Fatigue and mental underload further pronounced in L3 conditionally automated driving: Results from an EEG experiment on a test track

Nikol Figalová
nikol.figalova@uni-ulm.de
Deptartment of Clinical and Health
Psychology, Ulm University
Ulm, Germany

Hans-Joachim Bieg
Robert Bosch GmbH
Renningen, Germany

Michael Schulz
Robert Bosch GmbH
Renningen, Germany

Jürgen Pichen
Department of Human Factors, Ulm
University
Ulm, Germany

Martin Baumann
Department of Human Factors, Ulm
University
Ulm, Germany

Lewis Chuang
Department of Human and
Technology, Institute for Media
Research, Faculty of Humanities,
Chemnitz University of Technology
Ulm, Germany

Olga Pollatos
Deptartment of Clinical and Health
Psychology, Ulm University
Ulm, Germany

Figure 1: The test track which was used in the experiment.

ABSTRACT

Drivers' role changes with increasing automation from the primary driver to a system supervisor. This study investigates how supervising an SAE L2 and L3 automated vehicle (AV) affects drivers' mental workload and sleepiness compared to manual driving. Using an AV prototype on a test track, the oscillatory brain activity of 23 adult participants was recorded during L2, L3, and manual driving. Results showed decreased mental workload and increased sleepiness in L3 drives compared to L2 and manual drives, indicated by self-report scales and changes in the frontal alpha and theta power spectral density. These findings suggest that fatigue and mental underload are significant issues in L3 driving and should be considered when designing future AV interfaces.

CCS CONCEPTS

• **Human-centered computing** → **Field studies**; **Empirical studies in HCI**; • **Social and professional topics** → **User characteristics**.

KEYWORDS

automated driving, L3 driving, conditional automation, fatigue, drowsiness, mental underload, EEG

1 INTRODUCTION

With increasing vehicle automation, drivers' role changes. While L0 SAE [32] vehicles are operated solely by the driver, SAE L2 vehicles

can take over the driving task when certain conditions are met. Nevertheless, L2 vehicles must be constantly monitored. This is no longer necessary on SAE L3 - drivers may engage in non-driving-related activities while the vehicle operates in an automated mode. This will likely result in a significant shift in the skills and abilities drivers need to possess.

The success of the transition to automated vehicles (AVs) depends on drivers' ability to adapt to their new role as monitors rather than primary operators. However, drivers may not be well prepared for this change [2, 6, 27]. Automated driving systems can perform the active tasks of steering and navigating complex traffic situations, while drivers are expected to passively supervise the system. This can lead to mental underload [26, 35]. The lack of stimulation makes highly automated driving difficult to satisfy psychological fulfilment and user experience [11], resulting in drivers being unable to stay alert during extended periods of automated driving [3, 37]. Therefore, AV drivers experience more fatigue compared to those driving manually [22, 28].

Fatigue significantly impairs the driver's ability to operate a vehicle safely [13, 21, 24, 25], hinders their attention [4], and situation awareness [5]. Moreover, drivers tend to misuse or abuse the AV system, either intentionally or due to a lack of understanding of their tasks as AV operators [30, 31]. These issues should be addressed when designing an intelligent user interface (UI) that monitors and supports drivers' performance [10, 20, 36, 40] and manages their attention [1, 41, 42].

Drivers fatigue can be detected using computer vision [9, 39, 43] and artificial intelligence [17, 38] using measures such as eyelid closure [7, 44], electrocardiogram [16], or electroencephalography (EEG) [15, 29, 33]. Most EEG studies report an increase in frontal alpha activity as an indicator of fatigue, while an increase in frontal theta power indicates a higher mental workload (for an overview, see Lohani et al. [23]). Detecting the driver's state is crucial for adaptive UIs. However, we should use a prophylactic approach to design safe, desirable, and acceptable intelligent UIs. Therefore, we must understand what drivers experience while monitoring the AV to design UIs that meet their needs.

Most research on driver fatigue and mental underload has been conducted using driving simulators [8], which have limitations that may affect the relevance of findings in the real world [14]. Moreover, most studies used scenarios that differ from what drivers would encounter in actual AVs (e.g., critical situations are rather unusual in natural traffic conditions [18], although they occur relatively frequently in simulator experiments). We propose an experiment using a test track and an AV prototype to address these issues. Participants will experience a monotonous, non-stimulating task similar to operating real L2 and L3 AV. We will measure sleepiness and mental workload using self-report scales and EEG. The goal is to compare drivers' cognitive states when interacting with L2 and L3 automation in a realistic setting. Our results will serve as a knowledgebase for the designers of future intelligent UIs for highly automated vehicles.

2 METHOD

2.1 Apparatus and experimental conditions

We used an AV prototype (VW Golf 7) at a test track in Stuttgart, Germany (Figure 1). The AV travelled in loops of 2000 m at a speed of 50 km/h on straight sections and 20 km/h in curves. The EEG was recorded using 32 channels placed according to the international 10-20 system. We used active electrodes and kept the impedance below $25k\Omega$. The data were recorded with a 1000 Hz sampling rate. The perceived mental workload was assessed using the NASA-TLX [12]. Sleepiness was measured using the Karolinska sleepiness scale [34].

Participants experienced three counterbalanced conditions (17 minutes each): monitoring an SAE L3 vehicle; driving an SAE L2 vehicle; manual driving. Participants were primed about the differences in their tasks in each condition. In the **L3 ride**, participants were told that the vehicle would request them to take control in advance if necessary. In the **L2 ride**, participants were told to supervise the automated system constantly and intervene in case of failure. In the **manual ride**, participants were asked to drive at 50 km/h on straight portions and slow to 20 km/h in curves.

2.2 Participants

We recruited 23 participants (14 females; M = 41.24 years; SD = 14.71) with no prior AV experience. All participants had normal or corrected-to-normal vision, no known neurological or psychiatric disease, and a valid German driving license (on average for M = 18.12 years, SD = 12.63). All participants provided informed consent and received 90 Euros.

2.3 EEG signal processing and analysis

The EEG data were preprocessed in Matlab version R2022a according to the BeMoBil pipeline [19]. The cleaned data were bandpass-filtered between 0.5 and 30 Hz. The relative spectral power density was analyzed in the theta band (4-8 Hz) and the alpha band (8 to 12 Hz) on the Fz electrode.

3 RESULTS

3.1 Self-report mental workload and sleepiness

A repeated measures ANOVA with a Greenhouse-Geisser correction determined that the mean NASA-TLX score differed between the rides ($F(1.569, 34.514)$ = 7.313, p=.004, ω^2 = 0.070). Post-hoc testing using the Bonferroni correction revealed that the mental workload was lower in the L3 rides compared to the L2 rides (mean difference=10.957, p=.020, d=.537), and in the L3 rides compared to the manual rides (mean difference=14.043, p=.002, d=.688).

Figure 2a presents the mean score of the six dimensions of NASA-TLX measured after the L2, L3, and manual rides. The scores suggest higher mental, physical, and effort demands in the manual rides compared to the L3 rides. Furthermore, we found a higher performance demand for the L2 rides compared to the L3 rides. Moreover, we compared the self-report sleepiness. A repeated measures ANOVA with a Greenhouse-Geisser correction determined that the mean KSS score differed between the rides ($F(1.961, 43.147)$ = 3.386, p=.044, ω^2 = 0.019). Post-hoc testing using the Bonferroni correction revealed that drivers experienced more sleepiness in
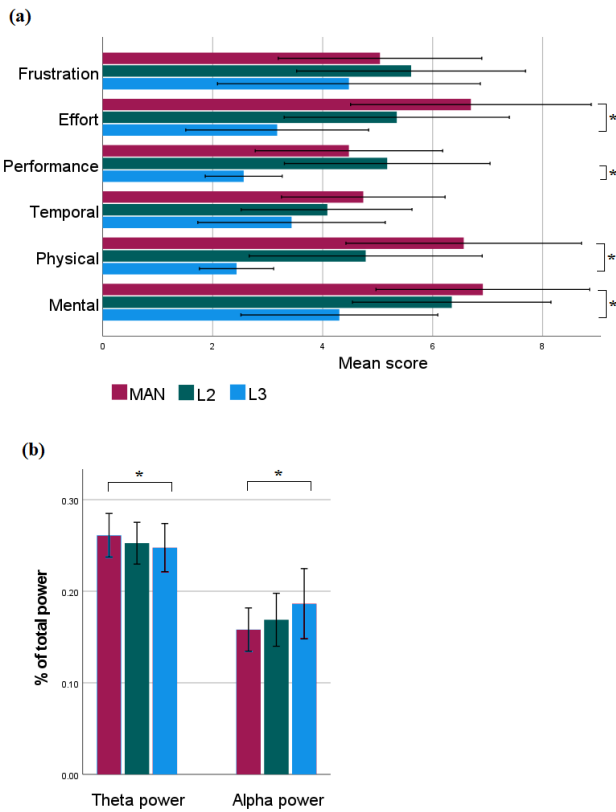
(a)

(b)

Figure 2: (a) The mean scores of the six dimensions of mental workload measured by NASA-TLX after each ride. (b) The relative power spectral density measured on the Fz electrode in the theta (4-8 Hz) and alpha (8-12 Hz) band. Significant differences in the mean scores are marked with an asterisk. The error bars represent the 95% confidence interval.

the L3 rides compared to the manual rides (mean difference=0.826, p=.039, d=.407).

3.2 EEG measures

We assessed the power spectral density in the theta (4-8 Hz) and alpha (8-12 Hz) bands. Results are visualised in 2b. A repeated measures ANOVA determined that the theta power differed between the rides (F(2.000, 42.000) = 4.937, p=.012, ω^2 = 0.008). Post-hoc testing using the Bonferroni correction revealed that the theta power was higher in the manual rides compared to the L3 rides (mean difference=0.013, p=.010, d=.246), which suggests that drivers experienced higher mental workload when driving manually.

Moreover, a repeated measures ANOVA determined that the alpha power differed between the rides (F(2.000, 42.000) = 4.417, p=.018, ω^2 = 0.021). Post-hoc testing using the Bonferroni correction revealed that the alpha power was higher in the L3 rides compared to the manual rides (mean difference=0.028, p=.016, d=.405), which suggests that drivers experienced more fatigue and drowsiness during the L3 rides.

4 DISCUSSION

The present study reports preliminary findings on the perception of L3 driving among drivers in a real AV. Our results suggest that self-reported mental workload was lowest during L3 rides, with no difference in mental workload between L2 and manual rides. The decreased frontal theta brain activity during L3 rides supports the idea that drivers experience a low mental workload while passively monitoring L3 AVs. Moreover, we found no difference between manual and L2 rides but a significant difference between manual and L3 rides. The increased alpha activity during L3 rides supports the argument that drivers become more fatigued and drowsy while passively monitoring L3 AVs.

Our findings are consistent with previous research indicating that vehicle automation reduces mental workload [3, 8, 22, 26, 35] and increases drowsiness [22, 28, 37]. However, previous studies primarily focused on L2 driving, and only scarce evidence has been published about L3 AVs outside the driving simulator. Our results suggest that the underload effect on mental workload may be further pronounced in L3 driving, which can be potentially dangerous given that L3 drivers are still partially responsible for the driving task [32]. This study was conducted on a test track with automation inexperienced drivers. Real traffic environment, as well as the effect of long-term experience with automation, should be addressed in future studies.

In conclusion, our experiment highlights the importance of designing UIs for L3 vehicles with consideration for drowsy and inattentive drivers. Dynamic UIs that manage drivers' attention and adapt to the current driving context are recommended. Issues such as attention shifting and cue saliency must be addressed. A driving monitoring system that detects inattention and engages with the driver in a timely manner ahead of a control transition should be included in the design of safe and desirable UIs for L3 vehicles. Further research is necessary to provide more specific recommendations addressing sleepiness and mental underload in the design of intelligent UIs for L3 AVs.

ACKNOWLEDGMENTS

REFERENCES

[1] Brian P Bailey and Joseph A Konstan. 2006. On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in human behavior* 22, 4 (2006), 685–708.

[2] Klaus Bengler, Klaus Dietmayer, Berthold Farber, Markus Maurer, Christoph Stiller, and Hermann Winner. 2014. Three decades of driver assistance systems: Review and future perspectives. *IEEE Intelligent transportation systems magazine* 6, 4 (2014), 6–22.

[3] Hans-Joachim Bieg, Constantina Daniilidou, Britta Michel, and Anna Sprung. 2020. Task load of professional drivers during level 2 and 3 automated driving. *Proceedings of the Human Factors and Ergonomics Society Europe* (2020), 41–52.

[4] Maarten AS Boksem, Theo F Meijman, and Monicque M Lorist. 2005. Effects of mental fatigue on attention: an ERP study. *Cognitive brain research* 25, 1 (2005),

107–116.

[5] Miriam Bongo and Rosemary Seva. 2022. Effect of Fatigue in Air Traffic Controllers' Workload, Situation Awareness, and Control Strategy. *The International Journal of Aerospace Psychology* 32, 1 (2022), 1–23.

[6] Stefan Brandenburg and E-M Skottke. 2014. Switching from manual to automated driving and reverse: Are drivers behaving more risky after highly automated driving?. In *17th international IEEE conference on intelligent transportation systems (ITSC)*. IEEE, Institute of Electrical and Electronics Engineers (IEEE), 2978–2983.

[7] Tang-Hsien Chang and Yi-Ru Chen. 2014. Driver fatigue surveillance via eye detection. In *17th international ieee conference on intelligent transportation systems (itsc)*. IEEE, 366–371.

[8] Joost CF De Winter, Riender Happee, Marieke H Martens, and Neville A Stanton. 2014. Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation research part F: traffic psychology and behaviour* 27 (2014), 196–217.

[9] Kartik Dwivedi, Kumar Biswaranjan, and Amit Sethi. 2014. Drowsy driver detection using representation learning. In *2014 IEEE international advance computing conference (IACC)*. IEEE, 995–999.

[10] Nikol Figalová, Lewis L Chuang, Jürgen Pichen, Martin Baumann, and Olga Pollatos. 2022. Ambient light conveying reliability improves drivers' takeover performance without increasing mental workload. *Multimodal Technologies and Interaction* 6, 9 (2022), 73.

[11] Anna-Katharina Frison, Philipp Wintersberger, Tianjia Liu, and Andreas Riener. 2019. Why do you like to drive automated? a context-dependent analysis of highly automated driving to elaborate requirements for intelligent user interfaces. In *Proceedings of the 24th international conference on intelligent user interfaces*. 528–537.

[12] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

[13] J Stephen Higgins, Jeff Michael, Rory Austin, Torbjörn Åkerstedt, Hans Van Dongen, Nathaniel Watson, Charles Czeisler, Allan I Pack, and Mark R Rosekind. 2017. Asleep at the wheel—the road to addressing drowsy driving. *Sleep* 40, 2 (2017).

[14] Philipp Hock, Johannes Kraus, Franziska Babel, Marcel Walch, Enrico Rukzio, and Martin Baumann. 2018. How to design valid simulator studies for investigating user experience in automated driving: review and hands-on considerations. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 105–117.

[15] Jianfeng Hu and Jianliang Min. 2018. Automated detection of driver fatigue based on EEG signals using gradient boosting decision tree model. *Cognitive neurodynamics* 12, 4 (2018), 431–440.

[16] Sang-Joong Jung, Heung-Sub Shin, and Wan-Young Chung. 2014. Driver fatigue and drowsiness monitoring system with embedded electrocardiogram sensor on steering wheel. *IET Intelligent Transport Systems* 8, 1 (2014), 43–50.

[17] Waldemar Karwowski, Przemysław Reszke, and Marian Rusek. 2020. Artificial Intelligence System for Drivers Fatigue Detection. In *International Conference on Computer Information Systems and Industrial Management*. Springer, 39–50.

[18] Moritz Klischat and Matthias Althoff. 2019. Generating critical test scenarios for automated vehicles with evolutionary algorithms. In *2019 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2352–2358.

[19] Marius Klug, Sein Jeung, Anna Wunderlich, Lukas Gehrke, Janna Protzak, Zakaria Djebbara, Andreas Argubi-Wollesen, Bettina Wollesen, and Klaus Gramann. 2022. The BeMoBIL Pipeline for automated analyses of multimodal mobile brain and body imaging data. *bioRxiv* (2022).

[20] Thomas Köhn, Matthias Gottlieb, Michael Schermann, and Helmut Krcmar. 2019. Improving take-over quality in automated driving by interrupting non-driving tasks. In *Proceedings of the 24th international conference on intelligent user interfaces*. 510–517.

[21] BM Kusuma Kumari and P Ramakanth Kumar. 2017. A survey on drowsy driver detection system. In *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*. IEEE, 272–279.

[22] Thomas Kundinger, Andreas Riener, Nikoletta Sofra, and Klemens Weigl. 2020. Driver drowsiness in automated and manual driving: insights from a test track study. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*. 369–379.

[23] Monika Lohani, Brennan R Payne, and David L Strayer. 2019. A review of psychophysiological measures to assess cognitive states in real-world driving. *Frontiers in human neuroscience* 13 (2019), 57.

[24] Alistair W MacLean, David RT Davies, and Kris Thiele. 2003. The hazards and prevention of driving while sleepy. *Sleep medicine reviews* 7, 6 (2003), 507–521.

[25] Gustav Markkula and J Engström. 2017. Simulating effects of arousal on lane keeping: Are drowsiness and cognitive load opposite ends of a single spectrum?. In *Tenth International Conference on Managing Fatigue, San Diego, CA*.

[26] Thomas McWilliams and Nathan Ward. 2021. Underload on the road: measuring vigilance decrements during partially automated driving. *Frontiers in psychology* 12 (2021), 631364.

[27] Natasha Merat and A Hamish Jamson. 2009. How do drivers behave in a highly automated car?. In *Driving Assesment Conference*, Vol. 5. University of Iowa.

[28] Rebecca Michael and Renata Meuter. 2006. Sustained attention and hypovigilance: The effect of environmental monotony on continuous task performance and implications for road safety. In *Proceedings of the Australasian road safety research, policing and education conference*, Vol. 10. Monash University.

[29] José M Morales, Carolina Díaz-Piedra, Héctor Rieiro, Joaquín Roca-González, Samuel Romero, Andrés Catena, Luis J Fuentes, and Leandro L Di Stasi. 2017. Monitoring driver fatigue using a single-channel electroencephalographic device: A validation study by gaze-based, driving performance, and subjective data. *Accident Analysis & Prevention* 109 (2017), 62–69.

[30] Raja Parasuraman and Victor Riley. 1997. Humans and automation: Use, misuse, disuse, abuse. *Human factors* 39, 2 (1997), 230–253.

[31] James Robinson. 2023. Police in Germany chase tesla for 15 minutes after driver turns on autopilot and 'goes to sleep'. https://news.sky.com/story/police-in-germany-chase-tesla-for-15-minutes-after-driver-turns-on-autopilot-and-goes-to-sleep-12778306

[32] SAE International. 2021. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles(SAE Standard J3016, Report No. J3016-202104)*. Technical Report. SAE International, Warrendale, PA.

[33] Misbah Kazi Salimuddin, Shraddha Panbude, et al. 2018. Driver drowsiness monitoring system using fusion of facial features & EEG. In *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*. IEEE, 1506–1510.

[34] Azmeh Shahid, Kate Wilkinson, Shai Marcu, and Colin M Shapiro. 2011. Karolinska sleepiness scale (KSS). In *STOP, THAT and one hundred other sleep scales*. Springer, 209–210.

[35] Jork Stapel, Freddy Antony Mullakkal-Babu, and Riender Happee. 2019. Automated driving reduces perceived workload, but monitoring causes higher cognitive load than manual driving. *Transportation research part F: traffic psychology and behaviour* 60 (2019), 590–605.

[36] Burkay Sucu and Eelke Folmer. 2013. Haptic interface for non-visual steering. In *Proceedings of the 2013 international conference on Intelligent user interfaces*. 427–434.

[37] Tobias Vogelpohl, Matthias Kühn, Thomas Hummel, and Mark Vollrath. 2019. Asleep at the automated wheel—Sleepiness and fatigue during highly automated driving. *Accident Analysis & Prevention* 126 (2019), 70–84.

[38] Esra Vural, Mujdat Cetin, Aytul Ercil, Gwen Littlewort, Marian Bartlett, and Javier Movellan. 2007. Drowsy driver detection through facial movement analysis. In *international workshop on human-computer interaction*. Springer, 6–18.

[39] Esra Vural, Müjdat Çetin, Aytül Erçil, Gwen Littlewort, Marian Bartlett, and Javier Movellan. 2008. Automated drowsiness detection for improved driving safety. (2008).

[40] Marc Wilbrink, Anna Schieben, and Michael Oehl. 2020. Reflecting the automated vehicle's perception and intention: Light-based interaction approaches for on-board HMI in highly automated vehicles. In *Proceedings of the 25th International Conference on Intelligent User Interfaces Companion*. 105–107.

[41] Philipp Wintersberger, Andreas Riener, Clemens Schartmüller, Anna-Katharina Frison, and Klemens Weigl. 2018. Let me finish before I take over: Towards attention aware device integration in highly automated vehicles. In *Proceedings of the 10th international conference on automotive user interfaces and interactive vehicular applications*. 53–65.

[42] Philipp Wintersberger, Clemens Schartmüller, and Andreas Riener. 2019. Attentive user interfaces to improve multitasking and take-over performance in automated driving: the auto-net of things. *International Journal of Mobile Human Computer Interaction (IJMHCI)* 11, 3 (2019), 40–58.

[43] Zhitao Xiao, Zhiqiang Hu, Lei Geng, Fang Zhang, Jun Wu, and Yuelong Li. 2019. Fatigue driving recognition network: fatigue driving recognition via convolutional neural network and long short-term memory units. *IET Intelligent Transport Systems* 13, 9 (2019), 1410–1416.

[44] Feng Zhou, Areen Alsaid, Mike Blommer, Reates Curry, Radhakrishnan Swaminathan, Dev Kochhar, Walter Talamonti, Louis Tijerina, and Baiying Lei. 2020. Driver fatigue transition prediction in highly automated driving using physiological features. *Expert Systems with Applications* 147 (2020), 113204.