

Article

Ambient Light Conveying Reliability Improves Drivers' Takeover Performance without Increasing Mental Workload

Nikol Figalová ^{1*}, Lewis L. Chuang ², Jürgen Pichen ³, Martin Baumann ³, and Olga Pollatos ¹

¹ Department of Clinical and Health Psychology, Institute of Psychology and Education, Ulm University, Germany

² Department of Human and Technology, Institute for Media Research, Faculty of Humanities, Chemnitz University of Technology, Germany

³ Department of Human Factors, Institute of Psychology and Education, Ulm University, Germany

* Correspondence: nikol.figalova@uni-ulm.de

Abstract: Drivers of L3 automated vehicles (AVs) are not required to continuously monitor the AV system. However, they must be prepared to take over when requested. Therefore, it is necessary to design an in-vehicle environment that allows drivers to adapt their levels of preparedness to the likelihood of control transition. This study evaluates ambient in-vehicle lighting that continuously communicates the current level of AV reliability, specifically on how it could influence drivers' take-over performance and mental workload (MW). We conducted an experiment in a driving simulator with 42 participants who experienced 10 take-over requests (TORs). The experimental group experienced a four-stage ambient light display that communicated the current level of AV reliability, which was not provided to the control group. The experimental group demonstrated better take-over performance, based on lower vehicle jerks. Notably, perceived MW did not differ between the groups, and the EEG indices of MW (frontal theta power, parietal alpha power, Task-Load Index) did not differ between the groups. These findings suggest that communicating the current level of reliability using ambient light might help drivers be better prepared for TORs and perform better without increasing their MW.

Keywords: ambient light; reliability; take-over request; mental workload; electroencephalography (EEG); transition of control

1. Introduction

Vehicle automation promises to enhance driver's comfort, traffic safety, and efficiency. However, to make automated vehicles (AVs) widely accepted and desirable, we must design in-vehicle systems that accommodate human capabilities. What are the driver's needs and preferences? What are the cognitive processes that support driver-vehicle communication and interaction? This study presents the development and empirical evaluation of a display that continuously informs drivers about the current reliability level of the AV system. Such a system could render AV behaviour more predictable for the driver and, hence, improve the joint driver-vehicle performance.

The Society of Automotive Engineers (SAE) defined six levels of automation [1]. Level 0 refers to no automation, while Level 5 refers to full automation. Increasing steps denote higher levels of vehicle autonomy and automation. The crucial distinction lies between partial automation (Level 2; L2) and conditional automation (Level 3; L3). For instance, an L2 vehicle can perform lateral and longitudinal control but requires the driver to supervise the system at all times. An L3 vehicle can, under certain conditions, operate autonomously and the driver is allowed to engage in non-driving tasks. However, the driver must reliably take over control when requested by the AV. Nevertheless, how can drivers be expected to resume stable vehicle control within a reasonable time, if they are engaged in non-driving tasks?

A driver remains calibrated to the vehicle dynamics, and environmental conditions during manual driving [2]. However, drivers' visual attention might be focused elsewhere when an L3 vehicle drives in an automated mode, resulting in impaired situation awareness [3,4] that is informed by peripheral ambient cues, e.g., vestibular information [5]. The lack of calibration of the vehicle and low situation awareness is not a problem as long as the vehicle operates in the automated mode. However, when an L3 vehicle reaches its operational boundary, a take-over request (TOR) is issued. Because the vehicle cannot operate safely by itself anymore, the driver is required to regain control. However, successful steering requires a frequently updated perceptual-motor loop, attuning gaze behaviour and steering to the current vehicle and environmental conditions [2]. The perceptual-motor control loop is disrupted during automated driving, which closely relates to the *out-of-the-loop* phenomena when the driver is not in physical control of the vehicle, and not monitoring the driving situation [6]. The lack of perceptual-motor calibration results in lower vehicle control and poorer driving performance, which can be represented by increased vehicle jerk [7–11]

To successfully implement L3 vehicles, we must ensure that drivers are ready to take over when requested. In addition, an appropriate user interface is necessary to ensure a smooth control transition. In the following sections, we will focus on concepts that lead us to design the ambient visual cue as an instrument to convey the current state of the AV system. First, we will briefly discuss why communicating reliable information in autonomous driving is beneficial. Next, we will introduce the "Multiple Resource Theory" [12], a crucial concept in our approach to delivering reliable information to the driver. Finally, we will present other work related to the design of the ambient visual cue.

1.1. Reliability communication

In the real world, the AV reliability will vary accordingly to its changing levels of certainty in coping with the dynamic environment [13]. Moray [14] suggests that, when interacting with an automated system, monitoring should be proportional to the reliability of the system. Therefore, conveying the current level of reliability could help drivers of L3 vehicles adapt their supervision strategy and the level of secondary task engagement [13] and help them better identify states that require their attention [14].

According to Norman [15], appropriate automation design should continually provide feedback and interact with operators. The feedback design in automation has been considered a potential solution for the out-of-the-loop problem [14]. Moreover, providing users with ongoing reliability feedback can better facilitate appropriate trust [16]. Previous studies showed that the continuous reliability information led to faster and more precise responses to a take-over request [17–19], better situation awareness [17], and more on-road glance behaviour in low-reliability situations [20]. Furthermore, Helldin et al. [19] found that drivers provided with uncertainty information could, to a higher degree, perform tasks other than driving without compromising driving safety.

1.2. Multiple resource theory

Multiple resource theory by Wickens [21] explains how dual-task performance decreases time-sharing ability. According to this theory, only limited resources are available for information processing. Multiple resource theory distinguishes two channels of visual processing, the focal and the ambient vision [22,23]. Focal vision supports object recognition and high acuity perception. Ambient vision, distributed across the visual field, is responsible for the perception of orientation and movement. Wickens [12] suggests that when two tasks use different channels, time-sharing will be better than the contrasting scenario where two tasks share usage of the same channel and compete for like resources.

Drivers involved in non-driving related tasks (NDRTs) are likely to look away from the road [24], e.g., at their smartphone or other passengers. Therefore, their focal vision will be occupied. Based on the multiple resource theory, peripheral displays can communicate AV reliability via ambient vision without inducing resource competition. Such peripheral

displays show a high potential to raise drivers' situation awareness and improve their take-over performance [25]. Moreover, peripheral displays convey information without demanding much attention or cognitive effort [26].

1.3. Ambient light

The ambient light was designed following the findings of the subsequent studies. Borojeni et al. [26] used a LED stripe behind the steering wheel to deliver a TOR. They found the ambient light to cue focal attention appropriately to the operational context without capturing focal attention itself. However, they did not include reliability information in this study.

Faltaous [13] conducted a user study to investigate the effect of communicating the system's reliability via a feedback bar. They used a LED strip with a continuous colour gradient from red to green. Based on the results, they derived five recommendations for reliability communication: (1) to provide training sessions with the interface; (2) to use distinct colours instead of a continuous colour spectrum; (3) to deliver TORs via a haptic or auditory channel and reliability via visual channel; (4) to adapt feedback modalities according to the nature of the driving task; and (5) to employ a relatively slow update rate of the reliability cue.

Finally, Zirk et al. [27] compared a four-stage likelihood alarm system (4-LAS), three-stage likelihood alarm system (3-LAS), conservative binary alarm system (con-BAS), and liberal alarm system (lib-BAS). They found that the con-BAS yielded the best performance, which they attribute to the absence of any cry wolf effect [28] and alarm fatigue. Contrary, the worst performance was yielded by the lib-BAS and a 3-LAS. Finally, the 4-LAS yielded significantly better results than both 3-LAS and the lib-BAS in terms of performance and absence of the cry wolf effect. The authors found that the 4-LAS induced the same behavioural strategies as the con-BAS system. They found that participants tend to collapse the four levels into two and treat the 4-LAS in the same way as a con-BAS. This finding is in accordance with the theory of categorical perception, which claims that human visual system separates colors into discrete categories rather than as a continuous function of luminance [29]. Hence, Zirk et al. recommend using a discrete con-BAS when possible (e.g., when a miss-prone system is not safety-critical) and the discrete 4-LAS when the fail-safe engineering approach must be followed[30].

1.4. Measuring driver's cognitive states

Performance-based (e.g., reaction time [31], vehicle jerk [11]) and self-assessment (e.g., questionnaires [32], interviews [33]) methods have been traditionally used to assess driver's cognitive states. With the development of low-cost, non-invasive sensors, psychophysiological measures became more available to be employed in driving research [34]. Electroencephalography (EEG) is perhaps one of the most widely used psychophysiological methods to study the neural indices of cognitive effort and attention allocation in the driving context. The most common use of EEG in the driving context is to study brain activity during states of mental over- or under-arousal (for a recent review, see [35]). Employing psychophysiological measures allows assessing cognitive states continuously, unobtrusively, and without disturbing the real-time task. Therefore, psychophysiological measures provide an additional value when combined with subjective and performance-based measures [34].

Most researchers jointly report that an increased mental workload (MW) leads to a decrease in parietal alpha power [36–39], and that alpha waves are associated with relaxation and tend to increase in states of lower alertness, and mental vigilance [40,41]. Frontal theta power was reported to increase with increasing MW [36,37,42–44], however, the findings in frontal theta power are less consistent. Some authors used a ratio between theta and alpha power to operationalize MW [45–47]. Holm et al. suggest a *Task-Load Index* (TLI), which is calculated as Fz theta/Pz alpha ratio and increases with increased MW

[48]. Pope et al. [49] proposed an *Engagement index* (EI), which is calculated as beta power divided by the sum of alpha power and theta power at the Cz, Pz, P3, and P4 electrodes.

1.5. The importance of conveying reliability

Drivers must be well prepared for the control transition because an inappropriate reaction to a TOR could result in a safety-critical event [50]. However, drivers of L3 AVs may focus on tasks which are not related to driving [1]. Hence, they might not pay attention to the vehicle and the traffic environment before a TOR is delivered. But how can we expect distracted drivers to respond adequately, with the required precision, and in the desired time frame?

Increasing the predictability of TORs could give drivers time to come back in the loop [51,52] and become calibrated to the vehicle dynamics [2]. Drivers would be able to perceive the complex traffic environment, comprehend the situation, and predict future states. Enhancing the situation awareness would lead to improved decision-making, therefore improving the take-over behaviour [3,52].

Future systems should provide continuous feedback on the automation states [15], and additional support before the control transition [52]. Using the peripheral vision channel to deliver the support feedback could deliver the necessary information without posing additional demands on drivers' processing capacity [12]. The four-stage setting seems to be the most ergonomic way of communicating the reliability information, ensuring that the system is neither too conservative (resulting in a miss-prone system) nor too liberal (resulting in the cry-wolf effect) [27]. When the TOR is predictable, drivers will have sufficient time to evaluate the current situation and get calibrated to the vehicle dynamics, which will result in a smooth, calm transition of control represented by low vehicle jerk [8–10].

1.6. Hypotheses and research questions

This study aims to evaluate the efficiency of a four-stage likelihood alarm system conveying the current level of reliability on drivers' performance. The alarm system was designed as an ambient light in the vehicle's cabin. We hypothesize that the ambient light will help drivers of L3 vehicles, who are distracted by a non-driving related task, better predict situations when the AV system requests drivers to take over. Increasing the predictability of TORs will therefore provide drivers with additional time to restore situation awareness and get calibrated to the vehicle dynamics, resulting in better take-over performance without increasing MW. To test our assumptions, we formulate the following hypotheses and research questions:

H1: The ambient light conveying reliability of the AV system does not increase drivers' mental workload.

H2: Drivers equipped with the ambient light have better take-over performance.

RQ1: Do drivers consider continuous reliability feedback useful in predicting TORs?

RQ2: How do drivers evaluate the four-stage settings of the ambient light?

2. Materials and Methods

2.1. Apparatus

The experiment took place in a highly immersive, fixed-base driving simulator with a realistic car mock-up. Only the front screen of the driving simulator was used due to technical issues. A LED stripe was mounted around the entire windshield of the car mock-up (see figure 1). This setup was chosen to provide a wide viewing angle, assuring that the light is equally distributed in the cabin of the vehicle and within the peripheral vision of participants, no matter where they focus or what is their seating position. The colours of the ambient light were optimised in a pre-study with five subjects evaluating the brightness and colour settings. We used four colour settings of the LED stripe: green (#00ff00), yellow (#ffff00), orange (#ff4500), and red (#ff0000).

A non-driving related task (NDRT) was presented on a dashboard screen with a size of 390 to 275 mm. The screen was placed in the middle console of the vehicle mock-up. We used the feature-based visual search task, which is grounded in the feature-integration theory of attention by Treisman and Gelade [53]. We chose this NDRT because it competes with the driving scene for limited spatial attention resources and is self-paced, hence every participant can adapt their effort according to the demands of the primary task and their level of fatigue. Thirty-nine icons with the letter "O" and one icon with the letter "Q" were presented. The icons were grouped into four fields (10 letters per field; each icon was 14 to 14 mm large), and the position of the target letter "Q" was randomised for each trial. To minimise the muscular noise in the EEG signal, participants responded to the NDRT on a keyboard placed under the dashboard screen in the reach of their right hand. The keyboard had four buttons corresponding to the four fields of the visual search task (see Figure 1). Participants were trained to press a button corresponding to the field where the letter "Q" is currently located.



Figure 1. The vehicle mock-up with LED light around the windshield and NDRT presented on the dashboard screen

2.2. Materials

We administered a German adaptation of the Driving Activity Load Index (DALI; [32]) to assess subjective MW. The scale consists of five items representing the factors contributing to MW (effort of attention, visual demand, auditory demand, temporal demand, and situational stress) in the context of driving.

2.3. Task and Experimental Design

Participants monitored an SAE L3 [1] vehicle. The experiment consisted of four driving scenarios, each approximately 10 minutes long. The scenarios were similar in length, road curvature, and the number of events. The order of the scenarios was counterbalanced using the balanced Latin square. There was light rain and fog consistently spread throughout the scenarios. Each scenario started once participants activated the automation. At the end of each scenario, the AV autonomously stopped in a designated parking area. To simulate the *out-of-the-loop phenomena* [6], which is expected in real-world L3 vehicles, participants conducted the NDRT while the vehicle was operating in an automated mode.

When operating in the automated mode, the ego vehicle traveled at 27.7 m/s on a rural road with mild oncoming traffic. We implemented 20 slower lead vehicles traveling on the same lane as the ego vehicle. The lead vehicles traveled at a speed of 13.85 m/s.

Once the ego vehicle was 100 m behind the slow lead vehicle, the lead vehicle either (a) accelerated and left or (b) initiated deceleration of -2 m/s^3 , which resulted in a TOR for the driver of the ego vehicle. We implemented ten vehicles that accelerated and ten vehicles which decelerated. Drivers were trained that, in case of a TOR, they are required to disengage from the secondary task, understand the current traffic situation, deactivate the automation, perform the necessary manoeuvre (overtake the slow lead vehicle), re-activate the automation, and return to the secondary task.

We implemented a between-group design. Both groups experienced the same training and driving scenarios. Nevertheless, we manipulated the settings of the ambient light delivered by the LED stripe around the windshield. In order to avoid bias in the EEG data originating from the absence of any light, participants from the group without the reliability information experienced green and red ambient light. The green light was continuously on during the automated drive and switched to red light once participants disabled the automation and drove manually. Therefore, no reliability information was conveyed via the LED light.

Participants from the group quipped with the ambient light conveying reliability were informed that the light changes its colour depending on the level of reliability on which the AV is currently operating. Participants were instructed that as the colour cascade (green-yellow-orange-red) progressed, the system's reliability decreased and the likelihood of a TOR increased. We explained that the green light indicates a high reliability of the AV system and low probability of a TOR; the yellow light shows slightly decreased reliability of the AV system; the orange light indicates moderately reduced reliability of the AV system; the red light indicates low reliability of the AV system and high probability of a TOR. We recommended participants from the experimental group use the reliability information to monitor the AV system and predict the TORs.

We implemented several false alarms (FAs) in the group equipped with the reliability light. In a FA situation, a decrease in reliability was cued via LED light, but no TOR followed. We implemented eight green-yellow-green FAs; four green-yellow-orange-green FAs; and two green-yellow-orange-red-green FAs. Nevertheless, we did not include any system failures. Therefore, all TORs in the experimental group were preceded by a gradual decline in reliability. The decline in reliability was conveyed via a discrete change to yellow, orange, and red light. The yellow and orange phases were three seconds long. The red phase was 1 second long, followed by an auditory TOR.

2.4. Procedure

To minimise the time spent in the simulator due to ongoing COVID pandemics, participants received an email with information about the experiment, informed consent, and a demographic questionnaire one day before the experiment. Upon arrival at the laboratory, participants were asked to sign the informed consent and data protection sheet. Then we introduced the driving simulator and the safety regulations. Comprehensive training in the transition of control was provided, lasting approximately 10 to 15 minutes. First, participants drove manually for several minutes to learn how to operate the vehicle, followed by several minutes of autonomous driving. Afterwards, participants learned how to conduct the NDRT. In the last stage of training, we explained the experimental task, and participants experienced two TORs. Further explanation and another driving trial followed whenever participants misunderstood the task and/or did not react accordingly after a TOR as instructed. Therefore, the take-over manoeuvre was trained, and every participant learned how to react after a TOR before the experiment started. After the training session, we asked participants to fill in the DALI, LETRAS-G, and the Cooperation scale.

After the training, we set up the EEG in another room. Participants were offered snacks and drinks during the EEG montage. Once the signal quality was sufficient, participants were seated in the simulator, and we explained the task again. Moreover, we explained the settings of the LED light, which was turned off during the training session.

Once the participants understood the instructions well, we initiated the experiment. First, participants experienced the four driving scenarios with a short break. At the end of each scenario, we asked participants to fill in the DALI and the Cooperation scale. Moreover, participants again filled in the LETRAS-G at the end of the last scenario. After that, we removed the electrodes, participants were offered to wash their hair, and the interview was conducted. In the end, participants received financial compensation.

2.5. Participants

We recruited 51 participants and divided them into two gender- and age-balanced groups. Nine participants were excluded from the final analysis due to technical problems recording the EEG data. The experimental group consisted of 21 participants (7 men and 14 women; age $M = 22.62$; $SD = 4.40$). The control group consisted of 21 participants (7 men and 14 women; age $M = 23.00$; $SD = 3.76$). All participants had normal or corrected to normal vision, no known neurological or psychiatric disease, and were fluent in the German language. All participants had a valid German car driver's license, on average for 5.40 years ($SD = 3.68$), drove up to 10 000 km per year, and 59.50 % of participants drove their car at least once a week. The majority of participants (71.4 %) were students of Ulm University, who were recruited via mailing lists and advertisements in the university building. All participants received financial compensation of 20 EUR or study credits.

2.6. EEG Signal Processing

The EEG was recorded using 32 Ag/AgCl active shielded electrodes placed according to the International 10-20 System with reference to the FCz position. We used a standard actiCAP snap elastic cap (BrainProducts, Munich, Germany). The impedances were kept below 25 kOhm. Data were recorded with a sampling rate of 500 Hz using a LiveAmp amplifier and the BrainVision Recorder (BrainProducts, Munich, Germany).

The raw data was exported from Brain Vision Analyzer to a .mat file and further pre-processed using MATLAB v2022a and the toolbox EEGLab v2022.0 [54]. In the offline processing, we high-pass filtered the signal at 0.5 Hz and low-pass filtered at 30 Hz. Bad channels were removed (i.e., maximum acceptable high-frequency noise $> 4 SD$ and minimum correlation with nearby channels $r = 0.8$) and interpolated. All channels were re-referenced to a common average. Finally, we used the independent component analysis with the runica algorithm to extract 32 components. Using the ICLabel plugin for EEGLab, we classified the origin of the components. We focused on the components one to seven and removed those flagged to originate in eye or muscular activity of more than 70 %. The spectral characteristics were computed using the Darbelai package version v2019.02.01.1. The alpha power is calculated in the bandwidth range of 8-13 Hz; the theta power is calculated in the bandwidth range of 4-8 Hz; the total power is calculated in the bandwidth range of 1-32 Hz. The frontal theta power was computed as a mean theta power on the Fz, F3, and F4 electrode location. The parietal alpha power was computed as a mean alpha power on the Pz, P3, and P4 electrode locations.

3. Results

3.1. Perceived mental workload

The perceived MW was measured using the DALI questionnaire [32]. We administered the questionnaire at five time points: after the simulator familiarization and after each of the four experimental rides. Using a mixed factor ANOVA with the Greenhouse-Geisser correction, we found a significant main effect of time ($F(1.78, 53.32) = 9.40$, $p < .001$, $\omega^2 = 0.07$). A Bonferroni corrected post-hoc test showed that there was a significant decline in perceived MW between the familiarization and Ride 2 ($p = .03$), Ride 3 ($p < .001$), and Ride 4 ($p < .001$). Moreover, we found a significant decline from Ride 1 to Ride 4 ($p = .01$). There was neither a main effect of group ($F(1.00, 30.00) = 0.33$, $p = .57$) nor the interaction ($F(1.78, 53.32) = 0.66$, $p = .50$). This finding suggests that the perceived MW was comparable for

participants with the ambient light and without the ambient light. The data are presented in Figure 2.

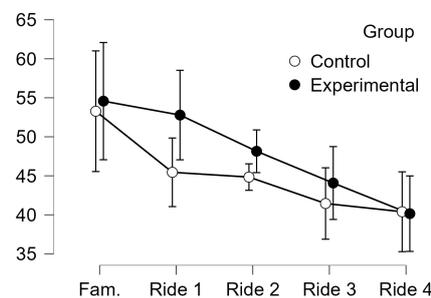


Figure 2. Results of the DALI questionnaire. The error bars represent 95 % confidence interval.

Furthermore, we conducted a Mann-Whitney U test to compare the perceived mental workload between men and women after the training session. This comparison was made to understand whether there are gender differences in how demanding is the first exposure to an AV. We found no difference between the groups [$U = 144.00, p = .511$]. We fail to reject the null hypothesis and conclude that there are no gender differences in perceived mental workload between men and women.

3.2. EEG measures of mental workload

To measure the objective MW, we analysed the relative frontal theta power and the relative parietal alpha power. The Task-Load Index (TLI) [48] was calculated as a quotient of frontal midline theta divided by parietal midline alpha. Figure 3 shows the topographical distribution of alpha and theta power over each of the four rides and the entire experiment.

To evaluate the frontal theta, we used a mixed-factor ANOVA with the Greenhouse-Geisser correction. We found a significant main effect of time ($F(2.19, 85.27) = 3.503, p = .03, \omega^2 = 0.008$). A Bonferroni corrected post-hoc test showed that there was a significant decrease in relative theta power between the first and the fourth ride ($p < .01$). There was neither a main effect of group ($F(1.00, 39.00) = 1.56, p = .22$) nor the interaction ($F(2.19, 85.27) = 1.30, p = .28$). The data are presented in Figure 4.

To evaluate the parietal alpha, we used a mixed-factor ANOVA with the Greenhouse-Geisser correction. We found a significant main effect of time ($F(1.79, 69.68) = 12.07, p < .001, \omega^2 = 0.001$). A Bonferroni corrected post-hoc test showed that there was a significant increase in relative alpha power between the first and the second ($p < .03$), third ($p = .001$), and fourth ($p < .001$) ride. Moreover, we found a significant increase between the second and the fourth ride ($p = .02$). There was neither a main effect of group ($F(1.00, 39.00) = 0.246, p = .62$) nor the interaction ($F(1.79, 69.68) = 0.33, p = .70$). The data are presented in Figure 4.

To evaluate the TLI, we used a mixed-factor ANOVA with the Greenhouse-Geisser correction. We found no main effect of time ($F(1.44, 56.08) = 2.05, p = .12$), group ($F(1.00, 39.00) = 1.56, p = .22$), or the interaction ($F(1.44, 56.09) = 0.13, p = .81$).

Both self-report and objective measures of MW suggest that there was no difference between the drivers equipped with the ambient light and those without the ambient light. Therefore, we conclude that the ambient light conveying reliability does not increase drivers' MW and accept hypothesis 1.

3.3. Take-over performance - vehicle jerk

We compared the take-over performance of drivers equipped with the ambient light and those without the ambient light. We analysed the vehicle jerk after a TOR, in the period when drivers deactivated the automation and overtook a slow-moving lead vehicle. The vehicle jerk was calculated as a second derivative of velocity. The signal was filtered using a five-point moving average filter. Then, the second derivative of vehicle velocity was calculated for every TOR. Furthermore, we calculated the coefficient of variance for every

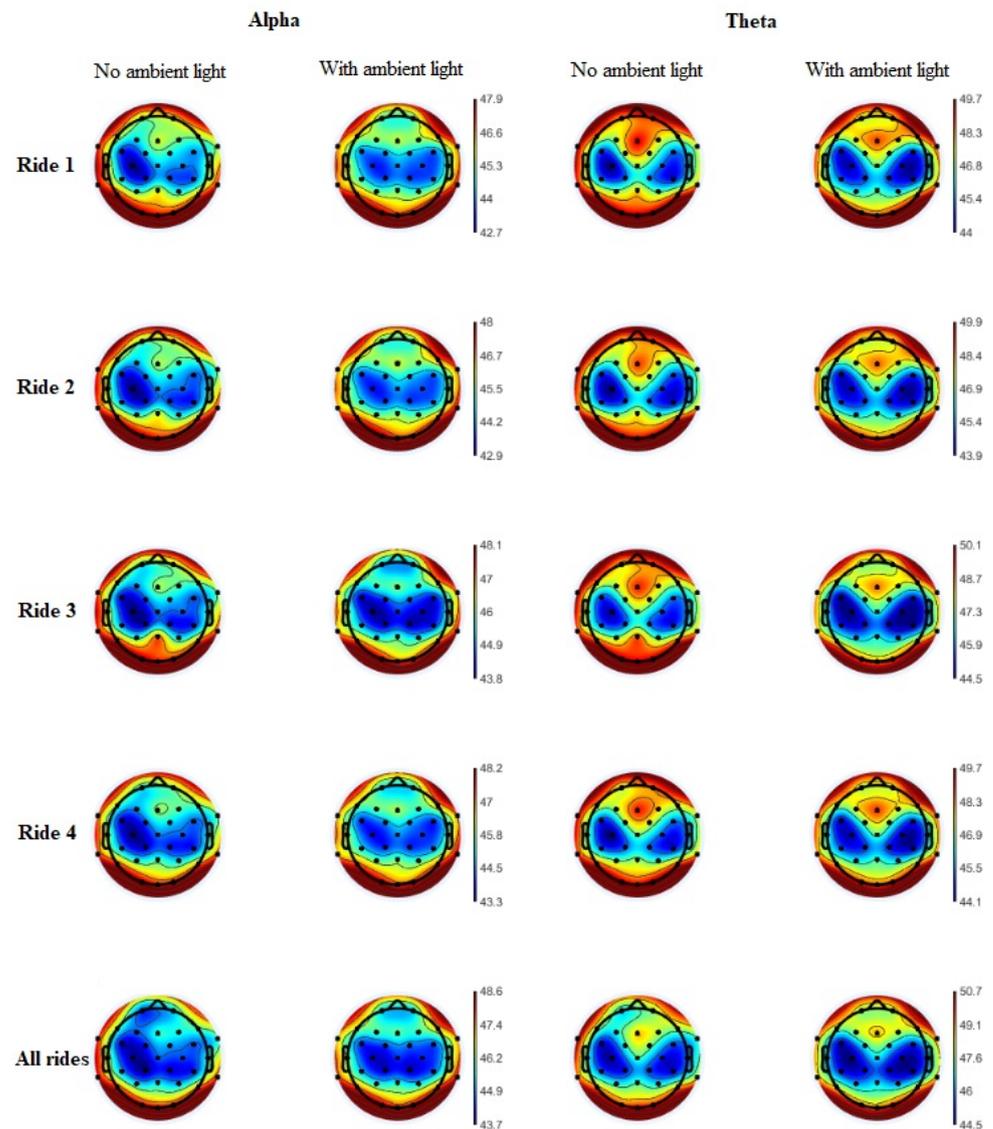


Figure 3. Averaged topography over the alpha (8-13 Hz) and theta (4-8 Hz) frequency range.

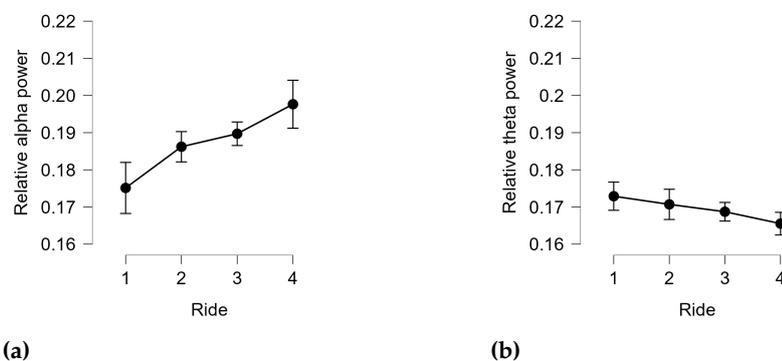


Figure 4. Relative power in the alpha and theta bandwidth over the four sequential rides. The error bars represent 95 % confidence interval.

TOR as the standard deviation divided by the mean. Finally, we standardised the values of 363

the coefficient of variance in order to compare relative low and high vehicle jerk within the sample. The results are shown in Figure 5.

Using a mixed-factor ANOVA with the Greenhouse-Geisser correction, we found a significant main effect of time ($F(3.53, 123.51) = 7.23, p < .001, \omega^2 = 0.016$) and a significant effect of the interaction between time and group ($F(3.53, 123.51) = 9.71, p < .001, \omega^2 = 0.023$). There was no main effect of group ($F(1, 35) = 0.02, p < .897, \omega^2 = 0.000$). The table with results of Bonferroni post-hoc tests can be found in Appendix A.

The results show that drivers with the ambient light performed similarly to the drivers without the ambient light in the first half of the experiment. This finding suggests that a period of training was necessary irrespective of the presence of ambient light conveying the reliability information. However, we observed increasing vehicle jerk of the drivers without the ambient light during take-overs; and there was no difference or even decrease in the vehicle jerk of the drivers with the ambient light in the second half of the experiment. This finding suggests that with training, participants equipped with the ambient reliability light performed better. Their driving style was calmer and more predictable, suggesting better control of the vehicle and better take-over performance. Therefore, we accept hypothesis 2.

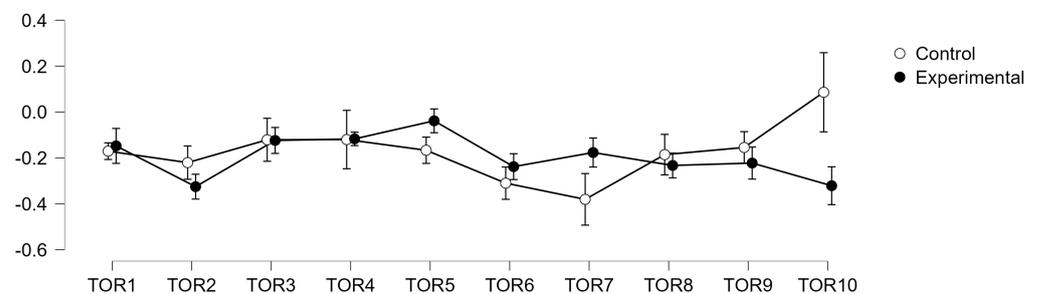


Figure 5. The mean standardized coefficient of variance of the vehicle jerk for both groups and each TOR

3.4. Overall assessment of the ambient light

To get a better understanding of drivers' opinions about the ambient light, we conducted an unstructured interview at the end of the experiment. We asked participants whether they considered the light to be useful, pleasant, obtrusive, and salient enough. Furthermore, we investigated whether they actually used the light in the way we intended it to be used, and what did they think about the four stages. Finally, we asked whether they have some recommendations on what to change.

Participants usually considered the ambient light to be a helpful tool to predict TORs. They mentioned that the light did not disturb or annoy them during the task, yet was salient enough to make them look up when there was a decrease in reliability. Most of the participants mentioned that over time, they learned to ignore the yellow-stage warning and only paid attention to the orange- and red-stage warnings. Most of the participants felt comfortable focusing on the NDRT because they *knew the system would warn them if something was going wrong*. One participant considered the ambient light to be *distracting because they could not focus on the NDRT*. Most of the participants of both groups mentioned that they needed more time to be comfortable in controlling the vehicle, even though we provided them training in the simulator. Overall, participants liked the ambient light and considered it a useful tool; however, some participants mentioned that the difference between the yellow and the orange stage was not salient enough.

4. Discussion

This study aimed to understand the effect of an ambient light conveying the current level of reliability on drivers' take-over performance and mental workload. Moreover, we aimed to get a qualitative evaluation of the ambient light design; and to understand how drivers perceive the reliability information. We conducted an experimental study in

a driving simulator. We employed a between-group design; one group was monitoring a vehicle equipped with a four-stage ambient light conveying reliability, and one group was monitoring a vehicle without the light. We analysed the vehicle jerk as a proxy of performance, perceived MW, and EEG signals as an objective measure of MW. Our data show that introducing the ambient light conveying reliability improved the take-over performance without increasing MW.

We hypothesized that the ambient light would lead to better take-over performance. We assumed that the continuous reliability information will help drivers identify states that require their attention [14] and that, therefore they will have more time to be prepared for the control transition [52]. Based on the literature, we used low vehicle jerk as an indicator of calm, relaxed driving style; and high vehicle jerk as an indicator of unpredictable, risky driving style [8,9]. We found that both groups performed similarly in the first half of the experiment. However, the vehicle jerk started increasing mid-experiment in the control group while remaining stable in the experimental group. This finding suggests a training period when drivers behave more carefully, irrespective of the ambient light. However, once drivers become used to the experimental situation, their driving style becomes riskier. Because the vehicle jerk of the experimental group remained stable throughout the experiment, we believe that the ambient light-induced more careful driving behaviour. Therefore, we assume that the ambient light could be useful to increase road safety by reducing risky driving behaviour.

To assess MW, we used both self-report and objective measures. We decided to combine the approaches to increase the robustness of our findings [34]. We observed that MW decreased over time, but there was no difference between drivers with and without the ambient light. This effect was observed using both self-report and psychophysiological measures. The learning effect can be considered as the main factor in the overall decrease of MW [41]. Our results suggest that the task was equally demanding for drivers irrelevant to the ambient light, and imply that the ambient light could support drivers and convey information without increasing MW. This outcome was expected based on the multiple resource theory [12,21,55]. Other authors came to similar conclusions about ambient displays, which seem to be useful to assist drivers with TORs [26], lane changing [56], or controlling the driving speed [57].

According to the reports and comments from the participants, the four-stage setting of the ambient light seemed adequate. We confirmed that people tend to combine the four stages into two, resulting in a sort-of-conservative binary alarm system. This is desirable [27], and we recommend using the four-stage likelihood setting in reliability communication. However, based on the feedback from the participants of the experiment, we recommend further investigating designing the light setting. However, the light design was not the main focus of this experiment.

4.1. Limitations

In this study, we did not use eye-tracking. However, it would be useful to understand the effect of ambient light on gaze behaviour. Moreover, it would be useful to understand how the ambient light influenced drivers' performance in the NDRT. Finally, our sample consists predominantly of young women, hence the generalizability of the results might be limited.

4.2. Future orientation

In future studies, it would be useful to study what are the best shades of each of the four stages; and how to locate the light in the cabin of the vehicle to increase efficiency. It is necessary to understand how to deal with different lighting conditions in relation to the brightness and colour of the ambient light; and to assess whether the light setting can be optimised for each driver. To understand whether this optimisation is necessary, we recommend studying how take over performance changes between participants of different gender, age, and personality. Moreover, before implementing such a system in real vehicles,

it would be necessary to understand the effect of a system failure on drivers' trust and acceptance of the ambient light.

4.3. Contribution

Development of a new alarm system with a setting that has not been tested before. Testing neuroergonomic methodology in the driving context. Evaluating a new method of increasing traffic safety and improving joint driver-vehicle performance.

5. Conclusions

Drivers with ambient light had a better take-over performance without increased mental workload. We found that the vehicle jerk was comparable for all drivers at the beginning of the experiment, which can be understood as a period of training. After drivers became familiar with the task, we observed a difference between drivers with the ambient light and without. The vehicle jerk of the drivers equipped with the ambient light remained low throughout the experiment, suggesting a calm driving style. The vehicle jerk of drivers without the ambient light increased in the second half of the experiment, suggesting that their driving style became more aggressive and unpredictable. Therefore, we argue that drivers with ambient light had a better take-over performance. Moreover, we measured the MW using both self-report data (the DALI questionnaire) and EEG indices of MW (frontal theta power, parietal alpha power, and the Task-Load Index). We found that the MW was comparable for all the drivers irrespective of the ambient light. Furthermore, most drivers reported the ambient light to be a useful tool to predict TORs. In conclusion, we argue that the ambient light conveying reliability improves drivers' take-over performance without increasing their mental workload. Therefore, it could improve the joint driver-AV performance during the control transition on SAE L3 automation.

Author Contributions: Conceptualization, N.F., L.C., M.B., and O.P.; methodology, N.F., L.C., M.B., and O.P.; software, N.F. and J.P.; validation, N.F., and J.P.; formal analysis, N.F.; investigation, N.F.; resources, N.F., L.C., J.P., M.B., and O.P.; data curation, N.F.; writing—original draft preparation, N.F.; writing—review and editing, J.P., M.B., O.P.; visualization, N.F.; supervision, L.C., M.B., and O.P.; project administration, M.B. and O.P.; funding acquisition, M.B. and O.P. All authors have read and agreed to the published version of the manuscript.

Funding: This project was funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement 860410.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Ulm University (protocol code 456120UMH, approved on February 15, 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data are available upon request.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AV	Automated vehicle
SAE	Society of Automotive Engineers
L2	Level 2
L3	Level 3
TOR	Take-over request
3-LAS	Three-stage likelihood alarm system
4-LAS	Four-stage likelihood alarm system
lib-BAS	Liberal binary alarm system
con-BAS	Conservative binary alarm system
EEG	Electroencephalography
EEG	Electroencephalography
EI	Engagement index
TLI	Task-Load Index
NDRT	Non-driving related task
DALI	Driving activity load index
FA	False alarm
MW	Mental workload

497

Appendix A

498

Table A1. Bonfferoni post-hoc comparisons on vehicle jerk - main effect of time

		Mean Difference	SE	t	P_{bonf}
TOR2	TOR3	-0.150	0.037	-4.029	.003
	TOR4	-0.154	0.037	-4.135	.002
	TOR5	-0.170	0.037	-4.562	< .001
	TOR10	-0.155	0.037	-4.159	.002
TOR3	TOR6	0.151	0.037	4.059	.003
	TOR7	0.156	0.037	4.181	.002
TOR4	TOR6	0.155	0.037	4.165	.002
	TOR7	0.160	0.037	4.287	.001
TOR5	TOR6	0.171	0.037	4.593	< .001
	TOR7	0.176	0.037	4.714	< .001
TOR6	TOR10	-0.156	0.037	-4.190	.002
TOR7	TOR10	-0.161	0.037	-4.311	< .001

Note. Only the significant results are presented

Table A2. Bonfferoni post-hoc comparisons on vehicle jerk - the interaction Group * Time

		Mean Difference	SE	t	P_{bonf}
Con., TOR1	Con., TOR10	-0.256	0.055	-4.675	< .001
Con., TOR2	Con., TOR10	-0.306	0.055	-5.584	< .001
Exp., TOR2	Exp., TOR5	-0.287	0.051	-5.667	< .001
Con., TOR3	Con., TOR7	0.259	0.055	4.728	< .001
Con., TOR3	Con., TOR10	-0.207	0.055	-3.772	.037
Exp., TOR3	Exp., TOR10	0.197	0.051	3.899	.022
Con., TOR4	Con., TOR7	0.260	0.055	4.747	< .001
Con., TOR4	Con., TOR10	-0.206	0.055	-3.753	.040
Exp., TOR4	Exp., TOR10	0.204	0.051	4.035	.013
Con., TOR5	Con., TOR7	0.214	0.055	3.897	.023
Con., TOR5	Con., TOR10	-0.253	0.055	-4.603	.001
Exp., TOR5	Exp., TOR6	0.199	0.051	3.944	.019
Exp., TOR5	Exp., TOR8	0.194	0.051	3.842	.028
Exp., TOR5	Exp., TOR10	0.283	0.051	5.588	< .001
Con., TOR6	Con., TOR10	-0.396	0.055	-7.214	< .001
Con., TOR7	Con., TOR9	-0.226	0.055	-4.119	.009
Con., TOR7	Con., TOR10	-0.466	0.055	-8.500	< .001
Con., TOR8	Con., TOR10	-0.272	0.055	-4.950	< .001
Con., TOR9	Con.rol, TOR10	-0.240	0.055	-4.381	.003

Note. Con. = Control group, Exp. = experimental group, only the significant results are presented

References

1. SAE, T. Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems-J3016. *Society of Automotive Engineers: On-Road Automated Vehicle Standards Committee* **2013**. 499
2. Mole, C.D.; Lappi, O.; Giles, O.; Markkula, G.; Mars, F.; Wilkie, R.M. Getting back into the loop: the perceptual-motor determinants of successful transitions out of automated driving. *Human factors* **2019**, *61*, 1037–1065. 500
3. Endsley, M.R. Toward a theory of situation awareness in dynamic systems. *Human factors* **1995**, *37*, 32–64. 501
4. De Winter, J.C.; Happee, R.; Martens, M.H.; Stanton, N.A. 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* **2014**, *27*, 196–217. 502
5. Sadeghian Borojeni, S.; Boll, S.C.; Heuten, W.; Bülthoff, H.H.; Chuang, L. Feel the movement: Real motion influences responses to take-over requests in highly automated vehicles. In Proceedings of the Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 2018, pp. 1–13. 503
6. Merat, N.; Seppelt, B.; Louw, T.; Engström, J.; Lee, J.D.; Johansson, E.; Green, C.A.; Katazaki, S.; Monk, C.; Itoh, M.; et al. The “Out-of-the-Loop” concept in automated driving: proposed definition, measures and implications. *Cognition, Technology & Work* **2019**, *21*, 87–98. 504
7. Kim, H.S.; Choi, M.H.; Choi, J.S.; Kim, H.J.; Hong, S.P.; Jun, J.H.; Tack, G.R.; Kim, B.; Min, B.C.; Lim, D.W.; et al. Driving performance changes of middle-aged experienced taxi drivers due to distraction tasks during unexpected situations. *Perceptual and motor skills* **2013**, *117*, 411–426. 505
8. Murphey, Y.L.; Milton, R.; Kiliaris, L. Driver’s style classification using jerk analysis. In Proceedings of the 2009 IEEE workshop on computational intelligence in vehicles and vehicular systems. IEEE, 2009, pp. 23–28. 506
9. Feng, F.; Bao, S.; Sayer, J.R.; Flannagan, C.; Manser, M.; Wunderlich, R. Can vehicle longitudinal jerk be used to identify aggressive drivers? An examination using naturalistic driving data. *Accident Analysis & Prevention* **2017**, *104*, 125–136. 507
10. Shahini, F.; Zahabi, M. Effects of levels of automation and non-driving related tasks on driver performance and workload: A review of literature and meta-analysis. *Applied Ergonomics* **2022**, *104*, 103824. 508
11. Du, N.; Kim, J.; Zhou, F.; Pulver, E.; Tilbury, D.M.; Robert, L.P.; Pradhan, A.K.; Yang, X.J. Evaluating effects of cognitive load, takeover request lead time, and traffic density on drivers’ takeover performance in conditionally automated driving. In Proceedings of the 12th international conference on automotive user interfaces and interactive vehicular applications, 2020, pp. 66–73. 509
12. Wickens, C.D. Multiple resources and mental workload. *Human factors* **2008**, *50*, 449–455. 510
13. Faltaous, S.; Baumann, M.; Schneegass, S.; Chuang, L.L. Design guidelines for reliability communication in autonomous vehicles. In Proceedings of the Proceedings of the 10th international conference on automotive user interfaces and interactive vehicular applications, 2018, pp. 258–267. 511
14. Moray, N. Monitoring, complacency, scepticism and eutactic behaviour. *International journal of industrial ergonomics* **2003**, *31*, 175–178. 512
15. Norman, D. The ‘Problem’ with Automation: Inappropriate Feedback and Interaction, not ‘Over-Automation’. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **1990**, *327*, 585–93. 513
16. Hoff, K.A.; Bashir, M. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors* **2015**, *57*, 407–434. 514
17. Beller, J.; Heesen, M.; Vollrath, M. Improving the driver–automation interaction: An approach using automation uncertainty. *Human factors* **2013**, *55*, 1130–1141. 515
18. Chen, K.T.; Chen, H.Y.W. Manipulating music to communicate automation reliability in conditionally automated driving: A driving simulator study. *International Journal of Human-Computer Studies* **2021**, *145*, 102518. 516
19. Helldin, T.; Falkman, G.; Riveiro, M.; Davidsson, S. Presenting system uncertainty in automotive UIs for supporting trust calibration in autonomous driving. In Proceedings of the Proceedings of the 5th international conference on automotive user interfaces and interactive vehicular applications, 2013, pp. 210–217. 517
20. Ebnali, M.; Fathi, R.; Lamb, R.; Pourfalamatoun, S.; Motamedi, S. Using Augmented Holographic UIs to Communicate Automation Reliability in Partially Automated Driving. In Proceedings of the AutomationXP@ CHI, 2020. 518
21. Wickens, C.D. The structure of attentional resources. *Attention and performance VIII* **1980**, *8*, 239–257. 519
22. Leibowitz, H.; Post, R. The two modes of processing concept and some implications In Beck, J.(Ed.), *Organization and representation in perception*. Mahwah, NY: Erlbaum **1982**. 520
23. Previc, F.H. The neuropsychology of 3-D space. *Psychological bulletin* **1998**, *124*, 123. 521
24. Noble, A.M.; Miles, M.; Perez, M.A.; Guo, F.; Klauer, S.G. Evaluating driver eye glance behavior and secondary task engagement while using driving automation systems. *Accident Analysis & Prevention* **2021**, *151*, 105959. 522
25. Yang, Y.; Götze, M.; Laqua, A.; Dominioni, G.C.; Kawabe, K.; Bengler, K. A method to improve driver’s situation awareness in automated driving. *Proceedings of the human factors and ergonomics society Europe* **2017**, pp. 29–47. 523
26. Borojeni, S.S.; Chuang, L.; Heuten, W.; Boll, S. Assisting drivers with ambient take-over requests in highly automated driving. In Proceedings of the Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, 2016, pp. 237–244. 524

-
27. Zirk, A.; Wiczorek, R.; Manzey, D. Do we really need more stages? Comparing the effects of likelihood alarm systems and binary alarm systems. *Human factors* **2020**, *62*, 540–552. 557
 28. Breznitz, S. *Cry wolf: The psychology of false alarms*; Psychology Press, 2013. 558
 29. Goldstone, R.L.; Hendrickson, A.T. Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science* **2010**, *1*, 69–78. 559
 30. Parasuraman, R.; Riley, V. Humans and automation: Use, misuse, disuse, abuse. *Human factors* **1997**, *39*, 230–253. 560
 31. Sena, P.; d'Amore, M.; Brandimonte, M.A.; Squitieri, R.; Fiorentino, A. Experimental framework for simulators to study driver cognitive distraction: brake reaction time in different levels of arousal. *Transportation research procedia* **2016**, *14*, 4410–4419. 561
 32. Pauzié, A. A method to assess the driver mental workload: The driving activity load index (DALI). *IET Intelligent Transport Systems* **2008**, *2*, 315–322. 562
 33. Khamis, N.K.; Deros, B.M.; Nuawi, M.Z.; Omar, R.B. Driving fatigue among long distance heavy vehicle drivers in Klang Valley, Malaysia. In Proceedings of the Applied Mechanics and materials. Trans Tech Publ, 2014, Vol. 663, pp. 567–573. 563
 34. Figalová, N.; Nasser, M.; Jokhio, S.; Mbelekani, N.Y.; Zang, C.; Yang, Y.; Peng, C.; Liu, Y.C.; Muhammed, A.P.; Tabone, W.; et al. Methodological Framework for Modelling and Empirical Approaches (Deliverable D1. 1 in the H2020 MSCA ITN project SHAPE-IT) **2021**. 564
 35. Lohani, M.; Payne, B.R.; Strayer, D.L. A review of psychophysiological measures to assess cognitive states in real-world driving. *Frontiers in human neuroscience* **2019**, *13*, 57. 565
 36. Mun, S.; Whang, M.; Park, S.; Park, M.C. Effects of mental workload on involuntary attention: A somatosensory ERP study. *Neuropsychologia* **2017**, *106*, 7–20. 566
 37. Brouwer, A.M.; Hogervorst, M.A.; Van Erp, J.B.; Heffelaar, T.; Zimmerman, P.H.; Oostenveld, R. Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of neural engineering* **2012**, *9*, 045008. 567
 38. Fairclough, S.H.; Venables, L. Prediction of subjective states from psychophysiology: A multivariate approach. *Biological psychology* **2006**, *71*, 100–110. 568
 39. Gevins, A.; Smith, M.E.; Leong, H.; McEvoy, L.; Whitfield, S.; Du, R.; Rush, G. Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human factors* **1998**, *40*, 79–91. 569
 40. Chen, J.; Ren, B.; Song, X.; Luo, X. Revealing the "Invisible Gorilla" in Construction: Assessing Mental Workload through Time-frequency Analysis. In Proceedings of the ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction. IAARC Publications, 2015, Vol. 32, p. 1. 570
 41. Figalová, N.; Delp, A.L.; Nick, A.; Chuang, L.L.; Baumann, M.; Pollatos, O. Alpha and Theta Power in Repeated Take-over Requests. In Proceedings of the Proceedings of the Neuroergonomics conference, 2021. 571
 42. Borghini, G.; Astolfi, L.; Vecchiato, G.; Mattia, D.; Babiloni, F. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews* **2014**, *44*, 58–75. 572
 43. Käthner, I.; Wriessnegger, S.C.; Müller-Putz, G.R.; Kübler, A.; Halder, S. Effects of mental workload and fatigue on the P300, alpha and theta band power during operation of an ERP (P300) brain–computer interface. *Biological psychology* **2014**, *102*, 118–129. 573
 44. Liao, Y.; Li, G.; Li, S.E.; Cheng, B.; Green, P. Understanding driver response patterns to mental workload increase in typical driving scenarios. *IEEE Access* **2018**, *6*, 35890–35900. 574
 45. Borghini, G.; Aricò, P.; Di Flumeri, G.; Salinari, S.; Colosimo, A.; Bonelli, S.; Napoletano, L.; Ferreira, A.; Babiloni, F. Avionic technology testing by using a cognitive neurometric index: a study with professional helicopter pilots. In Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2015, pp. 6182–6185. 575
 46. Kamzanova, A.T.; Kustubayeva, A.M.; Matthews, G. Use of EEG workload indices for diagnostic monitoring of vigilance decrement. *Human factors* **2014**, *56*, 1136–1149. 576
 47. Gevins, A.; Smith, M.E. Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical issues in ergonomics science* **2003**, *4*, 113–131. 577
 48. Holm, A.; Lukander, K.; Korpela, J.; Sallinen, M.; Müller, K.M. Estimating brain load from the EEG. *TheScientificWorldJOURNAL* **2009**, *9*, 639–651. 578
 49. Pope, A.T.; Bogart, E.H.; Bartolome, D.S. Biocybernetic system evaluates indices of operator engagement in automated task. *Biological psychology* **1995**, *40*, 187–195. 579
 50. Holländer, K.; Pflöging, B. Preparing drivers for planned control transitions in automated cars. In Proceedings of the Proceedings of the 17th international conference on mobile and ubiquitous multimedia, 2018, pp. 83–92. 580
 51. Merat, N.; Jamson, A.H.; Lai, F.C.; Daly, M.; Carsten, O.M. Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation research part F: traffic psychology and behaviour* **2014**, *27*, 274–282. 581
 52. Vogelpohl, T.; Kühn, M.; Hummel, T.; Gehlert, T.; Vollrath, M. Transitioning to manual driving requires additional time after automation deactivation. *Transportation research part F: traffic psychology and behaviour* **2018**, *55*, 464–482. 582
 53. Treisman, A.M.; Gelade, G. A feature-integration theory of attention. *Cognitive psychology* **1980**, *12*, 97–136. 583
 54. Delorme, A.; Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods* **2004**, *134*, 9–21. 584
 55. Wickens, C.D. Multiple resources and performance prediction. *Theoretical issues in ergonomics science* **2002**, *3*, 159–177. 585
 56. Löcken, A.; Heuten, W.; Boll, S. Supporting lane change decisions with ambient light. In Proceedings of the Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, 2015, pp. 204–211. 586

-
57. Meschtscherjakov, A.; Döttlinger, C.; Rödel, C.; Tscheligi, M. ChaseLight: ambient LED stripes to control driving speed. In Proceedings of the proceedings of the 7th international conference on automotive user interfaces and interactive vehicular applications, 2015, pp. 212–219. 616
617
618