6, 290–303. Available online at: <http://www.joetsite.com/wp-content/uploads/2017/07/Vol.-62-24-2017.pdf>
- Rajkowski, J., Kubiak, P., and Aston-Jones, G. (1994). Locus coeruleus activity in monkey: phasic and tonic changes are associated with altered vigilance. *Brain Res. Bull.* 35, 607–616. doi: 10.1016/0361-9230(94)90175-9
- Rastgoo, M. N., Nakisa, B., Rakotonirainy, A., Chandran, V., and Tjondronegoro, D. (2018). A critical review of proactive detection of driver stress levels based on multimodal measurements. *ACM Comput. Surv.* 51:88. doi: 10.1145/3186585
- Recarte, M. A., and Nunes, L. M. (2000). Effects of verbal and spatial-imagery tasks on eye fixations while driving. *J. Exp. Psychol.* 6, 31–43. doi: 10.1037/1076-898X.6.1.31
- Reilly, J., Kelly, A., Kim, S. H., Jett, S., and Zuckerman, B. (2018). The human task-evoked pupillary response function is linear: implications for baseline response scaling in pupillometry. *Behav. Res. Methods* 1–14. doi: 10.3758/s13428-018-1134-4
- Reimer, B., Mehler, B., Coughlin, J. F., Godfrey, K. M., and Chuanzhong, T. (2009). “An on-road assessment of the impact of cognitive workload on physiological arousal in young adult drivers,” in *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, (ACM), 115–118. doi: 10.1145/1620509.1620531
- Ruscio, D., Bos, A. J., and Ciceri, M. R. (2017). Distraction or cognitive overload? using modulations of the autonomic nervous system to discriminate the possible negative effects of advanced assistance system. *Accid. Anal. Prev.* 103, 105–111. doi: 10.1016/j.aap.2017.03.023
- Sahayadhas, A., Sundaraj, K., and Murugappan, M. (2012). Detecting driver drowsiness based on sensors: a review. *Sensors* 12, 16937–16953. doi: 10.3390/s121216937
- Saus, E. R., Johnsen, B. H., Eid, J., and Thayer, J. F. (2001). Who benefits from simulator training: Personality and heart rate variability in relation to situation awareness during navigation training? *Comput. Hum. Behav.* 28, 1262–1268. doi: 10.1016/j.chb.2012.02.009
- Savchenko, V. V., and Poddubko, S. N. (2018). “Development approach to a method for monitoring of driver’s ability of resumption of control over the vehicle by on-board systems in automatic mode,” in *IOP Conference Series: Materials Science and Engineering*. doi: 10.1088/1757-899X/386/1/012007
- Saxby, D. J., Matthews, G., Hitchcock, E. M., Warm, J. S., Funke, G. J., and Gantzer, T. (2008). “Effect of active and passive fatigue on performance using a driving simulator,” in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 52, (Los Angeles, CA: Sage Publications), 1751–1755. doi: 10.1177/154193120805202113
- Schäfer, A., and Vagedes, J. (2013). How accurate is pulse rate variability as an estimate of heart rate variability? a review on studies comparing photoplethysmographic technology with an electrocardiogram. *Int. J. Cardiol.* 166, 15–29. doi: 10.1016/j.ijcard.2012.03.119
- Schier, M. A. (2000). Changes in EEG alpha power during simulated driving: a demonstration. *Int. J. Psychophysiol.* 37, 155–162. doi: 10.1016/S0167-8760(00)00079-9
- Schmidt, E., and Bullinger, A. C. (2017). Mitigating passive fatigue during monotonous drives with thermal stimuli: Insights into the effect of different stimulation durations. *Accid. Anal. Prev.* 4575, 30443–30448. doi: 10.1016/j.aap.2017.12.005
- Schmidt, E., Decke, R., Rassehofer, R., and Bullinger, A. C. (2017). Psychophysiological responses to short-term cooling during a simulated monotonous driving task. *Appl. Ergon.* 62, 9–18. doi: 10.1016/j.apergo.2017.01.017
- Schmidt, E. A., Schrauf, M., Simon, M., Buchner, A., and Kincses, W. E. (2011). The short-term effect of verbally assessing drivers’ state on vigilance indices during monotonous daytime driving. *Transport. Res. Part F* 14, 251–260. doi: 10.1016/j.trf.2011.01.005
- Schmidt, E. A., Schrauf, M., Simon, M., Fritzsche, M., Buchner, A., and Kincses, W. E. (2009). Drivers’ misjudgement of vigilance state during prolonged monotonous daytime driving. *Accid. Anal. Prev.* 41, 1087–1093. doi: 10.1016/j.aap.2009.06.007
- Schneegass, S., Pfleging, B., Broy, N., Heinrich, F., and Schmidt, A. (2013). “A data set of real world driving to assess driver workload,” in *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (Eindhoven). doi: 10.1145/2516540.2516561
- Schwalm, M., Keinath, A., and Zimmer, H., D. (2008). “Pupillometry as a method for measuring mental workload within a simulated driving task,” in *Human Factors for assistance and automation*, eds F. Flemisch, B. Lorenz, H. Oberheid, K. A. Brookhuis, and D. de Waard (Maastricht: Shaker Publishing).
- Seo, S. H., Kwak, S. H., Chung, S. C., Kim, H. S., and Min, B. C. (2017). Effect of distraction task on driving performance of experienced taxi drivers. *Asian Biomed.* 11, 315–322. doi: 10.1515/abm-2018-0003
- Shaffer, F., and Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Front. Public Health* 5:258. doi: 10.3389/fpubh.2017.00258
- Shakouri, M., Ikuma, L. H., Aghazadeh, F., and Nahmens, I. (2018). Analysis of the sensitivity of heart rate variability and subjective workload measures in a driving simulator: the case of highway work zones. *Int. J. Ind. Ergon.* 66, 136–145. doi: 10.1016/j.ergon.2018.02.015
- Shapiro, D., Jamner, L. D., Lane, J. D., Light, K. C., Myrtek, M., Sawada, Y., et al. (1996). Blood pressure publication guidelines. *Psychophysiology* 33, 1–12. doi: 10.1111/j.1469-8986.1996.tb02103.x
- Sheridan, T., Meyer, J., Roy, S., and Decker, K. (1991). *Physiological and Psychological Evaluations of Driver Fatigue During Long Term Driving*. SAE Technical Paper. doi: 10.4271/910116
- Shirazi, A. S., Hassib, M., Henze, N., Schmidt, A., and Kunze, K. (2014). “What’s on your mind?: mental task awareness using single electrode brain computer interfaces,” in *Proceedings of the 5th Augmented Human International Conference* (Kobe). doi: 10.1145/2582051.2582096
- Sibi, S., Ayaz, H., Kuhns, D. P., Sirkin, D. M., and Ju, W. (2016). “Monitoring driver cognitive load using functional near infrared spectroscopy in partially autonomous cars,” in *Intelligent Vehicles Symposium (IV), 2016 IEEE* (Gothenburg). doi: 10.1109/IVS.2016.7535420
- Simon, M., Schmidt, E. A., Kincses, W. E., Fritzsche, M., Bruns, A., Aufmuth, C., et al. (2011). EEG alpha spindle measures as indicators of driver fatigue under real traffic conditions. *Clin. Neurophysiol.* 122, 1168–1178. doi: 10.1016/j.clinph.2010.10.044
- Simonson, E., Baker, C., Burns, N., Keiper, C., Schmitt, O. H., and Stackhouse, S. (1968). Cardiovascular stress (electrocardiographic changes) produced by driving an automobile. *Am. Heart J.* 75, 125–135. doi: 10.1016/0002-8703(68)90123-3
- Siren, A., Hakamies-Blomqvist, L., and Lindeman, M. (2004). Driving cessation and health in older women. *J. Appl. Gerontol.* 23, 58–69. doi: 10.1177/0733464804263129



- Sirevaag, E. J., Kramer, A. F., Coles, M. G. H., and Donchin, E. (1989). Resource reciprocity: an event-related brain potentials analysis. *Acta Psychol.* 70, 77–97. doi: 10.1016/0001-6918(89)90061-9
- Sirois, S., and Brisson, J. (2014). Pupillometry. *Wiley Interdiscip. Rev. Cogn. Sci.* 5, 679–692. doi: 10.1002/wcs.1323
- Smallwood, J., and Schooler, J. W. (2006). The restless mind. *Psychol. Bull.* 132, 946–958. doi: 10.1037/0033-2909.132.6.946
- Soares, V. C. G., Souza, J. K. S., Ginani, G. E., Pompéia, S., Tierra-Criollo, C. J., and Melges, D. B. (2013). “Identification of drowsiness and alertness conditions by means of spectral F-test applied to pupillometric signals,” in *Journal of Physics: Conference Series* (Gothenburg), Vol. 477. doi: 10.1088/1742-6596/477/1/012027
- Solis-Marcos, I., and Kircher, K. (2018). Event-related potentials as indices of mental workload while using an in-vehicle information system. *Cogn. Technol. Work* 1–13. doi: 10.1007/s10111-018-0485-z
- Solovey, E. T., Zec, M., Garcia Perez, E. A., Reimer, B., and Mehler, B. (2014). “Classifying driver workload using physiological and driving performance data: two field studies,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, ON). doi: 10.1145/2556288.2557068
- Spruston, N. (2008). Pyramidal neurons: dendritic structure and synaptic integration. *Nat. Rev. Neurosci.* 9, 206–221. doi: 10.1038/nrn2286
- Stemberger, J., Allison, R. S., and Schnell, T. (2010). “Thermal imaging as a way to classify cognitive workload,” in *2010 Canadian Conference on Computer and Robot Vision, IEEE* (Ottawa, ON). doi: 10.1109/CCRV.2010.37
- Strayer, D. L., and Drews, F. A. (2007). Multi-tasking in the automobile. *Attention: From theory to practice* 121–133.
- Strayer, D. L., Turrill, J., Cooper, J. M., Coleman, J. R., Medeiros-Ward, N., and Biondi, F. (2015). Assessing cognitive distraction in the automobile. *Hum. Factors* 57, 1300–1324. doi: 10.1177/0018720815575149
- Stuiver, A., Brookhuis, K. A., de Waard, D., and Mulder, B. (2014). Short-term cardiovascular measures for driver support: increasing sensitivity for detecting changes in mental workload. *Int. J. Psychophysiol.* 92, 35–41. doi: 10.1016/j.ijpsycho.2014.01.010
- Stuiver, A., de Waard, D., Brookhuis, K. A., Dijksterhuis, C., Lewis-Evans, B., and Mulder, L. J. M. (2012). Short-term cardiovascular responses to changing task demands. *Int. J. Psychophysiol.* 85, 153–160. doi: 10.1016/j.ijpsycho.2012.06.003
- Sugie, R., Arakawa, T., and Kozuka, K. (2016). Detection of fatigue in long-distance driving by heart rate variability. *Innovat. Comput. Inform. Control Exp. Lett.* 10, 1553–1559. Available online at: [https://www.researchgate.net/profile/Toshiya\\_Arakawa2/publication/304757070\\_Detection\\_of\\_Fatigue\\_in\\_Long-distance\\_Driving\\_by\\_Heart\\_Rate\\_Variability/links/5779da3608ae4645d611f3cb.pdf](https://www.researchgate.net/profile/Toshiya_Arakawa2/publication/304757070_Detection_of_Fatigue_in_Long-distance_Driving_by_Heart_Rate_Variability/links/5779da3608ae4645d611f3cb.pdf)
- Task Force of the European Society of Cardiology (1996). Heart rate variability, standards of measurement, physiological interpretation, and clinical use. *Circulation* 93, 1043–1065. doi: 10.1161/01.CIR.93.5.1043
- Tozman, T., Magdas, E. S., MacDougall, H. G., and Vollmeyer, R. (2015). Understanding the psychophysiology of flow: a driving simulator experiment to investigate the relationship between flow and heart rate variability. *Comput. Human Behav.* 52, 408–418. doi: 10.1016/j.chb.2015.06.023
- Tran, C. C., Yan, S., Habiaremye, J. L., and Wei, Y. (2017). “Predicting driver’s work performance in driving simulator based on physiological indices,” in *International Conference on Intelligent Human Computer Interaction*, (Cham: Springer), 150–162.
- Trivedi, M. M., Cheng, S. Y., Childers, E. M., and Krotosky, S. J. (2004). Occupant posture analysis with stereo and thermal infrared video: algorithms and experimental evaluation. *IEEE Trans. Vehicul. Technol.* 53, 1698–1712. doi: 10.1109/TVT.2004.835526
- Tsukahara, J. S., Harrison, T. L., and Engle, R. W. (2016). The relationship between baseline pupil size and intelligence. *Cogn. Psychol.* 91, 109–123. doi: 10.1016/j.cogpsych.2016.10.001
- Tsunashima, H., and Yanagisawa, K. (2009). Measurement of brain function of car driver using functional near-infrared spectroscopy (fNIRS). *Comput. Intell. Neurosci.* 2009:164958. doi: 10.1155/2009/164958
- Unni, A., Ihme, K., Jipp, M., and Rieger, J. W. (2017). Assessing the driver’s current level of working memory load with high density functional near-infrared spectroscopy: a realistic driving simulator study. *Front. Hum. Neurosci.* 11:167. doi: 10.3389/fnhum.2017.00167
- van Boxtel, A. (2001). Optimal signal bandwidth for the recording of surface EMG activity of facial, jaw, oral, and neck muscles. *Psychophysiology* 38, 22–34. doi: 10.1111/1469-8986.3810022
- van Gent, P., Farah, H., van Nes, N., and van Arem, B. (2018). Analysing noisy driver physiology real-time using off-the-self sensors: heart rate analysis software from the taking the fast lane project. *J. Open Res. Softw.* doi: 10.13140/RG.2.2.24895.56485
- Vicente, J., Laguna, P., Bartra, A., and Bailón, R. (2016). Drowsiness detection using heart rate variability. *Med. Biol. Eng. Comput.* 54, 927–937. doi: 10.1007/s11517-015-1448-7
- Villringer, A., Planck, J., Hock, C., Schleinkofer, L., and Dirnagl, U. (1993). Near infrared spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults. *Neurosci. Lett.* 154, 101–104. doi: 10.1016/0304-3940(93)90181-J
- Visnovcova, Z., Mestanik, M., Gala, M., Mestanikova, A., and Tonhajzerova, I. (2016). The complexity of electrodermal activity is altered in mental cognitive stressors. *Comput. Biol. Med.* 79, 123–129. doi: 10.1016/j.combiomed.2016.10.014
- Vogelpohl, T., Kühn, M., Hummel, T., and Vollrath, M. (2018). Asleep at the automated wheel—Sleepiness and fatigue during highly automated driving. *Accid. Anal. Prev.* 4575, 30117–30119. doi: 10.1016/j.aap.2018.03.013
- von Lüthmann, A., Herff, C., Heger, D., and Schultz, T. (2015). Toward a wireless open source instrument: functional near-infrared spectroscopy in mobile neuroergonomics and BCI applications. *Front. Hum. Neurosci.* 9:617. doi: 10.3389/fnhum.2015.00617
- Voytek, B., and Knight, R. T. (2015). Dynamic network communication as a unifying neural basis for cognition, development, aging, and disease. *Biol. Psychiatry* 77, 1089–1097. doi: 10.1016/j.biopsych.2015.04.016
- Wang, F., Wang, H., and Fu, R. (2018). Real-Time ECG-based detection of fatigue driving using sample entropy. *Entropy* 20:196. doi: 10.3390/e20030196
- Wang, H. (2015). “Detection and alleviation of driving fatigue based on EMG and EMS/EEG using wearable sensor,” in *Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare*. doi: 10.4108/eai.14-10-2015.2261628
- Wang, L., Wang, H., and Jiang, X. (2017). A new method to detect driver fatigue based on EMG and ECG collected by portable non-contact sensors. *PROMET-Traffic Transport.* 29, 479–488. doi: 10.7307/ptt.v29i5.2244
- Wang, X., and Chuan, X. (2016). Driver drowsiness detection based on non-intrusive metrics considering individual specifics. *Accid. Anal. Prev.* 95, 350–357. doi: 10.1016/j.aap.2015.09.002
- Wanyan, X., Zhuang, D., Lin, Y., Xiao, X., and Song, J. -W. (2018). Influence of mental workload on detecting information varieties revealed by mismatch negativity during flight simulation. *Int. J. Ind. Ergon.* 64, 1–7. doi: 10.1016/j.ergon.2017.08.004
- Wascher, E., Arnau, S., Gutberlet, I., Karthaus, M., and Getzmann, S. (2018). Evaluating pro- and re-active driving behavior by means of the EEG. *Front. Hum. Neurosci.* 12:205. doi: 10.3389/fnhum.2018.00205
- Wascher, E., Rasch, B., Sängler, J., Hoffmann, S., Schneider, D., Rinkenauer, G., et al. (2014). Frontal theta activity reflects distinct aspects of mental fatigue. *Biol. Psychol.* 96, 57–65. doi: 10.1016/j.biopsycho.2013.11.010
- Wei, C. S., Lin, Y. P., Wang, Y. T., Lin, C. T., and Jung, T. P. (2018). A subject-transfer framework for obviating inter- and intra-subject variability in EEG-based drowsiness detection. *Neuroimage* 174, 407–419. doi: 10.1016/j.neuroimage.2018.03.032
- Wekselblatt, J. B., and Niell, C. M. (2015). Behavioral state—getting in the zone. *Neuron* 87, 7–9. doi: 10.1016/j.neuron.2015.06.020
- Wijsman, J., Grundlehner, B., Penders, J., and Hermens, H. (2013). Trapezius muscle EMG as predictor of mental stress. *ACM Trans. Embedded Comput. Syst.* 12:99. doi: 10.1145/2485984.2485987
- Wikström, B.-O. (1993). Effects from twisted postures and whole-body vibration during driving. *Int. J. Ind. Ergon.* 12, 61–75. doi: 10.1016/0169-8141(93)90038-F
- Wilhelm, B. J., Widmann, A., Durst, W., Heine, C., and Otto, G. (2009). Objective and quantitative analysis of daytime sleepiness in physicians after night duties. *Int. J. Psychophysiol.* 72, 307–313. doi: 10.1016/j.ijpsycho.2009.01.008
- Yakobi, O. (2018). “Determinants of Association and Dissociation between Subjective and Objective Measures of Workload,” in *Proceedings of the*



- Human Factors and Ergonomics Society Annual Meeting*, (Los Angeles, CA: SAGE Publications).
- Yan, S., Tran, C. C., Wei, Y., and Habiyaemye, J. L. (2017). Driver's mental workload prediction model based on physiological indices. *Int. J. Occup. Saf. Ergon.* 15, 1–9. doi: 10.1080/10803548.2017.1368951
- Yang, S., Hosseiny, S. A. R., Susindar, S., and Ferris, T. K. (2016). "Investigating driver sympathetic arousal under short-term loads and acute stress events," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 60, (Los Angeles, CA: SAGE Publications).
- Yarkoni, T., and Westfall, J. (2017). Choosing prediction over explanation in psychology: lessons from machine learning. *Perspect. Psychol. Sci.* 12, 1100–1122. doi: 10.1177/1745691617693393
- Yerkes, R. M., and Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *J. Comp. Neurol. Psychol.* 18, 459–482. doi: 10.1002/cne.920180503
- Ying, L., Fu, S., Qian, X., and Sun, X. (2011). Effects of mental workload on long-latency auditory-evoked-potential, salivary cortisol, and immunoglobulin A. *Neurosci. Lett.* 491, 31–34. doi: 10.1016/j.neulet.2011.01.002
- Yoshida, Y., Ohwada, H., Mizoguchi, F., and Iwasaki, H. (2014). Classifying cognitive load and driving situation with machine learning. *Int. J. Mach. Learn. Comput.* 4:210. doi: 10.7763/IJMLC.2014.V4.414
- Yoshino, K., Oka, N., Yamamoto, K., Takahashi, H., and Kato, T. (2013). Functional brain imaging using near-infrared spectroscopy during actual driving on an expressway. *Front. Hum. Neurosci.* 7:882. doi: 10.3389/fnhum.2013.00882
- Zander, T. O., Andreessen, L. M., Berg, A., Bleuel, M., Pawlitzki, J., Zawallich, L., et al. (2017). Evaluation of a dry EEG system for application of passive brain-computer interfaces in autonomous driving. *Front. Hum. Neurosci.* 11:78. doi: 10.3389/fnhum.2017.00078
- Zhao, C., Zhao, M., Liu, J., and Zheng, C. (2012). Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator. *Accid. Anal. Prev.* 45, 83–90. doi: 10.1016/j.aap.2011.11.019

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Lohani, Payne and Strayer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Comparing the Relative Strengths of EEG and Low-Cost Physiological Devices in Modeling Attention Allocation in Semiautonomous Vehicles

Dean Cisler<sup>1\*</sup>, Pamela M. Greenwood<sup>1</sup>, Daniel M. Roberts<sup>1</sup>, Ryan McKendrick<sup>2</sup> and Carryl L. Baldwin<sup>1</sup>

<sup>1</sup>Department of Psychology, George Mason University, Fairfax, VA, United States, <sup>2</sup>Northrop Grumman, Falls Church, VA, United States

## OPEN ACCESS

### Edited by:

Bruce Mehler,  
Massachusetts Institute of  
Technology, United States

### Reviewed by:

Ben D. Sawyer,  
University of Central Florida,  
United States  
Bobbie Seppelt,  
Massachusetts Institute of  
Technology, United States  
Jose Manuel Ferrandez,  
Universidad Politécnica de  
Cartagena, Spain

### \*Correspondence:

Dean Cisler  
dcisler@masonlive.gmu.edu

**Received:** 17 June 2018

**Accepted:** 11 March 2019

**Published:** 29 March 2019

### Citation:

Cisler D, Greenwood PM,  
Roberts DM, McKendrick R and  
Baldwin CL (2019) Comparing the  
Relative Strengths of EEG and  
Low-Cost Physiological Devices in  
Modeling Attention Allocation in  
Semiautonomous Vehicles.  
*Front. Hum. Neurosci.* 13:109.  
doi: 10.3389/fnhum.2019.00109

As semiautonomous driving systems are becoming prevalent in late model vehicles, it is important to understand how such systems affect driver attention. This study investigated whether measures from low-cost devices monitoring peripheral physiological state were comparable to standard EEG in predicting lapses in attention to system failures. Twenty-five participants were equipped with a low-fidelity eye-tracker and heart rate monitor and with a high-fidelity NuAmps 32-channel quick-gel EEG system and asked to detect the presence of potential system failure while engaged in a fully autonomous lane changing driving task. To encourage participant attention to the road and to assess engagement in the lane changing task, participants were required to: (a) answer questions about that task; and (b) keep a running count of the type and number of billboards presented throughout the driving task. Linear mixed effects analyses were conducted to model the latency of responses reaction time (RT) to automation signals using the physiological metrics and time period. Alpha-band activity at the midline parietal region in conjunction with heart rate variability (HRV) was important in modeling RT over time. Results suggest that current low-fidelity technologies are not sensitive enough by themselves to reliably model RT to critical signals. However, that HRV interacted with EEG to significantly model RT points to the importance of further developing heart rate metrics for use in environments where it is not practical to use EEG.

**Keywords:** low-cost technology, attention, alpha-band, semiautonomous vehicles, eye-tracking, electrocardiography

## INTRODUCTION

Semiautonomous driving systems or “partial driving automation” (SAE Level 2; SAE International, 2016) are driver assistance systems that are increasingly available in passenger vehicles, with conditional driving automation (SAE level 3) still largely under development. As recently pointed out by Eriksson and Stanton (2017), SAE level 2 is commonly confused with highly automated driving, when in fact the semiautonomous level requires drivers to monitor the automation. For both SAE levels 2 and 3, drivers must be prepared to intervene when system limitations and failures

occur. These systems are intended to be advanced driver assistance systems (ADASs) and thus are not intended to supplant the need for drivers to maintain vigilant attention and intervene when necessary.

ADAS in passenger vehicles are urgently needed. Highway fatalities in the US declined steadily for five decades but increased more than 10% in the first 6 months of 2016 with only a slight decline (0.8%) from that peak in 2017 (NHTSA's National Center for Statistics and Analysis, 2017). Overall, the 2016 and 2017 fatality numbers are a troubling reversal of decades of improvement in highway fatalities. Importantly, an estimated 94% of fatal crashes are attributable to driver error, with 41% of those errors being recognition errors including inattention, internal and external distractions, and inadequate surveillance (Singh, 2015). The advent of semiautonomous systems in vehicles is already reducing crashes by reducing driver error. Automatic emergency braking reduced rear-end crashes by about 40% (Cicchino, 2017) and rear cross-traffic alerts reduced backing crashes by about 32% (Cicchino, 2018).

Despite the potential benefit for automation to reduce vehicle crashes, automation can have unpredictable effects on drivers. Increased vehicle automation changes how drivers pay attention and tend to decrease situation awareness (Sarter et al., 1997; Endsley, 2017). People use automation when they should not, over-rely on automation, over-trust automation, and fail to monitor automation closely (Parasuraman and Riley, 1997). In a prior meta-analysis, a greater degree of automation was found to be associated with reduced ability to recover from a system failure (Onnasch et al., 2014). Importantly, increased levels of vehicle automation shift the driver's role from one of active control to one of a supervisor of the automation (van den Beukel et al., 2016). It is imperative to understand how advanced vehicle automation affects the safety of drivers and passengers.

Although ADASs do reduce crashes, they also have a number of known operational limits. Misunderstanding or over-trust in these systems may result in drivers failing to monitor the automation and subsequently failing to detect critical signals related to the system's functionality (Parasuraman and Manzey, 2010). There have been recent news reports of fatal Tesla crashes that occurred when the automation failed to detect obstacles during a period when the driver was not monitoring the automation (CNBC, 2018). Current ADASs are not designed to brake effectively during "cut-in," "cut-out," or crossing-path scenarios. Pedestrian detection systems do not detect all pedestrians, notably those carrying large packages. These limits render driver inattention hazardous in all partially automated SAE 2 vehicles. Now that most new vehicles are equipped with some automation, it is important to understand how drivers respond to signals indicating automation disengagement. Inattentive drivers may require more urgent warnings—warnings that could annoy or startle the attentive driver. Therefore, warnings of automation faltering or failing should be tailored to the driver's attentional state to be most effective. Further, there is increasing recognition that under some conditions, safety considerations may require

automation to shut itself off to protect an inattentive driver. Such systems would depend on non-invasive sensors able to reliably detect driver attentional state. A major focus of the current work is to understand the predictive capabilities of non-invasive low-cost sensors, compared to well established but expensive and relatively cumbersome methods such as multi-channel EEG.

EEG obtained with standard EEG recording equipment has been shown to be sensitive to attentional state and is often considered the defacto physiological measure for attention. Previous EEG studies using high-fidelity EEG systems, have reported that alpha-band activity increases just before errors in processing that stimuli (Mazaheri et al., 2009; O'Connell et al., 2009; Brouwer et al., 2012; Ahn et al., 2016; Aghajani et al., 2017; Zhang et al., 2017). Increased prestimulus alpha-band has also been associated with mind wandering during driving (Baldwin et al., 2017). Although EEG is well-established as a measure of attention, it may not be practical for use in vehicles insofar as real-time scalp recording and analysis of alpha-band power would be needed. Portable EEG systems have shown promise in their ability to monitor driver engagement and drowsiness in a simulator study (Johnson et al., 2011). Even though portable EEG systems may be capable in field settings, they are expensive compared to other portable physiological measuring systems, thereby adding to consumer costs. Lower-cost technology systems exist for monitoring driver state that are more robust and less cumbersome than EEG and thus more likely to be adapted and installed into vehicles. For example, the General Motors Cadillac 2018 and 2019 CT6 models offer a super cruise feature that includes an infrared eye-tracking system which is used by the automation to determine driver attention (Clerkin, 2017). Similarly, low-cost, reliable heart rate monitors with signal quality comparable to that produced by Zephyr™ and KardiaMobile, could potentially be integrated into vehicles to record drivers' heart electrical activity (ECG). This raises the question of whether sufficient classification sensitivity to the attentional state can be achieved with low-fidelity, low-cost sensors such as heart-rate monitors and eye-trackers?

An existing body of research has investigated the use of metrics other than EEG to monitor operator state. For example, metrics of cardiovascular activity have been used to assess constructs such as mental workload, fatigue, and operator stress. In general, both heart rate increases and heart rate variability (HRV) decreases have been associated with increased mental effort (Mulder, 1992; Wilson, 1992). For example, Stuiver et al. (2014) found that 40 s periods of HRV were sensitive to increased effort expenditure due to driving in fog vs. clear visibility, with fog-inducing decreased HRV. Mehler et al. (2012) found that heart rate and skin conductance level increase as cognitive demand increases. HRV has also been used to classify fatigue during simulated driving (Patel et al., 2011). Metrics of HRV have been found to index changes in mental effort over time as participants adapt to a task and change task strategies and performance criteria. Short periods of high HRV reflecting primarily parasympathetic influences may, therefore, serve as a sensitive index of fluctuations in task effort and temporarily

lowered levels of effort on a trial by trial basis (Thayer et al., 2012). HRV as a workload measure is generally most sensitive in the mid-range, particularly around 0.10 Hz area (Mulder, 1992). The mid-range is most sensitive to the amount of mental effort invested in the task, not task complexity, *per se*. Hogervorst et al. (2014) directly compared three measures of HRV used to index workload: (a) high-frequency HRV measured in root mean square of successive differences (RMSSDs); (b) the spectral power in the range 0.15–0.5 Hz of the ECG R to R intervals; and (c) mid-frequency variability with spectral power between 0.07 and 0.15 Hz of the ECG R to R intervals. It should be noted that the third measure would be categorized as low frequency according to the (Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology, 1996). Hogervorst et al. (2014), found that, apart from EEG, only respiration frequency and RMSSD produced a significant classification of workload.

Metrics of eye movements have also shown promise in recent years as indices of attention. Metrics obtained from eye trackers, such as fixations, horizontal spread of fixations, and gaze concentration have been used successfully to index attention in several recent driving investigations. For example, Wang et al. (2014) compared a number of different eye gaze metrics and found that horizontal gaze concentration derived from the standard deviation of horizontal gaze position was robust and sensitive to changes to cognitive demand during driving on actual roads. Research by Fridman et al. (2018) used in-vehicle video recordings of eye movements in conjunction with either, Hidden Markov Models or three-dimensional convolutional neural network, to classify driver cognitive load during driving on an actual highway. Likewise, in a simulated vehicle automation task, Louw and Merat (2017) found horizontal gaze dispersion to be sensitive to increased task demand stemming from secondary task engagement. Dehais et al. (2011) and Zeeb et al. (2015) found that gaze concentration was a sensitive index of attentional focusing, found to predict the speed of “take-over” from automation.

Combinations of physiological measures have shown particular promise. For example, combinations of EEG, eye-tracking, and HRV have been used to: (a) classify operator states (Hogervorst et al., 2014); (b) determine whether a driver is on-task or mind wandering (Baldwin et al., 2017); and (c) to successfully adapt automation to improve driver performance (Wilson and Russell, 2003a,b). Hogervorst et al. (2014) provided a partial comparison, reporting that EEG measures obtained the highest classification accuracy compared to eye, heart, and respiratory measures. When EEG was combined with eye measures (pupil size and eyeblinks) there was not a significant improvement over EEG alone as predictors of workload in an n-back working memory task.

In light of evidence that RMSSD (Hogervorst et al., 2014) and eye-gaze (Dehais et al., 2011; Wang et al., 2014) were both found to be effective in predicting driver attentiveness, we hypothesized that these two measures in combination and when obtained from low-cost equipment could be as sensitive in predicting driver performance in a simulator during automated driving as EEG alpha-band, obtained from high-fidelity EEG equipment.

## MATERIALS AND METHODS

### Participants

Twenty-five participants were recruited through the George Mason University undergraduate research pool, in exchange for course credit. Participant requirements were to be above 18 years of age, have normal or corrected to normal vision and hearing, not currently taking psychoactive medications, and have a valid United States driver's license. Participants were also asked to not wear heavy eye makeup the day of their scheduled appointment or wear braids, wigs, or hair extensions as they affect contact between EEG electrodes and the scalp. In order to increase enrollment in the study, in addition to course credit, some participants were given a \$15.00 bonus upon completion of their scheduled session. **Table 1** provides an overview of participant demographic information.

### Materials

#### Simulated Drives

Five fully autonomous drives were programmed using a low-fidelity desktop simulator containing Internet Screen Assembler pro version 20 and Real Time Technologies Sim Creator version 3.2 simulator software on a Windows 7 computer with 64-bit operating system. Each of the drives was displayed on a Dell Monitor with screen size measuring 52 cm in length and 32.5 cm in height with a screen resolution of 1,920 × 1,200 pixels. Each of the drives was programmed to complete an automated lane changing task, adapted from Mattes (2003) and lasted approximately 10 min in duration. The 10-min duration was due to limitations in the Sim Creator software. During the drives, participants were instructed to respond with serial button presses every time the system indicated there was an automation failure. System functionality was represented by right or left facing arrows, appearing in the bottom right corner of the monitor that varied in the gradient of the color red to green and appeared on average every 13 s, with a jitter  $\pm 2$  s resulting in 4–5 lane changes per minute. Arrow duration was 150 ms. System reliability was indicated by the amount of red at the tip of the arrow. Arrows representing reliable system functionality, Reliable Automation Arrows (presented on 80% of trials) indicated the system was operating normally (the base of the arrow was green with a small amount of red at the tip). After the presentation of a Reliable Automation Arrow, the vehicle would respond by changing lanes correctly. Arrows indicating unreliable system functionality, Unreliable Automation Arrows (presented on 20% of trials or on 10 trials per drive) indicated that the system had failed (the arrow tip was completely filled in with red). After the presentation of the Unreliable Automation Arrow, the vehicle would respond by making one of three possible lane changes. Of the ten Unreliable Automation Arrows, on six of them the vehicle would fail to make a lane change, for two of them the vehicle

**TABLE 1** | Participant demographics.

| Total participants             | 25           | Female (12)            |
|--------------------------------|--------------|------------------------|
| Age                            | Range: 18–39 | Mean: 22.6 (6.01 SD)   |
| Driving experience (in months) | Range: 6–270 | Mean: 57.13 (60.75 SD) |



would respond by making an incorrect lane change (opposite of where the Unreliable Automation Arrow was pointing), and for two the vehicle would make a correct lane change. Participants were told to respond with a button press if the arrow was an Unreliable Automation Arrow then make a second button press to indicate which type of lane change the vehicle made after the presentation of the Unreliable Automation Arrow. Participants were exposed to a total of 50 arrows per 10-min drive.

Two secondary tasks were administered to participants in addition to the lane changing task. The point of these tasks was to keep participants engaged in the driving task and discourage participants from focusing their eyes on the icons in the interface. During each of the time periods, participants were asked to: (a) keep a running count of the number of Coca-Cola and Northrop Grumman signs they encountered; and (b) answer “driver engagement” questions regarding the vehicle’s status such as: speed changes, current lane position, or lane changes. In each time period, there were 25 total billboards and three “driver engagement” questions.

### Questionnaires

Participants were administered a demographics questionnaire, the Trust Between People and Automation (Jian et al., 2000), Merritt (2011) Trust Scale Items, the Merritt (2011) scale based on Liking Items, and the Propensity to Trust Scale Items (Merritt et al., 2013).

### EEG Recording

Each participant was equipped with a 40-channel NuAmps EEG cap with silver/silver-chloride electrodes. Data were recorded from a subset of electrodes: Fz, Cz, Pz, Oz, F1, F2, P1, P2, Ground (at location AFz), A1 (the left mastoid, serving as the online reference), and A2 (the right mastoid), as well as EOG electrodes placed above and below the left eye as well as at the outer canthus of both eyes. Data were collected at a sampling rate of 500 Hz with an online high-pass filter of 0.1 Hz and an online low-pass filter of 70 Hz.

### Eye-Tracking

Gaze dispersion was recorded using the Pupil Pro headset developed by Pupil Labs. This is a low-cost eye-tracker that monitors the participant’s right pupil with a camera as well as the environment with a head-mounted camera. The data was recorded using Pupil Lab recording software. Sensor settings for the cameras were as follow: the pupil camera was set to  $640 \times 480$  with a frame rate of 120 fps maximum resolution and the world camera was set  $1,920 \times 1,080$  with a frame rate of 30 fps maximum resolution.

### Heart Rate Monitor

A low-cost Zephyr BioPatch heart rate monitor was attached to the participant using ECG electrodes in order to collect heart rate activity during each of the time periods.

### Lab-Streaming Layer

The lab-streaming layer (LSL) software library<sup>1</sup> was used to synchronize the timestamps through a network connection

between the driving simulation, as well as our physiological devices: the eye-tracker and heart rate monitor.

## Procedure

After providing written informed consent of a protocol approved by George Mason University’s Human Subjects Institutional Review Board, participants were introduced to the heart rate monitor, eye-tracker, the EEG cap. Procedures were used to lower impedance of the scalp EEG electrodes.

### ECG Setup

Participants were handed the Zephyr Heart Rate Monitor and asked to place it so that it was centered with their sternum so the ECG electrodes could acquire heart rate activity at the fourth intercostal space located at the left and right sternal border. Next, the heart rate monitor was synced with the BioHarness software on a nearby laptop computer.

### EEG Setup

Next, participants were fitted with the Neuroscan 40 channel EEG cap. Impedance was lowered to 5 k $\Omega$  or below by applying electroconductivity gel between the electrodes and the scalp then lightly abrading the scalp using a blunt needle (Luck, 2005). Next participants were shown how excessive movement can introduce noise into EEG waveforms and asked to remain as still in their chair as possible for the duration of the experiment.

### Eye-Tracking

We used the Pupil Pro headset to monitor eye movements and gaze patterns for the duration of the drives. After placing the headset on each participant, the pupil camera was adjusted to better capture their pupil. Once the camera was able to accurately track the participant’s pupil, they underwent a calibration process in order to synchronize the pupil tracking camera with the world facing camera *via* Pupil Pro software. This allowed us to track the location of the display in order to convert the gaze position to display coordinates. A confidence value is estimated for each sample of eye data that ranges from 0 to 1 indicating a level of certainty that the pupil was accurately identified for that sample. Only samples with confidence at or above 0.8 were used for further data analyses.

After setting up the participant with the physiological metrics, participants were seated 75 cm away from the monitor. At the start of the training drive, each participant was read the instructions aloud and introduced to the controls on the gear shift. Participants were instructed to immediately press the button labeled as *U* as soon as they saw an Unreliable Automation Arrow, then make a second button press indicating the type of error that occurred (*N* = No lane change, *I* = Incorrect lane change, and *C* = Correct lane change). Participants were asked to only respond to the Unreliable Automation Arrows. Participants were also instructed to pay attention to the images on each of the billboards and count the number of times they saw logos for Coca-Cola and Northrop Grumman as well as answer the “Yes” or “No” driver engagement questions (DEQ; e.g., “Speed increased after last arrow?”, “I am currently traveling 67 mph?”, “I am currently in the far right lane?” presented during each trial. Participants were allowed to complete the practice as

<sup>1</sup><https://github.com/scn/labstreaminglayer>

many times as they needed to feel comfortable responding to the task. After training, participants were then administered the five time periods in counterbalanced order. After each time period participants were asked to report the total number of Coca-Cola and Northrop Grumman billboards to the experimenter. Upon completion of the simulated driving session, participants were administered the questionnaires.

## Data Analysis

### Heart Rate Variability (HRV)

Data collected from the driving simulator was synced in time with the data collected from the Zephyr Heart Rate Monitor. We sampled the heart rate data starting 10 s before the onset of the Unreliable Automation Arrow until the presentation of the arrow. As reported in Klinger (1978), shifts in thought patterns can happen on average every 14 s. A maximum window of 10 s was chosen for ECG activity as that would allow us to maximize the number of sampled beats per second without extending too far back to potentially sample HRV due to the previous Automation Arrow. Based on previous work by Hogervorst et al. (2014), HRV was the measure of interest because it has been shown to be a robust classifier in identifying low vs. high workload compared to spectrally defined medium and high HRV. We calculated the ECG R-wave peak to peak interval for each trial using the MATLAB wavelet toolbox, using the maximum overlap discrete wavelet transform (MODWT). The squared absolute value of the signal approximation was calculated allowing for the use of an algorithm to identify R peaks for further analysis. Mean R to R was calculated by averaging the time between R peaks (meanRR). HRV was calculated using the RMSSDs.

### EEG Processing

EEG spectral data were processed using MATLAB with EEGLAB toolbox version 12.0.2.4b (Delorme and Makeig, 2004). EEG channels were mapped using the BESA file, a four shell DIPFIT spherical model of the channel locations. Data were re-referenced to the average of the two mastoid electrodes. Unreliable Automation Arrows were labeled within the waveform of the EEG data. Data were filtered at a high-pass filter of 1 Hz cutoff and 2 Hz transition bandwidth, and a low-pass filter of 40 Hz and 10 Hz transition bandwidth. Data was decomposed *via* independent component analysis (ICA), and components representing blinks or eye movements were visually identified and removed. Electrodes exceeding  $\pm 2$  standard deviations were identified as artifactual and rejected. Additionally, data exceeding  $\pm 100 \mu\text{V}$  was rejected from the data to remove artifacts caused by large movements or other noise. Data from electrodes rejected due to artifacts that exceeded two standard deviations were subjected to spherical interpolation. Dummy markers were placed in the EEG data 1 s before each unreliable signal event to the presentation of the arrow and the data were epoched to those markers. The 1-s window was chosen to capture the mental state of participants immediately prior to the onset of the Unreliable Automation Arrow. Previous research on what is termed “prestimulus alpha” have shown increases in alpha spectral power, prior to a failure in detecting a signal, using a

time window of 800 ms to 1,000 ms prior to stimulus onset (Busch et al., 2009; Mazaheri et al., 2009). Each epoch was linearly detrended, and a hamming windowed Fourier transform was used to convert the data from the time-domain to the frequency domain, as implemented in the MATLAB function *pwelch*. The data were then converted into decibel power using  $10 \cdot \log_{10}(\text{power})$  in order to get a better approximation of the normal distribution. The FFT bin nearest to 10 Hz, here 9.76 Hz, was used to analyze alpha activity at electrodes Pz, Cz, and Fz.

### Eye-Tracking

Gaze dispersion data collected from the eye-tracker was synced in time with each drive through LSL and sampled 3 s before each onset of an Unreliable Automation Arrow to the presentation of the arrow. This time window was selected in order to maximize the number of sampled eye movements prior to the onset of the Unreliable Automation Arrow while avoiding potential contamination from eye movements that occurred due to the billboard task. Horizontal and vertical gaze dispersion were calculated by computing the standard deviation of a measure of pixels over which the eyes moved for the X (horizontal) or Y (vertical) dimension of the raw data identified with a confidence value of 0.8 or higher. Horizontal and vertical gaze dispersion was then transformed using the natural log of their values ( $\ln X$  and  $\ln Y$ , respectively) to approximate the normal distribution.

### Behavioral Data

Responses to the presentation of the Unreliable Automation Arrows and responses indicating the type of error (second button presses) were extracted to assess changes in performance over a time period during the experimental session. Participants were instructed to immediately respond as soon as they saw an Unreliable Automation Arrow. Due to high accuracy shown by participants in identifying Unreliable Automation Arrows, the latency of response to Unreliable Automation Arrows was the measure of interest. We first calculated the grand mean for our entire data set and standard deviation. We set all response times higher than 2,600 ms to equal 2,600 ms. In order to get a better idea of how well participants were able to distinguish between critical events and reliable events, the *A* measure of sensitivity was used. Since the measure of  $d'$  is calculated by taking the difference of hits and false alarms that have been converted from probabilities into z-scores, the inclusion of a 1 or a 0 can lead to a value that does not fall below the ROC curve. Use of non-parametric sensitivity calculated using the *A* statistic, as described in Zhang and Mueller (2005), eliminates the reliance of converting probabilities to z-scores and obtains the measure of sensitivity by calculating the average of the minimum-area and maximum-area proper ROC curves as constrained by false alarms and hits. Analysis of the accuracy of the second button press that indicated the type of lane change the vehicle made (incorrect, correct, or no lane change) were calculated for further analysis in SPSS.

For the billboard task, the probability of hits and false alarms was calculated for each 10 min time period of the drive. The *A* statistic was calculated for further analysis in SPSS. For the DEQs, accuracy was calculated by averaging the responses of the questions for each 10 min episode of the drive. With

the trust questionnaire data, statements identified as being negative were reverse coded allowing us to average the scores for further analyses.

## RESULTS

Data were analyzed using SPSS and the R statistical package (R Core Team, 2017). To assess how well participants were able to discriminate between Unreliable Automation and Reliable Automation Arrows, we calculated the *A* statistic for each time period. To assess speed-accuracy tradeoff, a correlation analysis was conducted comparing *A* to reaction time (RT) for the discrimination task. That analysis produced a significant, 2-tailed, negative correlation ( $R^2 = -0.50$ ,  $p < 0.05$ ), indicating that participants did not slow their responses in order to achieve higher accuracy scores. Since accuracy was at the ceiling for participants, discrimination RT was the behavioral measure of interest.

In order to model RT to Unreliable Automation Arrow over the five time periods, linear-mixed effects models were carried out. These models were constructed using the R package *lme4* (Bates et al., 2012). We conducted interactive models of RT to Unreliable Automation Arrow across the five time periods for each measure (alpha-band  $\times$  time period, HRV  $\times$  time period, meanRR  $\times$  time period, lnX  $\times$  time period). These were random intercept and slope models. Participant and trial (10 trials in each time period) were random factors. For each variable, only time period significantly modeled RT ( $p < 0.05$ ). Only alpha-band interacted with the time period in modeling RT. A likelihood ratio test (LRT) comparing the interactive model (alpha-band  $\times$  time period) to a null additive model (alpha-band + time period) produced a significant Chi-square ( $X^2_{(1)} = 5.251$ ,  $p = 0.0219$ ), suggesting that the interaction was important in modeling RT.

Linear-mixed effects models were also used to model RT. An interactive model of RT was constructed with alpha-band, meanRR, HRV, lnX, and time period as fixed factors (Formula:  $RT \sim 1 + (Pz\ Alpha + meanRR + HRV + lnX + TimePeriod)^3 + (1|Participant) + (1|Trial)$ ). Participant and trial (10) were random factors. That model produced two significant interactions (AIC = 2727.7, BIC = 2873.3,  $p < 0.05$ ), indicating the likelihood of alpha-band  $\times$  time period ( $\beta = 0.04158$ ) and alpha-band  $\times$  HRV ( $\beta = -0.1588$ ) in modeling RT. As horizontal gaze dispersion (lnX) did not contribute significantly to the model, lnX was dropped from the model and a reduced model was fitted (Bolker et al., 2009). The reduced LME was conducted to model RT using alpha-band, meanRR, HRV, and time period as the fixed factors (Formula:  $RT \sim 1 + (Pz\ Alpha + meanRR + HRV + TimePeriod)^3 + (1|Participant) + (1|Trial)$ ). Interactions were limited to two- and three-way. That model produced a significant three-way interaction (AIC = 2713.2, BIC = 2803.5,  $p < 0.05$ , marginal  $R^2 = 0.02$ , conditional  $R^2 = 0.42$ ) indicating the likelihood of alpha-band, HRV, and time period ( $\beta = 0.03861$ ) in modeling RT. The  $R^2$  values, calculated and reported as described in Nakagawa et al. (2017), indicate that 2% of the variance was explained by the fixed factors

alone while 42% of the variance was explained by random effects included in the model. The model also produced a significant two-way interaction of MeanRR  $\times$  Time Period ( $\beta = -0.03588$ ,  $p < 0.05$ ). A LRT comparing the interactive model (alpha-band  $\times$  meanRR  $\times$  HRV  $\times$  time period) with an additive null model produced a significant Chi-square ( $X^2_{(10)} = 21.092$ ,  $p = 0.021$ ), indicating the interactions were important in modeling RT. Since the only significant three-way interaction involved alpha-band, HRV, and time period, LRTs were conducted to test the interactions: (a) alpha-band  $\times$  HRV; (b) alpha-band  $\times$  Time Period; and (c) HRV  $\times$  Time Period. The three LRT tests showed that alpha-band  $\times$  HRV ( $X^2_{(9)} = 18.649$ ,  $p = 0.0284$ ) and HRV  $\times$  Time Period ( $X^2_{(9)} = 19.228$ ,  $p = 0.023$ ) were significant. The interaction of Alpha-band  $\times$  Time Period was not significant ( $X^2_{(9)} = 15.809$ ,  $p = 0.071$ ). Considered together, these results indicate that alpha-band, HRV, and time period are important factors in modeling RT, with meanRR a weaker factor. **Figure 1** shows the changes for the physiological measures over each time period. **Figure 2** provides a visual comparison of alpha-band power and HRV over time period.

## Response to Lane Change Accuracy

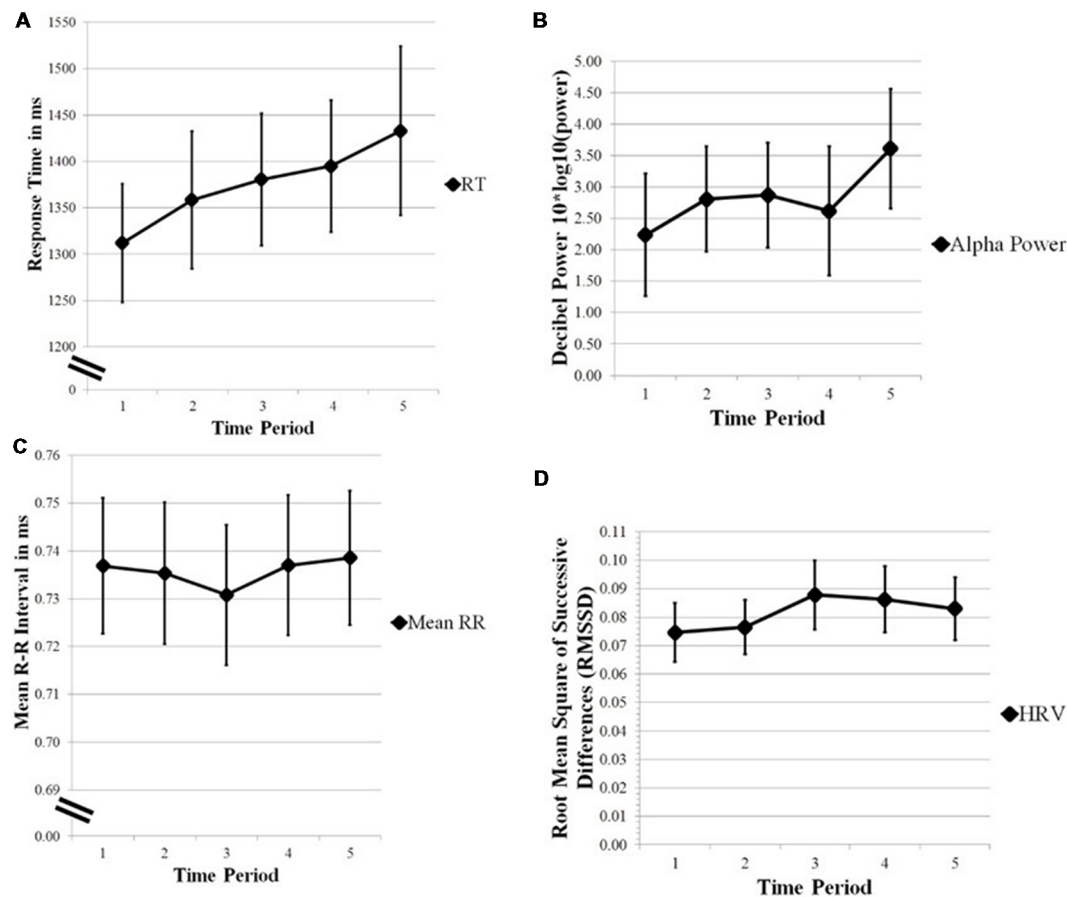
Accuracy scores calculated from the second button presses which identified the type of lane change made by the vehicle were submitted to a repeated measures ANOVA to assess the change in accuracy over time. There was no statistical significance in the analysis of changes over time in accuracy of deciding which type of lane change was made by the vehicle ( $F_{(4,92)} = 0.404$ ,  $p = 0.806$ ).

## Billboard Task

Preliminary analyses of *A* sensitivity scores were calculated looking at the changes in sensitivity to identifying the Northrup Grumman and Coca Cola billboards over time. The *A* statistic was calculated for the billboard task, as shown in **Figure 3**. Due to high accuracy for the billboard responses and in the absence of a hypothesis on an effect of the two billboard types, *A* scores were collapsed across Northrup Grumman and Coca Cola billboards. A repeated measures ANOVA was conducted in SPSS looking at changes in *A* sensitivity scores as a function of time. Statistical significance was not observed ( $F_{(4,92)} = 1.495$ ,  $p = 0.210$ ).

## Driver Engagement Questions

Accuracy was calculated for each time period by averaging the responses for the DEQs. As shown in **Figure 4**, participants increased in accuracy in their responses to the questions before showing a performance decrease at the third time period and an increase in performance for the fourth and fifth time period. A repeated measures ANOVA analyzed the change in accuracy over time. Mauchly's test of sphericity indicated that the assumption of sphericity had not been violated and therefore sphericity was assumed. There was a marginal effect of time on accuracy of response ( $F_{(4,96)} = 2.353$ ,  $p = 0.059$ ).



**FIGURE 1 | (A)** Reaction time (RT). **(B)** Alpha-band power. **(C)** Mean RR. **(D)** Heart rate variability (HRV) plotted over 10 min time periods. Error bars are standard error of the mean. Alpha-band power, Mean RR, and HRV are important factors in modeling RT over time.

## Trust Questionnaires

Five correlation analyses were conducted to assess the relationship between our questionnaires (Trust Between People and Automation, Merritt Trust Scale Items, Merritt scale based on Liking Items, Propensity to Trust Scale Items), physiological metrics selected based on the LME (alpha-band, Mean RR, and HRV), and behavioral metrics (RT and A). Of our five correlation analyses, there was a statistically significant negative bivariate correlation between alpha-band activity at midline parietal site Pz and the Merritt et al.'s (2013) Propensity to Trust Scale Items ( $r = -0.430$ ,  $p < 0.05$ ).

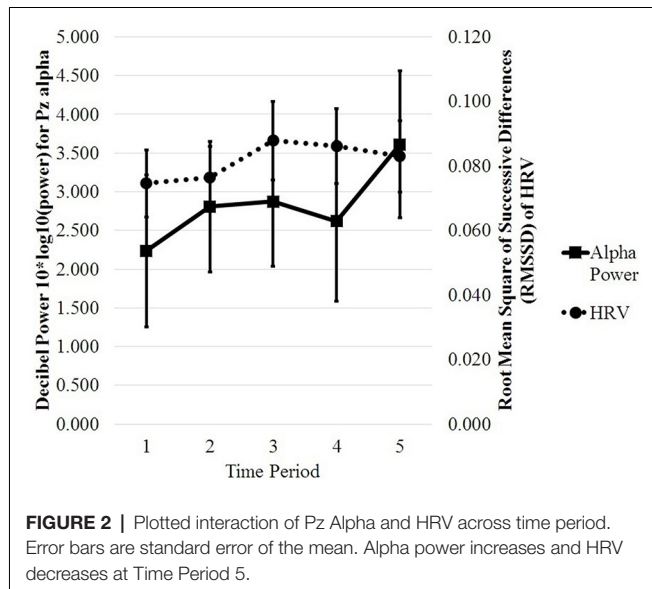
## DISCUSSION

We obtained partial support for our hypothesis. We found that HRV interacted with alpha-band activity and time period to model the speed of processing signals of automation unreliability. Gaze dispersion did not model the speed of processing signals of automation unreliability, either alone or in combination with other measures. Mean RR (heart rate measured in R-R intervals) did model RT in interaction with time period but not in interaction with alpha-band or HRV. Our findings confirm

previous evidence that prestimulus alpha-band activity is the most effective measure of mental processing (Hogervorst et al., 2014) but extend that work in showing HRV increased the predictive capability of parietal alpha-band. The readiness of the brain to process signals of system unreliability was affected by the combined effects of HRV and alpha-band activity. This evidence that HRV modulates alpha-band activity with consequences for automation signal processing argues for the importance of developing heart rate metrics in operational environments where EEG is not practical.

Regarding the time course, HRV initially increased over the session of autonomous vehicle driving, but then decreased near the end. Based on the existing HRV literature, the effect of workload on HRV depends in part on the duration of the workload demand. Mulder (1992) has argued that the cardiac response to 5–10 min periods of increased workload reflects preparation for fight-or-flight activation of the sympathetic nervous system with increased HR and decreased HRV. In contrast, a short-lasting increase in workload (25–30 s) was reflected in short-lasting increases in heart rate and blood pressure in combination with corresponding decreases in HRV and blood pressure variability (Stuiver et al., 2012). For our task,

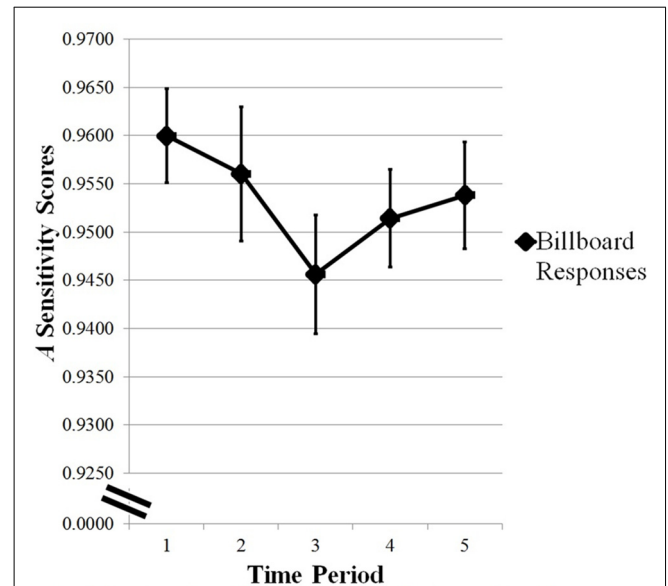




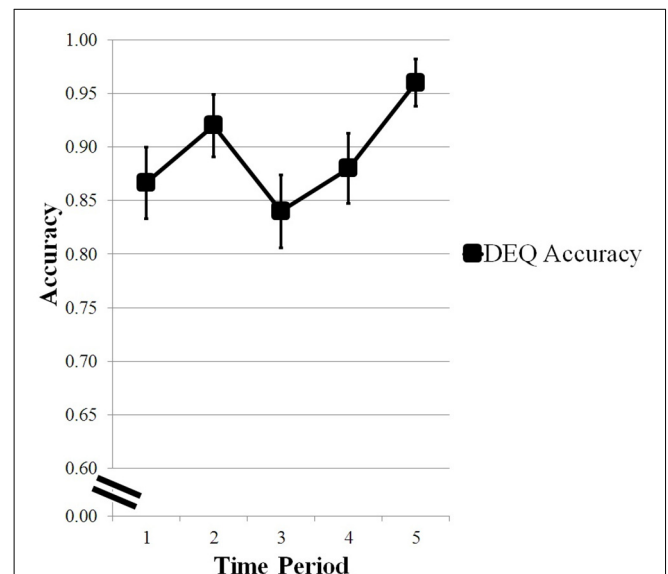
the workload may have increased when Unreliable Automation Arrow signals were presented. However, the present study measured HRV prior to those unpredictable signals indicating unreliable automation. Therefore we could not determine whether those signals transiently increased workload. The slowing of RT linearly over the session and the initial increase in HRV during autonomous vehicle operation are consistent with an interpretation that workload increased over the session.

HRV has previously been associated with emotional regulation (Appelhans and Luecken, 2008). HRV has been found to be higher in those people who were better able to regulate their emotions in social interactions (Butler et al., 2006) and in marital interactions (Smith et al., 2011). Our finding that high-frequency HRV interacted with alpha-band to model the speed of responding to unreliable signals points to a role for individual differences in emotional response regulation in processing automation signals. Further, in operational environments, it might be interesting to determine whether very low and low-frequency HRV also predicts RT of responding to signals of automation reliability.

RT to the signals of unreliable automation slowed fairly linearly over the 55-min drive. Use of RT to measure processing of signals from automation during a simulated drive is very relevant to the topic of real-world driving of vehicles equipped with ADASs. In ADAS-equipped vehicles in the real world, the driver receives frequent signals from various automation systems [e.g., drowsy driving, lane departure, lane keeping, and (more rarely) sensor failure warnings]. The slowing of RT to automation signals over the simulated driving session could suggest a vigilance decrement. However, the sensitivity index *A* from the discrimination task did not change over the driving session and accuracy of responses to the lane changing task was high. Moreover, the driving session was interrupted briefly every 10 min or so (due to limitations of the software), which would not be conducive to the development of a vigilance decrement. Therefore, we do not interpret our findings of slowed RT as



**FIGURE 3 |** Changes in A sensitivity scores for the Billboard task across time period. Error bars are standard error of the mean. Statistical significance was not observed for A sensitivity scores across time period.



**FIGURE 4 |** Changes in accuracy scores of Driver Engagement Questions (DEQ) across time period. Error bars are standard error of the means. Marginal effect of time on response accuracy.

consistent with a vigilance decrement. Workload is another possible explanation for slowing RT to signals of automation. The decrease in accuracy on the DEQs between the second and third time points, despite the high accuracy of the secondary billboard task do suggest a slight increase in workload or possible depletion of cognitive resources, such as that commonly found in vigilance tasks. However, that result was marginally significant.

Alpha-band showed a more complex pattern than RT over the session, with an overall increase in power over the driving

session, interrupted by a temporary drop in power in the 4th time period. Other investigations of alpha-band activity during vehicle operation have found increases in alpha-band power over time. Simon et al. (2011) observed an increase in alpha-band over a driving session between the first 20 min of driving and the last 20 min. That was measured only in people who claimed to be very fatigued. Craig et al. (2012) found increases in alpha-band power at frontal, central, and posterior regions over time as participants engaged in a monotonous simulated driving task. A literature review by Lal and Craig (2001) concluded that alpha band activity changed as drivers become fatigued. Since we had participants engage in a fully autonomous drive, it is possible that some became passively fatigued or drowsy during the session. An attempt by the participant to maintain engagement despite the passive nature of monitoring the automation may partially explain the high accuracy of detection of the Unreliable Automation Arrows. Further, attention to a spatial location (Worden et al., 2000) and to features (Snyder and Foxe, 2010) also modulates alpha-band activity when participants are required to detect changes in the spatial location or visual features of stimuli when they are actively suppressing irrelevant stimuli. This has been observed over dorsal areas when the color was cued but over ventral areas when motion was cued (Snyder and Foxe, 2010). In the present study in which participants were required to discriminate stimuli defined by color, the modulation of alpha-band activity could, therefore, reflect the anticipated need to discriminate based on color.

We speculate that the interaction between alpha-band, HRV, and time period that was observed in the LME model may reflect changed influences of workload and/or attention over time. The increase in alpha-band activity from time period 4 to time period 5 may reflect lapses of attention to the arrow task during the last time period. This is similar to previous findings from O'Connell et al. (2009) in which they report increased alpha-band activity prior to missing a target. As discussed above, the increase in HRV may reflect the response to workload demands placed on participants. This increase in workload in addition to reduced attention may have affected participants' response times to the Unreliable Automation Arrows indicating that the automation was in an unreliable state.

In contrast to previous work, we did not find that eye gaze measures predicted RT to signals of an unreliable automation state. Greater concentration of gaze (lower variance) has been associated with a higher workload (Victor et al., 2005). He et al. (2011) found that smaller horizontal gaze dispersion was an indication of mind wandering. As horizontal gaze dispersion did not contribute to modeling RT in the present study, we speculate that the billboard task and DEQs forced participants to maintain awareness of stimuli in the road environment and thereby remain attentive to the driving task. Further, the problem of "looking but not seeing" in driving may limit the usefulness of gaze concentration as a monitor of driver attentional state in the real world. In a real-world driving environment, operators may be less likely to detect a signal if they are not familiar with the automation. Further, in the current study, the reliability cue was not continuous. Rather, it appeared and remained on for a discrete amount of time (150 ms). This familiarity and the sudden

onset of the cue likely heightened participants' awareness of the Unreliable Automation Arrows and could have contributed to the high discrimination accuracy since participants were expecting the arrows to appear. In future studies, it would be useful to examine detection performance when changes were more gradual in a continuous display.

The present study has several limitations. First, driving in a simulator differs in a number of ways from on-road driving and the present design was an automated lane-changing task which did not require any active driving. Therefore, during the simulated drive, participants did not need to respond to sudden events common in everyday driving such as behavior of other drivers or pedestrians. Participants only needed to complete the tasks given to them. Further, the arrow task required participants to frequently monitor the automation display which changed the role of the driver from being an active participant to being a monitor of the automation. Monitoring the automation display, in conjunction with the secondary tasks, may have introduced additional noise making horizontal gaze dispersion less sensitive to operator state. We would note, however, that current SAE 2 vehicles do require the driver to monitor the automation display frequently. A second limitation was the absence of a measure of workload which makes it difficult to interpret the slowing of RT over time periods of the simulated drive. Third, the interruption of driving every 10 min makes the present study more relevant to city driving than to highway driving. Fourth, this study used a low-fidelity desk-top driving simulator. In future work, a high fidelity motion-based simulator with better automation capabilities allowing for longer automated drives will be used. Fifth, it could be argued that the high accuracy of target discrimination is a limitation. However, making the icons harder to discriminate would not be consistent with real-world driving demands which requires signals from an automation interface to be easily discriminable. Moreover, the speed of responding to those signals is an appropriate measure for driving performance. Despite these limitations, the present study provides insight into the feasibility of using portable, low-cost physiological measures to assess driver state in operational environments, including automated driving.

In sum, both EEG alpha-band and the interaction of HRV with alpha-band successfully modeled drivers' readiness to respond to signals of automation unreliability. This suggests that both those measures reflect the ability to attend to important events during driving. Our results suggest that cardiac metrics obtained from low-cost wearable sensors can be further developed for in-vehicle monitoring of driver state. Such monitoring could be used to tailor alerts or even turn off the automation (as in certain General Motors models) if the operator is judged to not be attending sufficiently to the road or monitoring the automation.

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of The GMU Internal Review Board with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of

Helsinki. The protocol was approved by the The GMU Internal Review Board. No vulnerable populations were included.

## AUTHOR CONTRIBUTIONS

DC was involved with the study design, data collection, data analysis, and write-up of this study. CB and PG were involved with the design, analysis, and write-up of this study. RM and DR aided in the study design.

## REFERENCES

- Aghajani, H., Garbey, M., and Omurtag, A. (2017). Measuring mental workload with EEG+fNIRS. *Front. Hum. Neurosci.* 11:359. doi: 10.3389/fnhum.2017.00359
- Ahn, S., Nguyen, T., Jang, H., Kim, J. G., and Jun, S. C. (2016). Exploring neurophysiological correlates of drivers' mental fatigue caused by sleep deprivation using simultaneous EEG, ECG, and fNIRS data. *Front. Hum. Neurosci.* 10:219. doi: 10.3389/fnhum.2016.00219
- Appelhans, B. M., and Luecken, L. J. (2008). Heart rate variability and pain: associations of two interrelated homeostatic processes. *Biol. Psychol.* 77, 174–182. doi: 10.1016/j.biopsycho.2007.10.004
- Baldwin, C. L., Roberts, D. M., Barragan, D., Lee, J. D., Lerner, N., and Higgins, J. S. (2017). Detecting and quantifying mind wandering during simulated driving. *Front. Hum. Neurosci.* 11:406. doi: 10.3389/fnhum.2017.00406
- Bates, D. M., Maechler, M., and Bolker, B. (2012). *lme4: Linear Mixed-Effects Models Using Eigen and C++*. R package version 0.999999-0. Available online at: <http://cran.r-project.org/web/packages/lme4/index.html>
- Brouwer, A. M., Hogervorst, M. A., van Erp, J. B. F., Heffelaar, T., Zimmerman, P. H., and Oostenveld, R. (2012). Estimating workload using EEG spectral power and ERPs in the *n*-back task. *J. Neural Eng.* 9:045008. doi: 10.1088/1741-2560/9/4/045008
- Busch, N. A., Dubois, J., and VanRullen, R. (2009). The phase of ongoing EEG oscillations predicts visual perception. *J. Neurosci.* 29, 7869–7876. doi: 10.1523/JNEUROSCI.0113-09.2009
- Butler, E. A., Wilhelm, F. H., and Gross, J. J. (2006). Respiratory sinus arrhythmia, emotion and emotion regulation during social interaction. *Psychophysiology* 43, 612–622. doi: 10.1111/j.1469-8986.2006.00467.x
- Cicchino, J. B. (2017). Effectiveness of forward collision warning and autonomous emergency braking systems in reducing front-to-rear crash rates. *Accid. Anal. Prev.* 99, 142–152. doi: 10.1016/j.aap.2016.11.009
- Cicchino, J. B. (2018). Effects of lane departure warning on police-reported crash rates. *J. Safety Res.* 66, 61–70. doi: 10.1016/j.jsr.2018.05.006
- Clerkin, B. (2017). *New GM Technology Lets Cars Go an Eye for an Eye*. Available online at: <https://www.dmv.org/articles/general-motors-eye-tracking-technology.php> [Retrieved February 4, 2019]
- CNBC. (2018). *Feds to Investigate Tesla Crash Driver Blamed on Autopilot (Reported by Robert Ferris and Lora Koldny)*. Available online at: <https://www.cnbc.com/2018/01/23/tesla-on-autopilot-crashes-into-fire-truck-on-california-freeway.html> [Retrieved June 5, 2018]
- Craig, A., Tran, Y., Wijesuriya, N., and Nguyen, H. (2012). Regional brain wave activity changes associated with fatigue: regional brain wave activity and fatigue. *Psychophysiology* 49, 574–582. doi: 10.1111/j.1469-8986.2011.01329.x
- Dehais, F., Causse, M., and Tremblay, S. (2011). Mitigation of conflicts with automation: use of cognitive countermeasures. *Hum. Factors* 53, 448–460. doi: 10.1177/0018720811418635
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009
- Endsley, M. R. (2017). Autonomous driving systems: a preliminary naturalistic study of the tesla model S. *J. Cogn. Eng. Decis. Mak.* 11, 225–238. doi: 10.1177/1555343417695197

## FUNDING

This work was supported by Northrop Grumman Research Grant 222953 to CB.

## ACKNOWLEDGMENTS

We would like to acknowledge and thank Steven Chong and Jasmine Dang in helping with data collection.

- Eriksson, A., and Stanton, N. A. (2017). Takeover time in highly automated vehicles: noncritical transitions to and from manual control. *Hum. Factors* 59, 689–705. doi: 10.1177/0018720816685832
- Fridman, L., Reimaer, B., Mehler, B., and Freeman, W. T. (2018). “Cognitive load estimation in the wild,” in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal, Canada: ACM), 1–9.
- He, J., Becic, E., Lee, Y.-C., and McCarley, J. S. (2011). Mind wandering behind the wheel: performance and oculomotor correlates. *Hum. Factors* 53, 13–21. doi: 10.1177/0018720810391530
- Hogervorst, M. A., Brouwer, A., and van Erp, J. (2014). Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. *Front. Neurosci.* 8:322. doi: 10.3389/fnins.2014.00322
- Jian, J. Y., Bisantz, A. M., and Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *Int. J. Cogn. Ergon.* 4, 53–71. doi: 10.1207/s15327566ijce0401\_04
- Johnson, R. R., Popovic, D. P., Olmstead, R. E., Stikic, M., Levensowski, D. J., and Berka, C. (2011). Drowsiness/alertness algorithm development and validation using synchronized EEG and cognitive performance to individualize a generalized model. *Biol. Psychol.* 87, 241–250. doi: 10.1016/j.biopsycho.2011.03.003
- Klinger, E. (1978). “Modes of normal conscious flow,” in *The Stream of Consciousness. Emotions, Personality, and Psychotherapy*, eds. K. S. Pope, and J. L. Singer (Boston, MA: Springer), doi: 10.1007/978-1-4684-2466-9\_9
- Lal, S. K., and Craig, A. (2001). A critical review of the psychophysiology of driver fatigue. *Biol. Psychol.* 55, 173–194. doi: 10.1016/s0301-0511(00)00085-5
- Louw, T., and Merat, N. (2017). Are you in the loop? Using gaze dispersion to understand driver visual attention during vehicle automation. *Trans. Res. Part C* 76, 35–50. doi: 10.1016/j.trc.2017.01.001
- Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique*. Cambridge: MIT Press.
- Mattes, S. (2003). “The lane-change-task as a tool for driver distraction evaluation,” in *Quality of Work and Products in Enterprises of the Future*, ed. H. Strasser, K. Kluth H. Rausch and H. Bubbs (Stuttgart: Ergonomia), 57–60.
- Mazaheri, A., Nieuwenhuis, I. L., van Dijk, H., and Jensen, O. (2009). Prestimulus  $\alpha$  and  $\mu$  activity predicts failure to inhibit motor responses. *Hum. Brain Mapp* 30, 1791–1800. doi: 10.1002/hbm.20763
- 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. *Hum. Factors* 54, 396–412. doi: 10.1177/0018720812442086
- Merritt, S. M., Heimbaugh, H., LaChapell, J., and Lee, D. (2013). I trust it, but I don't know why: effects of implicit attitudes toward automation on trust in an automated system. *Hum. Factors* 55, 520–534. doi: 10.1177/0018720812465081
- Merritt, S. M. (2011). Affective processes in human-automation interactions. *Hum. Factors* 53, 356–370. doi: 10.1177/0018720811411912
- Mulder, L. J. M. (1992). Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biol. Psychol.* 34, 205–236. doi: 10.1016/0301-0511(92)90016-n
- Nakagawa, S., Johnson, P. C. D., and Schielzeth, H. (2017). Coefficient of determination  $R^2$  and intra-class correlation coefficient from generalized linear

- mixed-effects models revisited and expanded. *J. R. Soc. Interface* 134:20170213. doi: 10.1098/rsif.2017.0213
- NHTSA's National Center for Statistics and Analysis. (2017). *2016 Fatal Motor Vehicle Crashes: Overview (Traffic Safety Facts Research Note. Report No. DOT HS 812 456)*. Washington, DC: National Highway Traffic Safety Administration.
- O'Connell, R. G., Dockree, P. M., Robertson, I. H., Bellgrove, M. A., Foxe, J. J., and Kelly, S. P. (2009). Uncovering the neural signature of lapsing attention: electrophysiological signals predict errors up to 20 s before they occur. *J. Neurosci.* 29, 8604–8611. doi: 10.1523/JNEUROSCI.5967-08.2009
- Onnasch, L., Wickens, C. D., Li, H., and Manzey, D. (2014). Human performance consequences of stages and levels of automation: an integrated meta-analysis. *Hum. Factors* 56, 476–488. doi: 10.1177/0018720813501549
- Parasuraman, R., and Manzey, D. H. (2010). Complacency and bias in human use of automation: an attentional integration. *Hum. Factors* 52, 381–410. doi: 10.1177/0018720810376055
- Parasuraman, R., and Riley, V. (1997). Humans and automation: use, misuse, disuse, abuse. *Hum. Factors* 39, 230–253. doi: 10.1518/001872097778543886
- Patel, M., Lal, S. K. L., Kavanagh, D., and Rossiter, P. (2011). Applying neural network analysis on heart rate variability data to assess driver fatigue. *Expert Syst. Appl.* 38, 7235–7242. doi: 10.1016/j.eswa.2010.12.028
- SAE International. (2016). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. Michigan: SAE International
- R Core Team. (2017). “R: A language and environment for statistical computing,” in *R Foundation for Statistical Computing* (Austria: Vienna), Available online at: <http://www.R-project.org/>
- Sarter, N. B., Woods, D. D., and Billings, C. E. (1997). “Automation surprises,” in *Handbook of Human Factors and Ergonomics*, ed. G. Salvendy (New York, NY: Wiley), 1926–1943.
- Simon, M., Schmidt, E. A., Kincses, W. E., Fritzsche, M., Bruns, A., Aufmuth, C., et al. (2011). EEG  $\alpha$  spindle measures as indicators of driver fatigue under real traffic conditions. *Clin. Neurophysiol.* 122, 1168–1178. doi: 10.1016/j.clinph.2010.10.044
- Singh, S. (2015). *Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey. (Traffic Safety Facts CrashStats. Report No. DOT HS 812 115)*. Washington, DC: National Highway Traffic Safety Administration.
- Smith, T. W., Cribbet, M. R., Nealey-Moore, J. B., Uchino, B. N., Williams, P. G., MacKenzie, J., et al. (2011). Matters of the variable heart: respiratory sinus arrhythmia response to marital interaction and associations with marital quality. *J. Pers. Soc. Psychol.* 100, 103–119. doi: 10.1037/a0021136
- Snyder, A. C., and Foxe, J. J. (2010). Anticipatory attentional suppression of visual features indexed by oscillatory  $\alpha$ -band power increases: a high-density electrical mapping study. *J. Neurosci.* 30, 4024–4032. doi: 10.1523/JNEUROSCI.5684-09.2010
- Stuiver, A., Brookhuis, K. A., de Waard, D., and Mulder, B. (2014). Short-term cardiovascular measures for driver support: increasing sensitivity for detecting changes in mental workload. *Int. J. Psychophysiol.* 92, 35–41. doi: 10.1016/j.ijpsycho.2014.01.010
- Stuiver, A., de Waard, D., Brookhuis, K. A., Dijksterhuis, C., Lewis-Evans, B., and Mulder, L. J. M. (2012). Short-term cardiovascular responses to changing task demands. *Int. J. Psychophysiol.* 85, 153–160. doi: 10.1016/j.ijpsycho.2012.06.003
- Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology. (1996). Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. *Circulation* 93, 1043–1065. doi: 10.1161/01.CIR.93.5.1043
- Thayer, J. F., Åhs, F., Fredrikson, M., Sollers, J. J. III., and Wager, T. D. (2012). A meta-analysis of heart rate variability and neuroimaging studies: implications for heart rate variability as a marker of stress and health. *Neurosci. Biobehav. Rev.* 36, 747–756. doi: 10.1016/j.neubiorev.2011.11.009
- van den Beukel, A. P., van der Voort, M. C., and Eger, A. O. (2016). Supporting the changing driver's task: exploration of interface designs for supervision and intervention in automated driving. *Trans. Res. Part F Psychol. Behav.* 43, 279–301. doi: 10.1016/j.trf.2016.09.009
- Victor, T. W., Harbluk, J. L., and Engström, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Trans. Res. Part F Psychol. Behav.* 8, 167–190. doi: 10.1016/j.trf.2005.04.014
- Wang, Y., Reimer, B., Dobres, J., and Mehler, B. (2014). The sensitivity of different methodologies for characterizing drivers' gaze concentration under increased cognitive demand. *Trans. Res. Part F Psychol. Behav.* 26, 227–237. doi: 10.1016/j.trf.2014.08.003
- Wilson, G. F. (1992). Applied use of cardiac and respiration measures: practical considerations and precautions. *Biol. Psychol.* 34, 163–178. doi: 10.1016/0301-0511(92)90014-1
- Wilson, G. F., and Russell, C. A. (2003a). Operator functional state classification using multiple psychophysiological features in an air traffic control task. *Hum. Factors* 45, 381–389. doi: 10.1518/hfes.45.3.381.27252
- Wilson, G. F., and Russell, C. A. (2003b). Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. *Hum. Factors* 45, 635–644. doi: 10.1518/hfes.45.4.635.27088
- Worden, M. S., Foxe, J. J., Wang, N., and Simpson, G. V. (2000). Anticipatory biasing of visuospatial attention indexed by retinotopically specific  $\alpha$ -band electroencephalography increases over occipital cortex. *J. Neurosci.* 20:RC63. doi: 10.1523/JNEUROSCI.20-06-j0002.2000
- Zeeb, K., Buchner, A., and Schrauf, M. (2015). What determines the take-over time? An integrated model approach of driver take-over after automated driving. *Accid. Anal. Prev.* 78, 212–221. doi: 10.1016/j.aap.2015.02.023
- Zhang, J., and Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika* 70, 203–212. doi: 10.1007/s11336-003-1119-8
- Zhang, J., Yin, Z., and Wang, R. (2017). Design of an adaptive human-machine system based on dynamical pattern recognition of cognitive task-load. *Front. Neurosci.* 11:219. doi: 10.3389/fnins.2017.00129

**Conflict of Interest Statement:** RM was employed by company Northrop Grumman.

The reviewer BS and handling Editor declared their shared affiliation.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Cisler, Greenwood, Roberts, McKendrick and Baldwin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Associating Vehicles Automation With Drivers Functional State Assessment Systems: A Challenge for Road Safety in the Future

Christian Collet<sup>1\*</sup> and Oren Musicant<sup>2</sup>

<sup>1</sup> Inter-University Laboratory of Human Movement Biology (EA 7424), Univ Lyon, Université Claude Bernard Lyon 1, Villeurbanne, France, <sup>2</sup> Department of Industrial Engineering and Management, Ariel University, Ariel, Israel

## OPEN ACCESS

### Edited by:

Karel Brookhuis,  
University of Groningen, Netherlands

### Reviewed by:

Luca Longo,  
Dublin Institute of Technology, Ireland  
Janet Lubertha Veldstra,  
University of Groningen, Netherlands

### \*Correspondence:

Christian Collet  
Christian.collet@univ-lyon1.fr

**Received:** 15 June 2018

**Accepted:** 01 April 2019

**Published:** 24 April 2019

### Citation:

Collet C and Musicant O (2019)  
Associating Vehicles Automation With  
Drivers Functional State Assessment  
Systems: A Challenge for Road Safety  
in the Future.  
*Front. Hum. Neurosci.* 13:131.  
doi: 10.3389/fnhum.2019.00131

In the near future, vehicles will gradually gain more autonomous functionalities. Drivers' activity will be less about driving than about monitoring intelligent systems to which driving action will be delegated. Road safety, therefore, remains dependent on the human factor and we should identify the limits beyond which driver's functional state (DFS) may no longer be able to ensure safety. Depending on the level of automation, estimating the DFS may have different targets, e.g., assessing driver's situation awareness in lower levels of automation and his ability to respond to emerging hazard or assessing driver's ability to monitor the vehicle performing operational tasks in higher levels of automation. Unfitted DFS (e.g., drowsiness) may impact the driver ability respond to taking over abilities. This paper reviews the most appropriate psychophysiological indices in naturalistic driving while considering the DFS through exogenous sensors, providing the more efficient trade-off between reliability and intrusiveness. The DFS also originates from kinematic data of the vehicle, thus providing information that indirectly relates to drivers behavior. The whole data should be synchronously processed, providing a diagnosis on the DFS, and bringing it to the attention of the decision maker in real time. Next, making the information available can be permanent or intermittent (or even undelivered), and may also depend on the automation level. Such interface can include recommendations for decision support or simply give neutral instruction. Mapping of relevant psychophysiological and behavioral indicators for DFS will enable practitioners and researchers provide reliable estimates, fitted to the level of automation.

**Keywords:** driver functional state, automated vehicles, monitoring, drowsiness, level of automation, activation level, vigilance, road safety

## INTRODUCTION: THE PROMISE OF AUTOMATED VEHICLES

Road traffic crashes represent a leading cause of death world-wide, more than 1.35 million lives each year, 48% of them in four-wheeled vehicles in Europe (World Health Organization, Global Status Report on Road Safety – Summary, 2018, pp. 2 and 6). Driving is a highly complex activity requiring considerable perceptual, physical, and cognitive demands on the driver (Sawyer et al., 2012) despite each of us has learned to drive a car. The human nervous system shows limitations

in controlling much information in parallel and the human driver is one of main factors in over 90% of the crashes (Sabey and Staughton, 1975; Treat, 1977; Hendricks et al., 2001; Otte et al., 2009; Singh, 2015).

With the aim to increase safety, Advanced Driving Assistance Systems (ADAS) have progressively been integrated into vehicles and can either warn the driver or actively intervene in the vehicle operation. Many systems are now likely to assist the drivers both in usual driving (e.g., cruise control or electronic stability program) and in critical situations (e.g., antilock braking system, collision avoidance system). Merat and Lee (2012) considered that the automation process is now inevitable, and rapidly evolving vehicle automation will change vehicles more in the next 5 years than during the preceding fifty, until the driver may no longer be needed (Ivanco, 2017). To date, Waldrop (2015) underlined that automation is one of the main topics that could yield completely driverless cars within the next decade.

Until driverless cars are available, there is an urgent need to consider the effect of increasingly automated vehicles on the ability of drivers to operate the vehicle, monitor both environment and automation, and efficiently take over driving responsibility. These tasks require allocating mental resources to help the process of information from multiple cues (e.g., the environment, in-cabin signals). Ironically, while automation may free the driver from some of the traditional driving tasks, new operations are added (monitoring automation, responding to “take over” requests) and attention (the main focus of the driver mental resources) is expected to more frequently be directed to secondary tasks (Jamson et al., 2013; Llaneras et al., 2013). Thus, it is likely that automation will have mixed effects on the amount of mental resources drivers are now required to allocate. Hockey et al. (2003) refer to the general concept of “operator functional state” dealing with the operator ability to allocate the required resources to meet the task demands. The overall load originating from such demands impacts the operator functional state. Determining the extent by which the driver functional state (DFS) is suitable for the current driving challenge is most imperative.

The recording of physiological indices seems appropriate while considering the level of automation, but also environmental conditions (e.g., traffic density, type of roadways or weather conditions), driver characteristics (e.g., driving experience, automation intrusiveness, and trust in automation). All the aforementioned categories are likely to influence the DFS. The importance of selecting the appropriate physiological indices determines the reliability of assessing the DFS accurately. Future vehicles will need to incorporate a DFS estimation system that can potentially support interventions to maintain safety. Some examples for such interventions include switching to a more acceptable level of automation, issuing alerts to the driver or nearby road users, and applying interventions to increase arousal.

The main objective of this article is to review how associating vehicles automation with drivers functional state assessment systems. This literature review will be organized along with the five following research areas: We will first describe how different levels of vehicle automation should mediate the allocation of attentional resources to driving. The next section will detail

the available methods of assessing the DFS. The complexity of assessing the DFS should point out the need to rely on different methodological solutions that must be integrated into a unique system. We will then propose a multimodal dataset acquisition requiring a close collaboration between the fields of engineering and behavioral neurophysiology thus leading to the redefinition of usual theoretical models. The whole of the preceding analysis will also have to take into account the singular characteristics of the drivers but also the external driving conditions. We will conclude by highlighting the contributions of our study to better understanding the relationships between vehicles automation and drivers functional state. We will also underline its limitations by acknowledging the path that remains to be done before we can propose complete autonomous driving solutions. This will not be done without the close collaboration between engineering sciences, neurophysiological and behavioral sciences.

## Levels of Automation and Allocation of Mental Resources

The Society of Automotive Engineers (SAEs) ranges vehicles automation capabilities from no automation (level 0) to complete automation (level 5). Level 0 accounts for most vehicles on the road today, where all driving tasks are manually handled. In level 1 (driving assistance), the vehicle has a single aspect of automation that assists the driver. Such automation level control either steering, speed (e.g., adaptive cruise control), or braking (e.g., automated emergency braking), but no more than one of these. In level 2 (partial automation), the vehicle can control both the steering and acceleration/deceleration, although the driver must always remain in complete control of the vehicle. This includes, among others, helping vehicles to stay in lanes and self-parking features. In level 3 (conditional automation) vehicles can make decisions for themselves such as overtaking slower moving vehicles. However, unlike the higher rated autonomous vehicles, this requires human override when the vehicle is unable to execute the task, or when the system fails. In this level, the driver must monitor automation and allocate attention to the driving as no information is provided about system failure. Level 4 (high automation) differs from level 3 in the sense that vehicles can intervene themselves in case of system failure. Thus, level 4 vehicles do not need human intervention in specific situations and will inform the driver on the need to take over in other situation as in occurrences of system breakdown or somehow underperformed or when in unfamiliar conditions (e.g., off-road driving, extreme weather). In level 5, complete automation does not require human interaction. Level 5 vehicles provide a much more responsive and refined service. These include off-road driving and other terrains that level 4 vehicles may not necessarily be able to detect or intelligently comprehend. In sum, the vehicle ability to monitor and “understand” the vehicle surroundings determines the level of automation. The main leaps in automation is between levels 2 and 3 in which the vehicle is already able to take complex tactical maneuvering decisions (e.g., changing lanes), and between levels 3 and 4 when human interaction is, in some circumstances, not required.

Whether one accepts the SAE scale of automation or proposes a different one, the discussion on the safety benefits of automation should consider the level of automation. While there is a broad agreement on the generally positive effect of automation, not all agree on the magnitude of this effect. As early as, Young and Stanton (2002) underlined that vehicle automation systems could reduce the required mental resources for driving and preserve safety by allowing the drivers to delegate some of their actions to the driving automation system. Therefore, drivers' functions are shifting from operating their vehicles to supervising their automation (Shen and Neyens, 2017) and would require a lower level of general activation in the central nervous system and a more relaxed functional state. It is thus believed that monitoring a system cost less than operating it. However, no real comparison of the involvement of mental resources has been provided by the scientific literature and workload may be higher since the driver is now responsible for monitoring not only the environment but also the way in which the vehicle operates. Monitoring a highly complex system without a situated mental model or the requisite diagnostic skills may be proven challenging. Caldwell et al. (1994) defined 'vigilance' as the "sustained readiness to detect and respond to changes in the environment" (p. 14) and linked it to general arousal. On the one hand, arousal impacts vigilance in the sense that we cannot be vigilant if we are not sufficiently aroused. On the other hand, being activated does not imply that we adequately orient our attention toward useful indices, while inhibiting competing indices (distractors). People who actively generate responses in a system have greater situation awareness than those who passively monitor the same outputs performed by an automated agent (Metzger and Parasuraman, 2001). Many studies pointed out the risk for disengagement and distraction from the road scene and the driving task (Lewis et al., 2018). Increases in automation reduced driver vigilance as shown by braking reaction time, emergency steering (Saxby et al., 2013), and in decreased ability to maintain lane position (Shen and Neyens, 2017). Young and Stanton (2007) also observed decrements in attentional resources negatively affecting driving performance. Another aspect of impaired vigilance is the possible increasing involvement in secondary tasks (Shen and Neyens, 2017) that would possibly increase the whole allocation of mental resources but not due to the requirements of the main task.

The above review suggests that driver capacities as maneuvering, managing secondary tasks, situational awareness, vigilance in monitoring automation, and responding to take-over requests at least partly depend on the DFS. We argue that estimating the DFS (as we subsequently described in section "Estimating the DFS") may have different strategies depending on the level of automation. To develop this argument, we refer to **Figure 1** presenting three radar subplots, each corresponding to a different level of automation. Each radar subplot specifies a list of driving capacities (maneuvering, situational awareness...). Black line indicates the level of capacity that is required in each of the selected driving aspects. The **Figure 1** presents how, with increased automation, maneuvering (i.e., correctly perform basic driving actions as braking and accelerating) and situational awareness capacities are becoming less and less

required. The **Figure 1** also presents the capacity of the driver according to his functional state (in blue).

If the DFS allows greater driving capacity (in blue) than what is required (in black), the probability of a crash remains low. However, a sudden increase in required capacity will also increase the risk of a critical situation. As the DFS can change from time to time, the reader should view the information suggested by the figure as an example for an arbitrary driver in an arbitrary time. To demonstrate that the figure presents plausible scenarios, we added references (indicated by the brackets []) for studies indicating when DFS (in blue) did not meet the requirements (in black). But clearly, more research is needed to accurately detect the relevant driving aspects, and their required capacities in the various automation levels. The information in **Figure 1**, therefore remains a schematic illustration of a possible future. Merat and Lee (2012) have also pointed out that little research has considered the consequences of high level of automation with most focusing on the effects of specific ADAS as lane-keeping or speed control (adaptive cruise control). This is an important concern despite some optimistic viewpoints (Merat and Lee, 2012; Waldrop, 2015), at this stage of autonomous vehicles development, automated driving is not yet reliable and safe (Dixit et al., 2016). Thus, research should study different levels of automation and accurately evaluate the effects of each on the DFS and consequently on drivers' performance. For example, Eriksson and Stanton (2017b) tried to determine the time drivers needed to take-over control from a highly automated vehicle when confronted with non-critical driving scenarios.

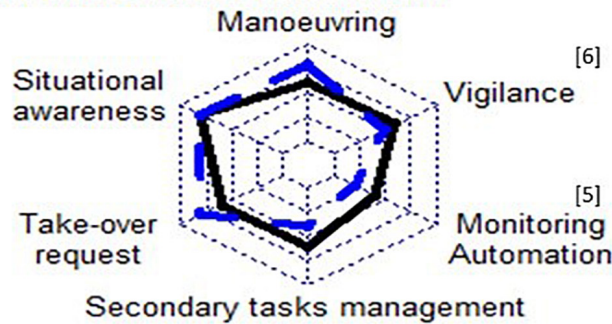
As described in **Figure 2**, the ability to take-over is not required in automation levels 0 and 4 but may prove critical in levels 2 to 3. Whether the DFS is well-adapted when the need to take-over occurs is one of the key-points determining the "DFS/levels of automation" interrelationships. Several hypotheses may be stated:

- The driver's arousal level decreases under the automated mode as there is no need for him (her) to be aroused at that time. Taking-over may thus take more time than when the driver is less aroused and this is generally observed when taking-over occurs on urgent scenarios which could not be anticipated (Eriksson and Stanton, 2017a).
- Conversely, monitoring the automated system working may be demanding and may require an arousal level higher than at rest. In this case, the relationships between DFS and levels of automation are rather complex. When drivers were under reduced time constraint, Eriksson and Stanton (2017a) observed that taking over did not affect driving performance, however, with a large standard deviation thus attesting strong inter-subjects variations in behavior.

## ESTIMATING THE DFS

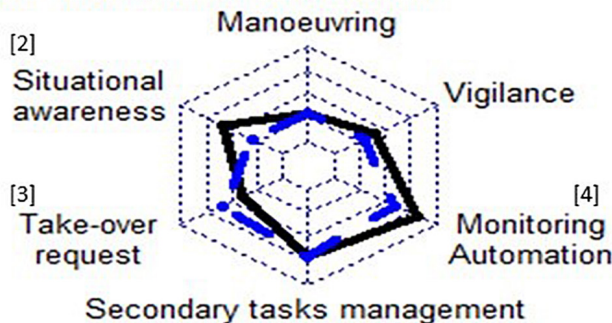
Estimating the DFS can take several approaches: in low automation levels, the DFS is visible by monitoring kinematic indices of driving. Such indices are based on vehicle dynamic, e.g., the intensity of braking events, driving speed, lane position and



**Level 1 and 2 Partial automation**

[1] with increase in automation secondary task engagement is becoming more frequent (Carsten et al, 2012, Jamson et al. 2013, Llaneras et al., 2013) and less demanding (de Winter et al. 2014)

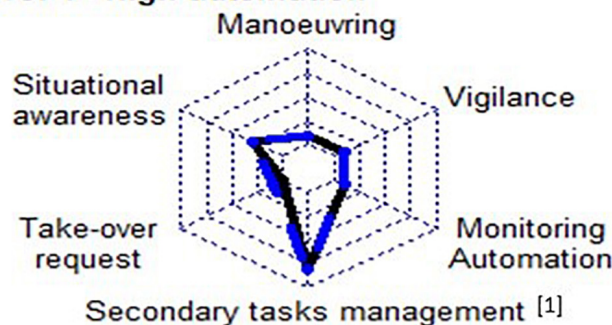
[2] In a series of simulator studies drivers in highly automated vehicle were less likely to gaze at the road center than manual drivers (de Winter et al. 2014, page 205)

**Level 3 - Conditional automation**

[3] Quality of take over performance depended on the secondary task (Zeeb et al. 2016)

[4] With an increasingly autonomous functionality, the driver now needs to monitor not only the environment but also the vehicle operation (Noy et al., 2018).

[5] In a series of simulator studies drivers failed to brake on time when ACC failed. (de Winter et al. 2014).

**Level 4 - High automation**

[6] Neubauer et al. (2012) postulated that full automation is likely to induce a state of fatigue associated with loss of alertness, probably due to the time spent in monitoring the system.

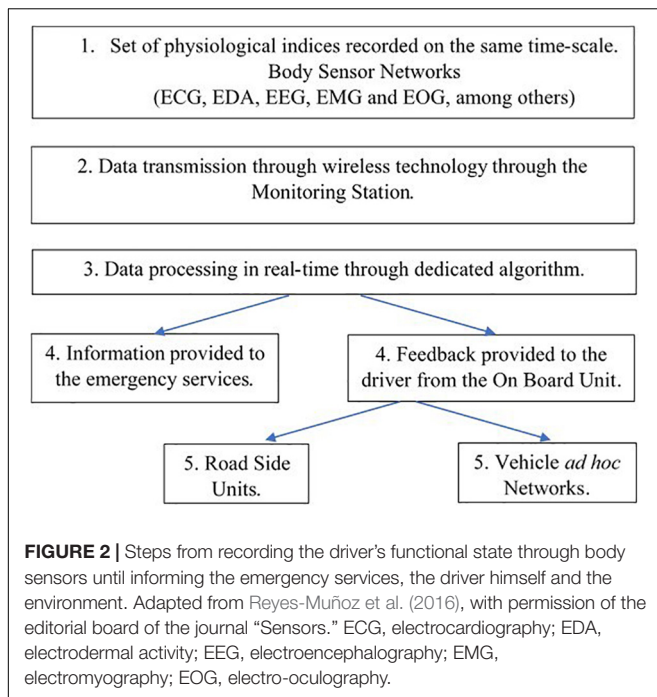
**FIGURE 1 |** Illustration of required capacities (black) and available capacities (blue) by level of automation (subplot). Level 5 is not present in the figure since driver involvement is not required in complete automation.

distance to the lead vehicle. However, with increasing automation some of these actions are automated and may not reflect the DFS. Thus, the automated system operates well while the DFS is with low levels. Another, and perhaps more direct approach to estimate the DFS aspects is to tap into driver physiological indices as heart rate (HR), heart rate variability (HRV), skin conductance, and electroencephalography (EEG).

There is a large body of research that links driving performance with physiological arousal which clearly influence sensorimotor performance (Hockey et al., 2003). We cannot perform well without being aroused enough because the arousal level (tonic activity of brain structures associated to adequate muscles activation) determines the choice of useful information, its processing, and the motor response to be then implemented

(Näätänen, 1973). Thus, functional state belongs to a conceptual framework including a quantitative dimension, i.e., energetic level supposing adequate (optimal) level of arousal which, in turn, influences a qualitative dimension, i.e., the ability to well process the information (adequate orientation of the attention, selection of useful cues, potential processing of concurrent information and inhibition of competing information). Boucsein and Backs (2009) elaborated an integrated model of arousal with four different levels, including sensory arousal, affective and memory arousal and arousal for action preparation. This is directly inspired from the earlier model by Näätänen (1973) supposing that performance directly depended upon both energetic and directional factors. On the basis of previous studies, general arousal is believed to impact behavioral efficiency since it involves





the ability to mobilize the energy of the organism to face task requirement. Thus, DFS may be described through tonic variations of physiological indices, i.e., quantitative dimension associated with phasic physiological variations of the same indices, thus attesting information perception and processing (see Näätänen, 1973 for historical reference and Caldwell et al., 1994, for defining the activation/vigilance interrelationships).

In this context, we have the potential to assess the cost of taking-over from a highly automated vehicle (SAE level 3 and 4), the time needed for this and the quality of taking the vehicle back in hands (Payre et al., 2016; Eriksson and Stanton, 2017b). Carsten et al. (2012) studied to which extent driver attention to the road scene was affected by the level of automation provided to assist or to take over the basic task of vehicle control. Autonomous vehicles may thus be viewed with skepticism in their ability to improve safety when automated driving fails, or is limited, the autonomous mode disengages and the drivers are expected to resume manual driving (Dixit et al., 2016). An accurate and comprehensive approach to these factors is necessary to assess their effects on DFS. Thus, studying human-automated system interaction should consider the need to maintain attention during prolonged periods. In this context, the ability to detect and respond to rare and unpredictable events is of highly importance (roadway hazards that automation may be ill equipped to detect, according to Greenlee et al., 2018). Recording DFS at the same time would allow to verify whether it is adapted for safely driving (during both continuous monitoring and periods where taking-over is necessary). Finally, we should also include environmental factors in our analysis, e.g., the impact of traffic density and any additional task which could be performed simultaneously by the driver in highly automated driving (Zeeb et al., 2016). Here, we see that DFS determination

depends on variable factors that are relatively difficult to identify. This tends to complicate the linking of the DFS with the level of automation of the vehicle.

In the following section, we will consider two main challenges:

- (i) Which physiological indices are the best candidates to determine the DFS under naturalistic conditions?
- (ii) How integrate redundant information into the recording system, redundancy ensuring its reliability?

A related requirement would be to eliminate false positives and negatives. If not, this will reduce driver's trust in the system, or worse, drivers will consider the system unreliable. In this context, neuroergonomics<sup>1</sup> can provide heuristic solutions since physiological indices can give useful information about DFS while being easily recordable with low intrusiveness. We could thus restrict the potential candidates to some central and peripheral indices (Lee et al., 2007; Clarion et al., 2009; Fernández et al., 2016).

## Physiological Indices From the Brain

At the central level, we should only consider ambulatory methods and not those from functional neuro-imagery (fMRI, MEG). Several tools with the ability to be used inside the vehicles are now available, e.g., electroencephalography (EEG – Lin et al., 2014; Damian et al., 2015) and functional near infra-red spectroscopy (fNIRS – Liu et al., 2016; Wang et al., 2016). EEG and fNIRS can provide information about DFS as they directly record intrinsic signals from the brain. Functional NIRS measures the cerebral microcirculation in the capillary networks and describes brain activations during actual driving sessions in real environments (Liu et al., 2016). Although it is premature to conclude that fNIRS will soon be integrated into real-time monitoring of DFS, several studies reported experimental designs both in simulated and actual driving (Liu et al., 2016; Wang et al., 2016).

Tonic variations of EEG waves are closely correlated to arousal states and can detect changes in brain activation. This is a real challenge to record EEG from inside vehicles (Caldero-Bardaji et al., 2016). Papadelis et al. (2007) requested sleep-deprived participants to drive in real field driving conditions and observed increase in brief paroxysmal bursts of alpha activity prior to severe driving errors. Anticipated EEG alpha bursts thus correlated with the risk to be involved in car crash. Damian et al. (2015) used mobile EEG to estimate the mental effort during a dual-task paradigm with EEG signal sent from wireless sensors during driving. Lin et al. (2014) assessed changes in drivers' arousal, fatigue, and vigilance with reference to variations in task performance, by evaluating associated EEG changes. The same team (Lin et al., 2010) developed a brain-computer interface integrating a dual module for physiological-acquisition and signal processing. The embedded modules can monitor DFS in real time and provide biofeedback to the driver as early as the drowsy state occurs. Wireless sensors associated with real-time data acquisition/processing, and with a dedicated

<sup>1</sup>Mehta and Parasuraman (2013) defined neuroergonomics as an emerging science studying human brain indices in relation to performance in a workplace and everyday settings.

algorithm are the main tools of a system monitoring DFS. One remaining concern is related to sensors themselves as conventional physiological measurements techniques required to have the sensors in close contact with the human body. These could nevertheless interfere with driving operations as body segments can come into contact with some elements and as these are very sensitive to noise and artifacts (mainly caused by head movements). Sun and Yu (2014) described a non-intrusive driver assistance system which is likely to detect ECG or EEG signals through clothes or hair without direct skin-contact. Thus, the last feature of a brain-computer interface would be to remotely detect the physiological signals with no physical contact with human skin. The near future will probably see the development of such systems. We must acknowledge that asking drivers to affix sensors on their skin could be perceived as constraining, by the potential inconvenience to driving, by the time spent placing sensors, the latter may be made more difficult by the wearing of certain clothes. Considering that drivers would be required to wear a recording device on the head, which would be a prohibitive constraint for many people, data from EEG and  $f$ NIRS have low practical properties at that time. However, they can be supplemented by information from the peripheral nervous system, in particular the autonomic and motor nervous systems.

## Peripheral Physiological Related to Driving Performance

Several indices from the autonomic nervous system (ANS) are sensitive to time-dependent variations in arousal level and to external stimuli (Clarion et al., 2009; Brookhuis and de Waard, 2010; Johnson et al., 2011; Rigas et al., 2011). As the systems recording ANS activity are ambulatory and weakly intrusive, these are good candidate for DFS assessment (Rada et al., 1995; Axisa et al., 2004; Ramon et al., 2008). HR and electrodermal activity (EDA, skin conductance) increase with each incremental increase in cognitive demand (Mehler et al., 2012) and are closely related to functional state (Hugdahl, 1996). Among others, Porges (1995) and Hugdahl (1996) early promoted the role of the ANS in cognition. Porges (1995, for a review) underlined the role of the parasympathetic branch and particularly the vagus nerve on attentional processes. Several indices from the peripheral motor system respect the aforementioned criteria and may be pooled into three main categories (i) indices from electromyography (EMG) monitoring, with a special focus on muscles from the neck and the back of the driver, (ii) indices from the oculomotor system aimed at giving information on palpebral, dilation of pupils and eye-gaze related features, and (iii) indices from facial mimics through emotional face recognition.

Heart rate is a very easily recordable variable even without bodily placed sensors. Lee et al. (2007) elaborated a non-intrusive measurement of HR by integrating dry sensors into the steering wheel with a wireless design for data transmission (the safety belt can also provide a naturalistic way for recording HR). No differences from usual HR recordings were found with the design the authors conceived thus attesting its reliability. Beside the basal values, HRV has close links with fatigue and drowsiness detection (Li and Chung, 2013). Yu-Lung et al. (2016)

recorded ECG from wireless thoracic sensors and process the cardiac signal using HRV. Several parameters (e.g., low-frequency power spectrum over high-frequency power spectrum or LF/HF ratio) were closely correlated to several changes in drivers' behavior, particularly with the frequency of yawning episodes. By comparison with rest state or high level of arousal, HRV presents specific alterations during drowsiness episodes (Vicente et al., 2016). The authors claimed that incorporating drowsiness assessment on the basis on HRV signal may improve the existing car safety systems.

Electrodermal activity is closely related with arousal as it is directly under the control of the sympathetic endings innervating sweat glands without any influence of the parasympathetic branch, thus derogating from the well-described principle of double innervation (Collet et al., 2013). Importantly, EDA is a witness of sympathetic functioning alone. By confronted drivers to an unexpected critical crash avoidance situation, Collet et al. (2005) showed that EDA was a predictive index of drivers' performance. The recording of EDA basal level along the whole session evidenced that drivers who avoided the obstacle pulled onto their traffic lane where those who exhibited the highest EDA basal values (about 30% above the reference EDA at rest). Conversely, the drivers who failed to avoid the obstacle showed a lower EDA level, at about 20% above the reference level at rest. Thus, drivers who performed well exhibited higher arousal and were more likely to perform adequately. More generally, when considering routine driving situations, there is a close positive relationships between EDA and cognitive demands (Mehler et al., 2012). Other indices can originate from basal EDA signal, e.g., the frequency of electrodermal responses was positively associated with decreased vigilance (Dementienko et al., 2001). When the drivers exhibited obvious signs of low vigilance, electrodermal response frequency decreased in parallel. We should nevertheless indicate that Dorrian et al. (2008) failed to evidence a relationship between EDA and participants state who were imposed one night of sustained wakefulness. While they rated increased levels of sleepiness and fatigue through paper and pencils tests, EDA did not present any difference between the reference period and the induced sleepiness and fatigue state. EDA usually range from 1.5 to 70  $\mu$ Siemens and data processing should be done with caution due to the high differences among people. Preventing metrologic errors due to individual differences may easily be overcome by normalizing data. Another way to increase reliability is to simultaneously record other physiological indices. This is usually done when experiments are designed to study complex human brain functions, such as DFS. There are thus many contributions presenting a data set of physiological indices (Rada et al., 1995; Ramon et al., 2008; Clarion et al., 2009; Lanatà et al., 2015; Taamneh et al., 2017). Lanatà et al. (2015) evaluated DFS by analyzing ANS changes through HR, EDA, and respiratory frequency along with performance indices of steering wheel angle corrections and response time. This study was performed under simulated driving conditions, but Healey and Picard (2005) already provided evidence of physiological recordings under actual driving conditions. They reported that EDA and HR were the most closely correlated with driver strain. Physiological monitoring could thus provide a continuous

assessment of how different driving contexts but also driver emotional states affect DFS. These studies clearly show the ability of peripheral physiological variables to closely correlate with the DFS. They can be supplemented by behavioral variables.

## Behavioral Indices Related to Driving Performance

Reyes-Muñoz et al. (2016) identified five behavioral indices that are close correlated to drowsiness, i.e., frequent yawning, frequent eye-blinking, pupil movement (gaze), head movement and facial expression. Fernández et al. (2016) also provided a thorough review focused on the role of computer vision technology applied to the development of monitoring systems. They considered that seven factors could evaluate the DFS with highly acceptance: reliability, real-time performance, low cost, small size, low power consumption, flexibility, and short time-to-market.

Yu-Lung et al. (2016) elaborated an intelligent driver assistance system including a camera in front of the driver for facial monitoring. Frequency of yawning was one of the main index predicting the occurrence of drowsiness (see also Sigari et al., 2014). Fernández et al. (2016) considered that the eyes are the most remarkable information sources in face analysis as they reflect affective states and focus of attention. There are nevertheless several methodological obstacles to overcome before providing a reliable set of information from the visual system (e.g., keeping the camera closely orienting on the eyes despite head movements). Song et al. (2013) described the main factors challenging accurate eyes localization, due to variations in facial expressions, variations of gaze direction, head/eyes movement coordination and surrounding lighting. The measures may be hindered by the wearing of glasses especially sunglasses and makeup (Fernández et al., 2016). Eye-blink and eyelid closure are of interest in detecting early signs of drowsiness, as these may be captured by a set of cameras placed on the dashboard (Hu and Zheng, 2009) and blinking has been reported to change during cognitive distraction phases (Fernández et al., 2016). Data acquisition and processing are provided by the seeing machines which *"continuously measure operator eye and eyelid behavior to determine the onset of fatigue and micro sleeps and deliver real-time detection and alerts"* (Fernández et al., 2016, p. 25 of 44).

Recordings of EMG activity have a high potential to bring information about the DFS. The alteration of muscles function may be associated with impairment in driving abilities and fatigue. Surprisingly, there are little scientific contributions from this field. Fu and Wang (2014) showed that the peak factor and the maximum of the cross-relation curve, two indices from surface EMG of the biceps femoris, were related to drivers' fatigue. EMG recorded from the neck and the back muscles are likely to provide information about sleepiness and driver fatigue. However, muscles activity is difficult to capture given the driver's sitting position, with the risk of sensors contact with the seat or headrest, thus affecting data reliability. Finally, head movements recordings by embedded cameras can provide similar information to that provided by EMG. Methodological difficulties may explain the weak number of works involving

EMG in actual driving. Despite behavioral indices of drowsiness occurrence are promising methods, Sahayadhas et al. (2012) underlined that the reliability and accuracy of driver drowsiness detection by a set of physiological indices is higher than that coming from other methods such as vehicle-based measures and behavioral measures.

## AN OBVIOUS REQUIREMENT: A MULTIMODAL DATASET ACQUISITION

Beside the methods used in laboratories, the challenge is to propose pragmatic, integrated systems, including a set of behavioral and physiological indices, simultaneously recorded in real time, both from the driver and the environment. This involves selecting indicators for their reliability and complementarity. Maglione et al. (2014) simultaneously recorded high resolution EEG data associated with heart and eye blinks rates. Then, fusion of data provided a robust method in studying complex human activities, involving several functions (Noori and Mikaeili, 2016). Rigas et al. (2011, 2012) described a set of physiological signals (ECG, EDA, and respiration) associated with driving history from the GPS and the vehicle's controller area network-bus (CAN) data. They incorporated these data into a Bayesian network (BN) and estimated that the system could detect stressful events with an accuracy of 82%. The development of an intelligent algorithm capable of recognizing the drivers' affective state was proposed by Singh et al. (2013). It was based on several physiological indices including EDA and blood flow through photo-plethysmography during on-road driving. Their neural networks are believed to predict DFS with a nearly 90% average precision. According to Reyes-Muñoz et al. (2016), recording physiological variables for DFS assessment could allow rescuers to make a faster and more accurate diagnosis in case of an accident, if the data is transmitted to the rescue services (**Figure 2** summarizes the successive steps from data acquisition/processing until provided feedback to the driver and eventually to the road control or emergency services).

## DRIVERS' INDIVIDUAL FEATURES AND EXTERNAL CONDITIONS

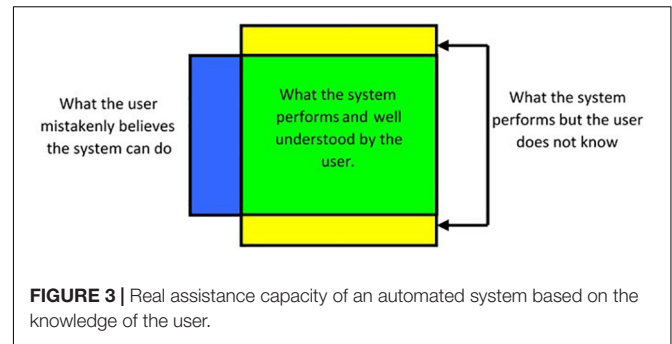
In addition to the variables used to evaluate the DFS, we must take into account two intrinsic factors, the individual characteristics of the drivers and the external driving conditions. One of the main concerns in providing feedback to the drivers is their high behavioral variability. There is thus considerable dispersion around the median behavior depending upon driver's characteristics in age, gender, driving experience and perhaps more importantly their psychological particularities or specific individual traits. Ranney (1994) underlined that the inherent variability of human behavior may be responsible of errors associated with an important rate of roadway crash causation. By comparison, systematic errors attributable to the well-known limits of the human information-processing system seems



rarer. In fact, all driving activities are believed to associate fast sensorimotor and automatic components with slower and more deliberate controlled cognitive processes. This refers to intra-individual behavioral variations as a function of time (according to specific individual differences, mental state and environmental context). Determining the boundaries around which the automated system could provide useful information to the drivers, i.e., with a high probability to take it into account for its meaning, is a key-component to be resolved in the next future. Little is known about how personality traits lead people to consider or ignore a given information. If the personality traits are stable features, and can be taken into account, the emotional states are more transient and therefore more difficult to detect. Yet, we know that they influence driving (Chan and Singhal, 2013) with a high probability of diverting the driver from the road scene. Another important concern is about how old drivers perceived the integration of more and more automated devices into their vehicles. First, Molnar and Eby (2017) question their motivation for technology use and assigned meanings. Second, they wonder whether the in-vehicle monitoring technology will be used and how transfer of control between automated and manual driving would occur in the elderly population. The role of trust in automation and its interaction with practice of partially or fully automated vehicles is also a key-variable. Payre et al. (2016) observed that drivers who had high trust in the automated vehicle exhibited longer reaction time when they were required to take-over by manual control recovery. Thus, over-trust may have deleterious effect on performance, a well-known effect of what high-technology is believed to bring (Collet et al., 2005). Overconfidence in vehicle equipment made drivers less efficient and this correlated well with a weak arousal level. This is well summarized by Endsley (2017): “*more autonomy is added to a system and its reliability and robustness increase, the lower the situation awareness of the driver and the less likely that he will be able to take over manual control if needed.*”

These examples clearly advocate for education in the use of automated systems. Strauch (2017) deplores that drivers have not gained enough expertise needed to effectively operate automated systems. Instead, they are forced to obtain the expertise *ad hoc* during system operations. We nevertheless suppose that the in-vehicle intelligent devices should identify the driver (through face identification), retrieve his previously stored profile from its data to then intelligently prescribe specific accident prevention tools and driving environment customizations, as proposed by Sawyer et al. (2012). At least, we should be informed and trained about how the automated device works so that we can improve take over whenever necessary. We should also change our representation about automated systems, as suggested by Figure 3 where the interactions with them can include three modes:

- The first is the adequacy when the driver understands some features of the system and uses them (in green).
- The second corresponds to the disjunction between the partial knowledge of system functionalities by the driver and the representation that s/he has some. In fact, the user does not



know some features of the automated system and thus can obviously not use them (in yellow).

- The third is false representations as the user wrongly thinks that the automated system can fill certain functions while it cannot (in blue).

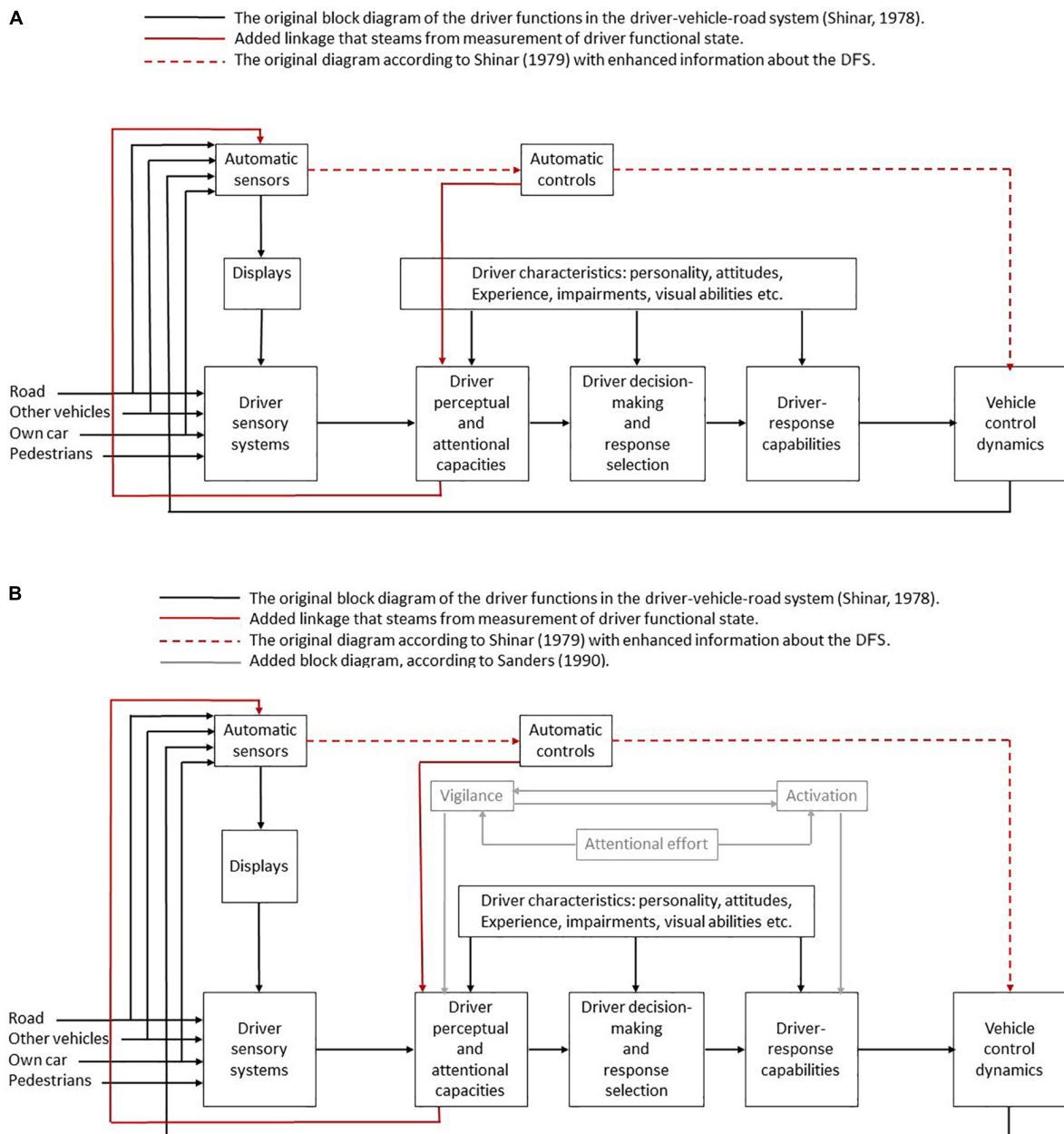
Reducing the discrepancy between drivers' representation of the system functioning and its actual abilities and functionalities (e.g., levels of automation) would probably imply to redefine the procedures of learning to drive.

## CONCEPTUAL CHANGES FOR CURRENT MODELS OF DRIVING PERFORMANCE AND LEARNING

Over the years, human factors research proposed several models for driver performance (Shinar, 1978; Michon, 1985; Endsley, 1995; Fuller, 2005; Wickens et al., 2015). These regard the driver as an information processing unit. Such information and attention models describe how the driver obtains data using his/her sensorial systems (vision, audition, etc.), process them to gain significant insights, apply a decision-making mechanism (e.g., slow down), adapt the decision to the actual context (e.g., adjust the braking intensity) and execute the decision with success determined by his/her abilities. According to these models, the driver limited capacity to collect and process the information from the environment explains driver error and misjudgment.

Here, we examine a model that was developed almost 40 years ago by Shinar (1978). We show how, in some ways, this model is still useful, and how it should be updated to incorporate new abilities to monitor the DFS. We explain how such updates may have potential safety benefits. In Figure 4A, the black lines depict the original connections in the model by Shinar (1978). The red dashed lines depict original connections that now serve to transfer information and driving decisions about the DFS as well as information and driving decisions that stem from knowledge about the environment. The red lines did not appear in the original model and represent new contributions. The original model (black lines) describes how the driver sensory system receives various cues, the information is then processed according to the driver perceptual and attentional capabilities to facilitate decision making and response. These cues do not





**FIGURE 4 | (A)** A limited-capacity model of driver information processing (adapted from Shinar, 1978) with added paths for DFS. Reproduction with permission of the author. **(B)** A hybrid limited-capacity model of driver information processing paths for DFS (adapted from Shinar, 1978 and Sanders et al., 1990). Here, we integrate all the factors that are supposed to make the DFS varying. The operations of information processing (perception, attention, memory access, decision making, and motor adjustments) require the mobilization of mental resources (attentional effort) and, thus change the DFS so that the driver is able to drive efficiently and safely. The DFS can be evaluated by a set of physiological and behavioral indices, and adjust as a function of the level of vehicle automation.

include DFS information. Next, the driver response impacts the vehicle dynamics. It is interesting that so long-ago, Shinar (1978) included a path for an autonomous system that can (1) display feedback to the driver (e.g., as done by level 0 systems as collision warning systems and navigation systems), and (2) control vehicle dynamics (e.g., as done by level 1 systems as adaptive cruise control and automatic emergency braking system). Despite the time passed since this model formulation, similar modes still

guide research teams in international meetings (Keith et al., 2005). The red lines represent the transfer of physiological cues related to the DFS. Determining the DFS can be based on ECG (HRV), EDA, and EEG. The driver himself may be aware of his physical and mental states through sensory feedbacks (e.g., fatigue, or other temporal impairments), and can decide to take measures to adjust (e.g., stop for a rest) or, in the near future, to engage an automated driving mode.

However, with the advances of physiological monitoring technology, these data can be picked up by electronic sensors (e.g., sit sensors or wearables). In automation level 0, the system can display this information back to the driver (e.g., a coffee icon can suggest that s/he should rest, other displays can warn about increase in mental effort, decrease in general activation or any other physiological condition). In levels 1–3, an automatic action can take two directions: the first is to increase driver perceptual and attentional capabilities. Numerous interventions can be suggested here. When mental workload is high, such intervention can lower the volume of the music, adjust air condition, shut off infotainment and message. In very low mental workload, the automated system can propose Trivia Games and even increase the volume of the music. On the second path, the autonomous system can control the vehicle dynamics to reduce demands from the environment. An interesting study by Hajek et al. (2013) investigated the safety benefit and acceptance of an adaptive cruise control that selected the optimal safe distance to the lead vehicle according to the DFS (estimated by the driver physiological indices). Another example is the ability to use physiological indices to predict intentions for emergency braking (Kim et al., 2015), such an ability may be use either to trigger the automatic emergency braking system (level 1) or to release a distress signal to the autonomous system (level 5) which can learn to avoid such stressful conditions for its driver (supervisor) in the future. In sections two and four, we mapped several physiological and behavioral indices that can be used to estimate these closely related aspects of driving. In **Figure 4B**, we offer that physiological indices for activation and vigilance also have a link for the automated sensors.

## CONCLUSION

Research into the effects of automation on DFS is expending due to understanding that in the near future, the human factor will remain an important component in driving and in monitoring automation. This manuscript points on: (1) The need of accurately assessing DFS, (2) Estimating the DFS may have different strategies given the level of automation, (3) Estimating the DFS can infer on interventions that are also related to automation (e.g., switching between levels of automation). Based on these understandings (points 1–3), we reviewed methods for estimating the DFS and described the potential characteristics of

an in-vehicle system. With regard to the first point, commuting executive functions usually performed by the driver to ADAS is likely to make her/him less concentrated on driving. The driver can monitor the system working or be engaged in other tasks with a connection with driving (supervising the route plans through the GPS) or not (reading or phoning or discussing with other passengers). Depending on these different activities, DFS may stay at a level comparable to that required for driving (parallel activity with the same demand as driving) or can change drastically and reach a level incompatible with driving (decrease in arousal level). This seems of particular importance in case of sudden need of taking-over. With regard to point 2, the extent by which driver's DFS remains at an adequate level is still pending and depends on different but interrelated factors (level of automation, driving conditions, driver's personality). With reference to point 3, an acceptable alternative would be to propose an intelligent system where we would choose the level of automation according to the objective of our trip (professional or leisure trip) of our state of fatigue (strong delegation or conservation of driving control) or conditions of external driving (traffic density, weather conditions). For example, traffic density has been shown as influencing the way in which the take-over is performed, with higher time to proceed and less accuracy when traffic is dense (Gold et al., 2016).

An integrated system capable of monitoring DFS in real time, should be based on several physiological indices recorded inside the vehicle. This would probably be the best way to ensure safety provided that it is built on sufficiently powerful algorithms capable of including all the driving scenarios that can potentially occur, thus depending on the external conditions (where driving takes place with traffic and weather conditions). It should also be able to provide useful feedback, from simple information about his functional state to even delivering graduated alerts, depending on their severity and urgency. Finally, we contributed to show how monitoring DFS can serve to update existing driving performance models to provide feedback to drivers and to automatically adjust autonomous behavior.

## AUTHOR CONTRIBUTIONS

OM and CC conceptualized and designed the study, organized the manuscript, and wrote the first draft. They contributed equally in writing the manuscript.

## REFERENCES

- Axisa, F., Gehin, C., Delhomme, G., Collet, C., Robin, O., and Dittmar, A. (2004). Wrist ambulatory monitoring system and smart glove for real time emotional, sensorial and physiological analysis. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 3, 2161–2164. doi: 10.1109/IEMBS.2004.1403632
- Boucsein, W., and Backs, R. W. (2009). "The psychophysiology of emotion, arousal and personality: methods and models," in *Handbook of Digital Human Modeling. Research for Applied Ergonomics and Human Factors Engineering*, ed. V. G. Duffy (Boca Raton, FL: CRC Press), 35.1–35.18.
- Brookhuis, K. A., and de Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. *Accid. Anal. Prev.* 42, 898–903. doi: 10.1016/j.aap.2009.06.001
- Caldero-Bardaji, P., Longfei, X., Jaschke, S., Reermann, J., Mideska, K. G., Schmidt, G., et al. (2016). Detection of steering direction using EEG recordings based on sample entropy and time-frequency analysis. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2016, 833–836. doi: 10.1109/EMBC.2016.7590830
- Caldwell, J. A., Wilson, G. F., Cetinguc, M., Gaillard, A. W. K., Gunder, A., Lagarde, D., et al. (1994). *Psychophysiological Assessment Methods*, Vol. 324. Brussels: North Atlantic Treaty Organization (NATO), 13–15.
- Carsten, O., Lai, F. C., Barnard, Y., Jamson, A. H., and Merat, N. (2012). Control task substitution in semi-automated driving: does it matter what aspects are automated? *Hum. Factors* 54, 747–761. doi: 10.1177/0018720812460246
- Chan, M., and Singhal, A. (2013). The emotional side of cognitive distraction: implications for road safety. *Accid. Anal. Prev.* 50, 147–154. doi: 10.1016/j.aap.2012.04.004

- Clarion, A., Ramon, C., Petit, C., Dittmar, A., Bourgeay, J. P., Guillot, A., et al. (2009). An integrated device to evaluate a driver's functional state. *Behav. Res. Methods* 41, 882–888. doi: 10.3758/BRM.41.3.882
- Collet, C., Di Rienzo, F., El Hoyek, N., and Guillot, A. (2013). Autonomic nervous system correlates in movement observation and motor imagery. *Front. Hum. Neurosci.* 7:415. doi: 10.3389/fnhum.2013.00415
- Collet, C., Petit, C., Priez, A., and Dittmar, A. (2005). Stroop color-word test, arousal, electrodermal activity and performance in a critical driving situation. *Biol. Psychol.* 205, 195–203. doi: 10.1016/j.biopsycho.2004.07.003
- Damian, A., Corona-Strauss, F. I., Hannemann, R., and Strauss, D. J. (2015). Toward the assessment of listening effort in real life situations: mobile EEG recordings in a multimodal driving situation. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2015, 8123–8126. doi: 10.1109/EMBC.2015.7320279
- De Winter, J. C. F., Happee, R., Martens, M. H., and Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: a review of the empirical evidence. *Transp. Res. Part F Traffic Psychol. Behav.* 27, 196–217. doi: 10.1016/j.trf.2014.06.016
- Dementienko, V. V., Markov, A. G., Koreneva, L. G., and Shakhnarovich, V. M. (2001). Automatic checkout systems to monitor driver's vigilance. *Crit. Rev. Biomed. Eng.* 29, 157–171. doi: 10.1615/CritRevBiomedEng.v29.i1.90
- Dixit, V. V., Chand, S., and Nair, D. J. (2016). Autonomous vehicles: disengagements, accidents and reaction times. *PLoS One* 11:e0168054. doi: 10.1371/journal.pone.0168054
- Dorrian, J., Lamond, N., Kozuchowski, K., and Dawson, D. (2008). The driver vigilance telemetric control system (DVTCS): investigating sensitivity to experimentally induced sleep loss and fatigue. *Behav. Res. Methods* 40, 1016–1025. doi: 10.3758/BRM.40.4.1016
- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Hum. Fact.* 37, 32–64. doi: 10.1518/001872095779049543
- Endsley, M. R. (2017). From here to autonomy. *Hum. Fact.* 59, 5–27. doi: 10.1177/0018720816681350
- Eriksson, A., and Stanton, N. A. (2017a). Driving performance after self-regulated control transitions in highly automated vehicles. *Hum. Fact.* 59, 1233–1248. doi: 10.1177/0018720817728774
- Eriksson, A., and Stanton, N. A. (2017b). Takeover time in highly automated vehicles: noncritical transitions to and from manual control. *Hum. Fact.* 59, 689–705. doi: 10.1177/0018720816685832
- Fernández, A., Usamentiaga, R., Carús, J. L., and Casado, R. (2016). Driver distraction using visual-based sensors and algorithms. *Sensors* 16:E1805. doi: 10.3390/s16111805
- Fu, R., and Wang, H. (2014). Detection of driving fatigue by using noncontact EMG and ECG signals measurement system. *Int. J. Neural Syst.* 24:1450006. doi: 10.1142/S0129065714500063
- Fuller, R. (2005). Towards a general theory of driver behaviour. *Accid. Anal. Preven.* 37, 461–472. doi: 10.1016/j.aap.2004.11.003
- Gold, C., Körber, M., Lechner, D., and Bengler, K. (2016). Taking over control from highly automated vehicles in complex traffic situations: the role of traffic density. *Hum. Fact.* 58, 642–652. doi: 10.1177/0018720816634226
- Greenlee, E. T., DeLucia, P. R., and Newton, D. C. (2018). Driver vigilance in automated vehicles: hazard detection failures are a matter of time. *Hum. Fact.* 60, 465–476. doi: 10.1177/0018720818761711
- Hajek, W., Gaponova, I., Fleischer, K. H., and Krems, J. (2013). Workload-adaptive cruise control – A new generation of advanced driver assistance systems. *Transp. Res. F Traffic Psychol. Behav.* 20, 108–120. doi: 10.1016/j.trf.2013.06.001
- Healey, J., and Picard, R. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Trans. Intell. Transp. Syst.* 6, 156–166. doi: 10.1109/TITS.2005.848368
- Hendricks, D. L., Fell, J. C., Freedman, M., Page, J. F., Bellis, E. S., Scheffle, T. G., et al. (2001). *The Relative Frequency of Unsafe Driving Acts Summary Technical Report*. Available at: <http://www.nhtsa.dot.gov/people/injury/research/UDAShortrpt/UDASummtechrept.pdf> (accessed February 26, 2019).
- Hockey, G. R. J., Gaillard, A. W. K., and Burov, O. (2003). *Operator Functional State: The Assessment and Prediction of Human Performance Degradation in Complex Tasks*, Vol. 355, Boston, MA: Ios Pr Inc., 392.
- Hu, S., and Zheng, G. (2009). Driver drowsiness detection with eyelid related parameters by support vector machine. *Expert Syst. Appl.* 36, 7651–7658. doi: 10.1016/j.eswa.2008.09.030
- Hugdahl, K. (1996). Cognitive influences on human autonomic nervous system function. *Curr. Opin. Neurobiol.* 6, 252–258. doi: 10.1016/S0959-4388(96)80080-8
- Ivanco, A. (2017). Fleet analysis of headway distance for autonomous driving. *J. Saf. Res.* 2017, 145–148. doi: 10.1016/j.jsr.2017.10.009
- Jamson, A. H., Merat, N., Carsten, O. M. J., and Lai, F. C. H. (2013). Behavioural changes in drivers experiencing highly-automated vehicle control in varying traffic conditions. *Transp. Res. C Emerg. Technol.* 30, 116–125. doi: 10.1016/j.trc.2013.02.008
- Johnson, M. J., Chahal, T., Stinchcombe, A., Mullen, N., Weaver, B., and Bédard, M. (2011). Physiological responses to simulated and on-road driving. *Int. J. Psychophysiol.* 81, 203–208. doi: 10.1016/j.ijpsycho.2011.06.012
- Keith, K., Trentacoste, M., Depue, L., Granda, T., Huckaby, E., Ibarguen, B., et al. (2005). *Roadway Human Factors and Behavioral Safety in Europe*. Washington, DC: The National Academies of Sciences, Engineering, and Medicine.
- Kim, I. H., Kim, J. W., Haufe, S., and Lee, S. W. (2015). Detection of braking intention in diverse situations during simulated driving based on EEG feature combination. *J. Neural Eng.* 12:016001. doi: 10.1088/1741-2560/12/1/016001
- Lanata, A., Valenza, G., Greco, A., Gentili, C., Bartolozzi, R., Bucchi, F., et al. (2015). How the autonomic nervous system and driving style change with incremental stressing conditions during simulated driving. *IEEE Trans. Intell. Transp. Syst.* 16, 1505–1517. doi: 10.1109/TITS.2014.2365681
- Lee, H. B., Choi, J. M., Kim, J. S., Kim, Y. S., Baek, H. J., Ryu, M. S., et al. (2007). Noninvasive biosignal measurement system in a vehicle. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2007, 2303–2306. doi: 10.1109/IEMBS.2007.4352786
- Lewis, B. A., Eisert, J. L., and Baldwin, C. L. (2018). Validation of essential acoustic parameters for highly urgent in-vehicle collision warnings. *Hum. Fact.* 60, 248–261. doi: 10.1177/0018720817742114
- Li, G., and Chung, W. Y. (2013). Detection of driver drowsiness using wavelet analysis of heart rate variability and a support vector machine classifier. *Sensors (Basel)* 13, 16494–16511. doi: 10.3390/s131216494
- Lin, C. T., Chang, C. J., Lin, B. S., Hung, S. H., Chao, C. F., and Wang, I. J. (2010). A real-time wireless brain-computer interface system for drowsiness detection. *IEEE Trans. Intell. Transp. Syst.* 4, 214–222. doi: 10.1109/TBCAS.2010.2046415
- Lin, C. T., Chuang, C. H., Huang, C. S., Tsai, S. F., Lu, S. W., Chen, Y. H., et al. (2014). Wireless and wearable EEG system for evaluating driver vigilance. *IEEE Trans. Biomed. Circ. Syst.* 8, 165–176. doi: 10.1109/TBCAS.2014.2316224
- Liu, T., Pelowski, M., Pang, C., Zhou, Y., and Cai, J. (2016). Near-infrared spectroscopy as a tool for driving research. *Ergonomics* 59, 368–379. doi: 10.1080/00140139.2015.1076057
- Llaneras, R. E., Salinger, J., and Green, C. A. (2013). “Human factors issues associated with limited ability autonomous driving systems: drivers' allocation of visual attention to the forward roadway,” in *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Bolton Landing, New York (Iowa City, IA: Public Policy Center, University of Iowa), 92–98. doi: 10.17077/drivingassessment.1472
- Maglione, A., Borghini, G., Aricò, P., Borgia, F., Graziani, I., Colosimo, A., et al. (2014). Evaluation of the workload and drowsiness during car driving by using high resolution EEG activity and neurophysiologic indices. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2014, 6238–6241. doi: 10.1109/EMBC.2014.6945054
- 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. *Hum. Fact.* 54, 396–412. doi: 10.1177/0018720812442086
- Mehta, R. K., and Parasuraman, R. (2013). Neuroergonomics: a review of applications to physical and cognitive work. *Front. Hum. Neurosci.* 7:889. doi: 10.3389/fnhum.2013.00889
- Merat, N., and Lee, J. D. (2012). Preface to the special section on human factors and automation in vehicles: designing highly automated vehicles with the driver in mind. *Hum. Fact.* 54, 681–686. doi: 10.1177/0018720812461374
- Metzger, U., and Parasuraman, R. (2001). The role of the air traffic controller in future air traffic management: an empirical study of active control versus passive monitoring. *Hum. Fact.* 43, 519–528. doi: 10.1518/001872001775870421
- Michon, J. A. (1985). “A critical view of driver behavior models: what do we know, what should we do?” in *Human Behavior and Traffic Safety*, eds L. Evans and R. C. Schwing (Boston, MA: Springer), 485–524. doi: 10.1007/978-1-4613-2173-6\_19
- Molnar, L. J., and Eby, D. W. (2017). Implications of advanced vehicle technologies for older drivers. *Accid. Anal. Prev.* 106, 457–459. doi: 10.1016/j.aap.2017.06.017

- Näätänen, R. (1973). "The inverted-U relationship between activation and performance: a critical review," in *Attention and Performance IV*, ed. S. Kornblum (Cambridge, MA: Academic Press), 4–155.
- Neubauer, C., Matthews, G., Langheim, L., and Saxby, D. (2012). Fatigue and voluntary utilization of automation in simulated driving. *Hum. Fact.* 54, 734–746. doi: 10.1177/0018720811423261
- Noori, S. M., and Mikaeili, M. (2016). Driving drowsiness detection using fusion of electroencephalography, electrooculography, and driving quality signals. *J. Med. Signals Sens.* 6, 39–46.
- Noy, I. Y., Shinar, D., and Horrey, W. J. (2018). Automated driving: safety blind spots. *Saf. Sci.* 102, 68–78. doi: 10.1016/j.ssci.2017.07.018
- Otte, D., Pund, B., and Jänsch, M. (2009). "A new approach of accident causation analysis by seven steps ACASS," in *Proceedings of the International Technical Conference on the Enhanced Safety of Vehicles*, Vol. 2009 (Washington, DC: National Highway Traffic Safety Administration).
- Papadelis, C., Chen, Z., Kourtidou-Papadelis, C., Bamidis, P. D., Chouvarda, I., Bekiaris, E., et al. (2007). Monitoring sleepiness with on-board electrophysiological recordings for preventing sleep-deprived traffic accidents. *Clin. Neurophysiol.* 118, 1906–1922. doi: 10.1016/j.clinph.2007.04.031
- Payre, W., Cestac, J., and Delhomme, P. (2016). Fully automated driving: impact of trust and practice on manual control recovery. *Hum. Fact.* 58, 229–241. doi: 10.1177/0018720815612319
- Porges, S. W. (1995). Orienting in a defensive world: mammalian modifications of our evolutionary heritage, a polyvagal theory. *Psychophysiology* 32, 301–318. doi: 10.1111/j.1469-8986.1995.tb01213.x
- Rada, H., Dittmar, A., Delhomme, G., Collet, C., Roure, R., Vernet-Maury, E., et al. (1995). Bioelectric and microcirculation cutaneous sensors for the study of vigilance and emotional response during tasks and tests. *Biosens. Bioelectron.* 10, 7–15. doi: 10.1016/0956-5663(95)96790-6
- Ramon, C., Clarion, A., Gehin, C., Petit, C., Collet, C., and Dittmar, A. (2008). An integrated platform to assess driver's physiological and functional states. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2008, 506–509. doi: 10.1109/IEMBS.2008.4649201
- Ranney, T. A. (1994). Models of driving behavior: a review of their evolution. *Accid. Anal. Prev.* 1994, 733–750. doi: 10.1016/0001-4575(94)90051-5
- Reyes-Muñoz, A., Domingo, M. C., López-Trinidad, M. A., and Delgado, J. L. (2016). Integration of body sensor networks and vehicular ad-hoc networks for traffic safety. *Sensors* 16:E107. doi: 10.3390/s16010107
- Rigas, G., Goletis, Y., Bougia, P., and Fotiadis, D. (2011). Towards driver's state recognition on real driving conditions. *Int. J. Veh. Technol.* 2011, 1–14. doi: 10.1155/2011/617210
- Rigas, G., Goletis, Y., and Fotiadis, D. I. (2012). Real-time driver's stress event detection. *IEEE Trans. Intell. Transp. Syst.* 13, 221–234. doi: 10.1109/TITS.2011.2168215
- Sabey, B. E., and Staughton, G. C. (1975). "Interacting roles of road environment, vehicle and road user in accidents," in *Proceedings of the 5th International Conference of the International Association for Accident and Traffic Medicine*, London.
- Sahayadhas, A., Sundaraj, K., and Murugappan, M. (2012). Detecting driver drowsiness based on sensors: a review. *Sensors* 12, 16937–16953. doi: 10.3390/s121216937
- Sanders, A. F. (1990). Issues and trends in the debate on discrete vs. continuous processing of information. *Acta Psychol.* 74, 123–167. doi: 10.1016/0001-6918(90)90004-Y
- Sawyer, B., Teo, G., and Mouloua, M. (2012). DriveID: safety innovation through individuation. *Work* 41(Suppl. 1), 4273–4278.
- Saxby, D. J., Matthews, G., Warm, J. S., Hitchcock, E. M., and Neubauer, C. (2013). Active and passive fatigue in simulated driving: discriminating styles of workload regulation and their safety impacts. *J. Exp. Psychol. Appl.* 19, 287–300. doi: 10.1037/a0034386
- Shen, S., and Neyens, D. M. (2017). Assessing drivers' response during automated driver support system failures with non-driving tasks. *J. Saf. Res.* 61, 149–155. doi: 10.1016/j.jsr.2017.02.009
- Shinar, D. (1978). *Psychology on the Road: The Human Factor in Traffic Safety*. Hoboken, NJ: John Wiley & Sons Inc.
- Sigari, M. H., Pourshahabi, M. R., Soryani, M., and Fathy, M. A. (2014). Review on driver face monitoring systems for fatigue and distraction detection. *Int. J. Adv. Sci. Technol.* 64, 73–100. doi: 10.14257/ijast.2014.64.07
- Singh, R. R., Conjeti, S., and Banerjee, R. A. (2013). Comparative evaluation of neural network classifiers for stress level analysis of automotive drivers using physiological signals. *Biomed. Signal Process. Control* 8, 740–754. doi: 10.1016/j.bspc.2013.06.014
- Singh, S. (2015). *Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey*. Washington, DC: National Highway Traffic Safety Administration, 1–2. doi: 10.1016/j.trf.2015.04.014
- Song, F., Tan, X., Chen, S., and Zhou, Z. H. (2013). A literature survey on robust and efficient eye localization in real-life scenarios. *Pattern Recognit.* 46, 3157–3173. doi: 10.1016/j.patcog.2013.05.009
- Strauch, B. (2017). The automation-by-expertise-by-training interaction. *Hum. Fact.* 59, 204–228. doi: 10.1177/0018720816665459
- Sun, Y., and Yu, X. B. (2014). An innovative nonintrusive driver assistance system for vital signal monitoring. *IEEE J. Biomed. Health Inform.* 6, 1932–1939. doi: 10.1109/JBHI.2014.2305403
- Taamneh, S., Tsiamyrtzis, P., Dcosta, M., Buddharaju, P., Khatr, A., Manser, M., et al. (2017). A multimodal dataset for various forms of distracted driving. *Sci. Data* 4:170110. doi: 10.1038/sdata.2017.110
- Treat, J. R. (1977). "Tri-level study of the causes of traffic accidents: an overview of final results," in *Proceedings of the American Association for Automotive Medicine Annual Conference*, vol. 21 (Chicago, IL: Association for the Advancement of Automotive Medicine), 391–403.
- Vicente, J., Laguna, P., Bartra, A., and Bailón, R. (2016). Drowsiness detection using heart rate variability. *Med. Biol. Eng. Comput.* 54, 927–937. doi: 10.1007/s11517-015-1448-7
- Waldrop, M. M. (2015). Autonomous vehicles: no drivers required. *Nature* 518, 20–23. doi: 10.1038/518020a
- Wang, W., Wang, B., Bu, L., Xu, L., Li, Z., and Fan, Y. (2016). Vigilance task-related change in brain functional connectivity as revealed by wavelet phase coherence analysis of near-infrared spectroscopy signals. *Front. Hum. Neurosci.* 10:100. doi: 10.3389/fnhum.2016.00400
- Wickens, C. D., Hollands, J. G., Banbury, S., and Parasuraman, R. (2015). *Engineering Psychology & Human Performance*. London: Psychology Press.
- World Health Organization, Global Status Report on Road Safety – Summary (2018). Available at: [https://www.who.int/violence\\_injury\\_prevention/road\\_safety\\_status/2018/GSRRS2018\\_Summary\\_EN.pdf](https://www.who.int/violence_injury_prevention/road_safety_status/2018/GSRRS2018_Summary_EN.pdf) (accessed February 26, 2019).
- Young, M. S., and Stanton, N. A. (2002). Malleable attentional resources theory: a new explanation for the effects of mental underload on performance. *Hum. Fact.* 44, 365–375. doi: 10.1518/0018720024497709
- Young, M. S., and Stanton, N. A. (2007). What's skill got to do with it? Vehicle automation and driver mental workload. *Ergonomics* 50, 1324–1339. doi: 10.1080/00140130701318855
- Yu-Lung, C., Yen-Cheng, F., and Chen, O. T. (2016). Real-time physiological and facial monitoring for safe driving. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2016, 4849–4852. doi: 10.1109/EMBC.2016.7591813
- 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. *Accid. Anal. Prev.* 92, 230–239. doi: 10.1016/j.aap.2016.04.002

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer JV and handling Editor declared their shared affiliation.

Copyright © 2019 Collet and Musicant. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Advantages of publishing in Frontiers



## OPEN ACCESS

Articles are free to read  
for greatest visibility  
and readership



## FAST PUBLICATION

Around 90 days  
from submission  
to decision



## HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,  
and constructive  
peer-review



## TRANSPARENT PEER-REVIEW

Editors and reviewers  
acknowledged by name  
on published articles

## Frontiers

Avenue du Tribunal-Fédéral 34  
1005 Lausanne | Switzerland

**Visit us:** [www.frontiersin.org](http://www.frontiersin.org)

**Contact us:** [info@frontiersin.org](mailto:info@frontiersin.org) | +41 21 510 17 00



## REPRODUCIBILITY OF RESEARCH

Support open data  
and methods to enhance  
research reproducibility



## DIGITAL PUBLISHING

Articles designed  
for optimal readership  
across devices



## FOLLOW US

[@frontiersin](https://twitter.com/frontiersin)



## IMPACT METRICS

Advanced article metrics  
track visibility across  
digital media



## EXTENSIVE PROMOTION

Marketing  
and promotion  
of impactful research



## LOOP RESEARCH NETWORK

Our network  
increases your  
article's readership