Task load of professional drivers during level 2 and 3 automated driving

Hans-Joachim Bieg[1], Constantina Daniilidou[2], Britta Michel[2], & Anna Sprung[2]
[1]Robert-Bosch GmbH, [2]MAN Truck & Bus SE
Germany

Abstract

As level 2 automated driving systems (SAE partial automation) become more elaborate, the similarity to a level 3 system (SAE conditional automation), from a driver's perspective, is gradually increasing. We examined differences in driver behaviour concerning level 2 and 3 automation in a driving simulator experiment with 31 professional truck drivers. All drivers received specific instructions concerning differences in the driver's role in both automation levels. Despite this, drivers had difficulties in adapting their behaviour to the different demands of level 2 vs. level 3 driving. An analysis of driver reactions shows potentially critical lapses in attention during level 2 drives, when drivers were performing an engaging non-driving related task while driving. A comparison of drivers' gaze distributions suggests that these lapses are likely due to a de-prioritisation of on-road glances during task performance. These results highlight the difficulties that may accompany improvements of level 2 automation performance and underline the need for measures to assist drivers in adapting their behaviour accordingly.

Introduction and previous work

Advancing sensor technology and signal processing methods lead to a gradual improvement of automation performance in automated level 2 (SAE 2016) vehicles, resulting in fewer driving mistakes that vindicate the driver's supervisory role. From a layman's perspective, well-functioning level 2 systems more and more seem like level 3 systems (Campbell et al., 2018). These systems seemingly need no supervision, despite the fact that the driver is still considered a crucial safety factor by its designers (SAE 2016).

In both partial (SAE level 2) and conditional automation (SAE level 3) the driver's main task can be described in terms of a vigilance task (Davies & Parasuraman, 1982): In partial automation drivers monitor longitudinal and lateral control to detect and respond to silent automation failures. In conditional and higher automation modes, drivers detect and respond to requests to intervene, which are issued by the automated system when it approaches a system boundary. System designers may adjust the saliency of requests to intervene such that the signal detection task in level 3

In D. de Waard, A. Toffetti, L. Pietrantoni, T. Franke, J-F. Petiot, C. Dumas, A. Botzer, L. Onnasch, I. Milleville, and F. Mars (2020). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2019 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

automation and higher becomes relatively easy. For example, relevant design guidelines mandate the use of multimodal warnings and offer advice on colour, symbolism, and warning tones or messages (e.g., Campbell et al., 2018). No such control over signal saliency is available for level 2 driving: silent failures may take on the form of lane drifts or non-reactions (NTSB 2017).

Previous work suggests that drivers may find it difficult to appreciate the demands of a (well-functioning) partially automated vehicle. For example, Omae et al. (2005) and Llaneras et al. (2013) found that drivers were more likely to engage in non-driving related tasks that restricted their monitoring ability such as interacting with a handheld electronic device. Such results may potentially be explained by assuming that drivers lacked exact information about the automated system's capabilities or their monitoring duties and may be combated by appropriate instructions (Campbell et al., 2018).

The presented work directly compares the behaviour of instructed, professional truck drivers to examine whether the drivers are able to adjust to the differential demands of level 2 and level 3 automation.

Materials and methods

The study was conducted in MAN's fixed-base, high-fidelity driving simulator with professional truck drivers.

Participants

Of 32 participants, one aborted the experiment whose data is excluded in the following. The remaining 31 participants of the study (all males, M=42.5 years, SD=14.6, range=22-70 years) were in possession of a valid driver's license for trucks or busses (German C/CE or D/DE license, first issued on average 20 years ago, SD=13 years). Most of the drivers were currently working full-time as professional truck or bus drivers (mainly long distance), 10 of the drivers were working in part-time. Half of the drivers reported a yearly mileage of more than 100.000 km, 11 participants a mileage between 10.000 up to 100.000 km and 4 participants between 500 and 10.000 km.

Procedure

Upon arriving, participants were informed about the nature and duration of the experiment as well as the safety instructions for the simulator. All participants provided written informed consent before testing and received monetary compensation for their participation. Participants were equipped with electrodes for measuring heart rate and skin conductance. After these preparations, an initial questionnaire was filled out and participants were instructed how to perform the non-driving related task in the experiment – a quiz task.

The experiment consisted of a familiarisation drive and two test drives (*L2*, *L3*, counterbalanced) in the simulator. In the familiarisation drive, drivers received instructions on how to operate the vehicle and automation and were familiarised with the Driving Activity Load Index (DALI) questionnaire. Each test drive was preceded by a short period to obtain a baseline for the physiological measures followed by 50 minutes of automated driving in each automation condition. The *L2* test drive was designed to resemble a drive with a partially automated vehicle (SAE level 2), the *L3* to resemble a drive with conditional automation (SAE level 3). During both test drives, participants experienced three non-driving related task conditions in randomized order: no task (*none*), an auditory quiz (*auditory*), and a visual-manual quiz (*visual*) for 10 minutes. After each condition, participants were asked to fill out the DALI questionnaire. A brief break was made in between both test drives.

Automation levels

The automation levels were implemented as follows: a peripheral detection/vigilance task (PVT) was embedded within the driving period as a proxy for a silent automation failure (partial, *L2*) or take-over request (high, *L3*). The PVT comprised of a small green rectangle that randomly appeared on the simulator screen (see Figure 1). In the high automation condition (*L3*), a short sound and a bright blue LED above the steering wheel (take-over cue) announced the appearance of the rectangle 10 seconds in advance. During partial automation (*L2)* no such cue was presented. Participants were instructed to respond to the PVT by pressing a button on the steering wheel as quickly as possible and to prioritize this task over others.

To avoid potential confounds, missed PVT prompts did not result in any feedback to the driver and no particularly arousing or aversive situations, (i.e., potential crashes due to inappropriate take-over behaviour) was presented in the experiment. The rather abstract implementation of the automation HMI and of the required responses made sure that differences between the two automation conditions were reduced to the core distinguishing differences of both vigilance tasks.

Figure 1. Top: peripheral vigilance task (PVT) in the L2 condition. Bottom: PVT in the L3 condition with the take-over LED near the steering wheel.

Non-driving related Tasks

The non-driving related task chosen in the current study is based on a quiz (Petermann-Stock et al., 2013). The quiz consisted of 240 questions covering the fields of common knowledge, proverbs, movies and TV shows, sports, geography, cars and trucks. The questions of the original quiz were adapted to reduce the skill level required and to cater to the targeted audience of truck drivers (e.g., by selecting trucking specific questions). For each question, three possible answers were presented whereas only one of them was correct. Two versions of the quiz were presented to engage the driver into tasks with similar characteristics to (hands-free) telephone conversation or using an electronic handheld device: In the *auditory* condition, the question as well as the answers were read to the participant. The participant was asked to provide a verbal answer. In the *visual*-manual condition, questions were presented visually on a tablet computer. In order to reveal a possible answer, the participant had to touch the screen where an indicator (A, B, C) was presented (Figure 2).

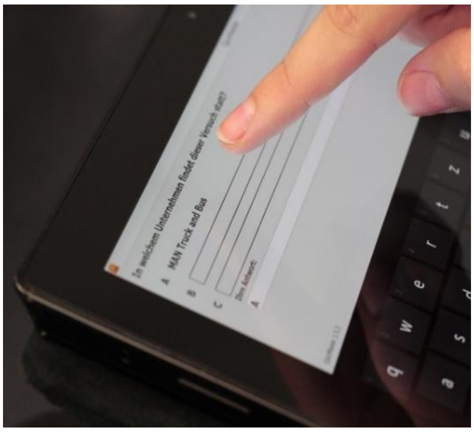

Figure 2. Visual-manual quiz.

Driving simulator & roadway

The study was conducted in the static driving simulator of MAN Truck & Bus AG. This simulator provides a 180° visual simulation as well as acoustic simulation of motor sounds and other vehicles. It is a mock-up of a full size TGX cabin built with aluminium profiles. The simulation software used was SILAB (Würzburg Institute for Traffic Sciences GmbH) which allows for the recording of vehicular data (e.g., velocity) as well as the integration of the physiological measurement equipment. The roadway implemented for the two test drives consisted of a 2-lane-highway with steady, but little traffic.

Eye-tracking

Two infrared video cameras (ON Semiconductor PYTHON1300, 1.3 MP) were installed in the simulator cabin near the A-pillar and centre console. Images were recorded throughout the experiment at a rate of 60 Hz. Processing of the images occurred off-line. Gaze direction information was computed using proprietary eye-tracking software (SmartEye embedded SDK v0.8.2). From this, gaze heading and pitch angles were computed. Both values were normalized using the mean gaze heading and pitch angles that were computed for each participant for the L2 automation drive without a non-driving related task. A road-centre region was defined with +/- 20° eccentricity and gaze information was classified as “on road” or “off road” based on this definition and the recorded, normalized heading and pitch angles. From this, the percentage of road centre gazes (PRC) for each condition was computed.

Self-assessment of workload

A self-assessment of workload was conducted using the Driving Activity Load Index (DALI). This questionnaire is designed for the assessment of workload during driving and addresses different factors such as perceptual load, mental workload and the driver’s state (Pauzié, 2008). All participants were asked to fill out the questionnaire during both test drives after each condition.

Physiology

In addition, participants' skin conductance and heart rate were recorded during the experiment, the analysis of which is omitted in the present paper.

Data analysis

Statistical data analysis was conducted using IBM SPSS (version 24.0) and R (version 3.5.2). If applicable, data was analysed using repeated measures ANOVAs with the factors automation level (*L2*, *L3*) and task (*none*, *visual*, *auditory*). For post-hoc analyses, t-tests for repeated measures were performed. Results of the ANOVA were corrected according to Greenhouse-Geisser whenever the Mauchly test of sphericity indicated heterogeneity of covariance. In the case of a violation of requirements, non-parametric ANOVAs (Friedman) and Wilcoxon Tests were used. Findings were considered statistically significant at $p < 0.05$.

Results

Monitoring ability

Participants' primary task consisted of an abstract detection task (PVT), which captured the core differences in terms of signal saliency between the two automation conditions (*L2/L3*). Figure 3 depicts the number of correct detection responses and missed signals per condition (three signals were presented per condition). Detection ability significantly differs between automation levels and groups ($\chi^2(31)=80.22$, $p < 0.001$). In *L3* most participants [77 – 90%] manage to react to all three out of three stimuli (3/3) which is significantly higher than in *L2* (Z = [-4.48;-2.36], all $p < 0.05$). Here, only 11-63% of the participants are able to react to all three presented stimuli. In *L2*, there are significant differences between the three conditions. When drivers are engaged in the *visual*-manual task, fewer signals are detected in comparison to the *auditory* task condition ($Z = -3.27$, $p < 0.01$) or when participants performed no (*none*) non-driving related task ($Z = -4.10$, $p < 0.001$). The difference between the *none* and *auditory* condition is not significant ($Z = -1.70$, $p = 0.09$). In *L3*, differences are also found for the *auditory* and *visual* task condition ($Z = -2.26$, $p < 0.05$).

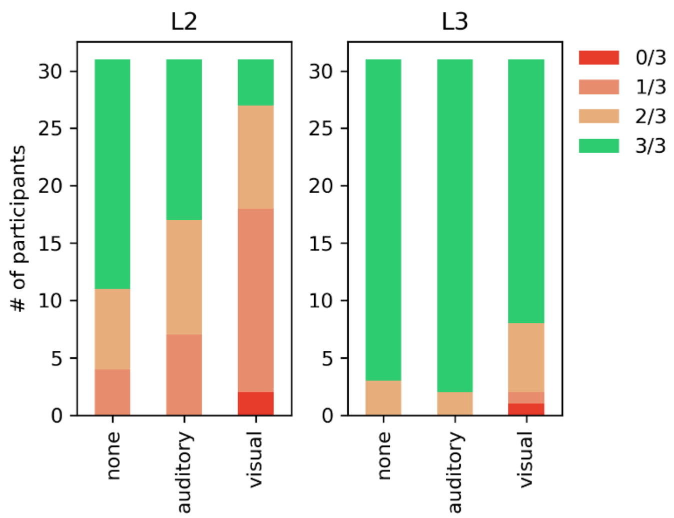

Figure 3. Number of correct responses to the detection task (PVT).

Gaze behaviour

Gaze behaviour was analysed by computing a percentage road centre (PRC) statistic per participant and drive based on angular gaze information. This analysis was performed for a subset of the 31 participants. Eight participants were excluded from this analysis either because of incomplete video recordings, issues with eye-tracker calibration and accuracy or because drivers partially made use of reading glasses. The following section presents the results for the data from the remaining 23 participants.

The analysis shows a significant difference between automation levels ($F(1,22) = 41.8$, $p < 0.01$), revealing that drivers are glancing at the roadway less frequently during *L3* (Figure 4). The analysis also shows a significant main effect of the type of non-driving related task ($F(2,44) = 268.6$, $p < 0.01$). Bonferroni corrected post-hoc comparisons corroborate the assumption that drivers' gaze is away from the road much more frequently during the *visual*-manual task condition in *L3* in comparison to no activity (mean of difference = 0.34, $t(22) = 14.9$, $p < 0.01$). Importantly, drivers' gaze is also off-road more frequently during the *visual*-manual task condition during *L2* (mean of difference = 0.36, $t(22) = 18.9$, $p < 0.01$). The comparison of PRCs in the *visual*-manual condition between *L3* and *L2* shows a small (mean of differences = 0.092) but significant difference ($t(22) = 4.2$, $p < 0.01$).

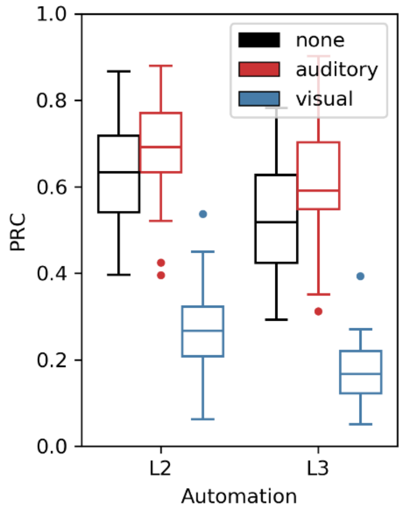

Figure 4. Percentage road centre (PRC) distributions.

Self-assessment of workload

For this analysis the data from one participant, who only partially completed all DALI questionnaires was removed. In general, the participants' ratings range in the lower end of the DALI scale. Significant differences in self-assessed workload are found between automation levels ($F(1,30) = 7.29$, $p < 0.05$) as well as between tasks ($F(2,60) = 11.15$, $p < 0.01$). Post-hoc tests show that workload is considered higher in the *visual*-manual task condition compared to the *none* condition across both automation conditions ($t(30) = [-4.50;-2.65]$, all $p < 0.05$). In addition, this condition is rated significantly lower in workload during *L3* automation ($t(30) = 4.0$, $p < 0.01$).

Individual DALI factor results are shown in Figure 6. Apart from an overall muting effect on all DALI factors, drivers' responses in *L3* in particular show a reduction of the stress factor, which in this automation condition seems to result in even lower stress ratings as the *none* condition. Note that the tactile factor was omitted in the present experiment.

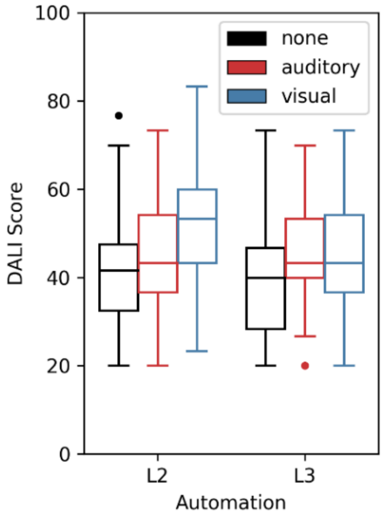

Figure 5. Subjective workload assessment results (DALI questionnaire score).

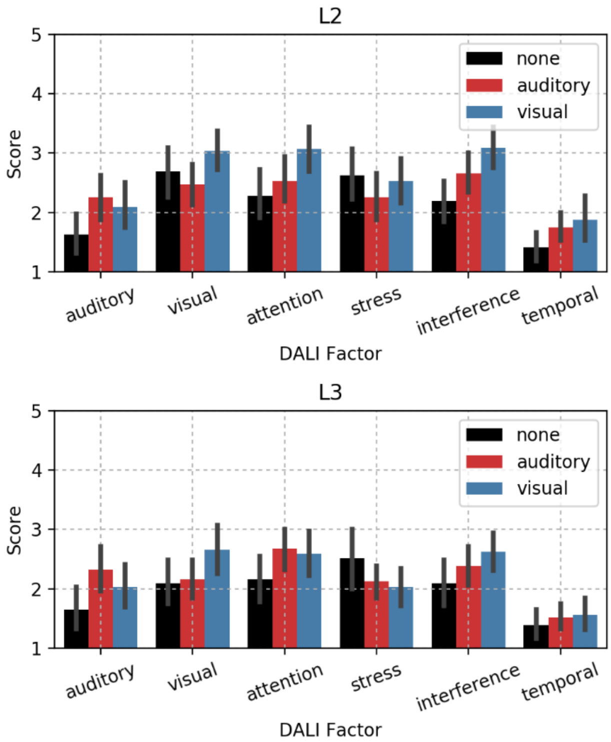

Figure 6. DALI factors.

Discussion

The present study investigated differences in professional truck drivers behaviour between level 2 (*L2*) and level 3 (*L3*) automated driving. Even though drivers received explicit instructions regarding their core, driving-related task, namely system supervision in *L2* and reaction to take-over requests in *L3*, and despite explicit requests by the experimenter to prioritise this task, drivers showed marked lapses in monitoring performance during *L2* – in particularly when they were engaged in a visual-manual non-driving related activity. For example, only two of the 31 drivers correctly responded to all three PVT (peripheral vigilance task) prompts that were presented as a proxy for a silent automation failure during the *L2* condition when they were also engaged in the visual-manual quiz task.

Failures to respond to PVT prompts (i.e., take-over requests) were also observed in *L3* driving, albeit at a much smaller frequency. In interpreting the *L3* response proportions, it must be noted that the system did not escalate the PVT prompt using e.g., additional or louder warning tones when drivers did not respond, as would be the case and easily feasible in a more comprehensive automation HMI (e.g., Llaneras et al., 2017).

PVT targets were solely presented visually in the *L2* condition, requiring participants to adapt their visual scanning behaviour: In this condition, participants should have prioritised the monitoring task in comparison to the L3 condition. Although the comparison of on-road glance distributions showed that drivers were looking at the road more frequently during L2, the difference to glance proportions in L3 was very small and comparatively low (ca. 30 % on average). This was the case despite participants noting the differential demands of both automation conditions as per the DALI questionnaire, where the visual task received lower workload ratings during L3 in comparison to L2.

Together, these findings highlight the fact that drivers seem to have difficulties in prioritizing their monitoring activity and non-driving related task adequately – despite clear instructions by the experimenter regarding the expected priorities. Reasons for this may be found in a lack of motivation regarding the primary monitoring task since the study was conducted in a driving simulator and not in a real vehicle. Yet, studies in real vehicles (e.g., Omae et al., 2005) and recent incidents in real traffic with *L2* vehicles (e.g., NTSB, 2016) suggest that drivers' priorities may be similarly misguided. Granted, the latter results and incidents were observed for non-professional drivers and it stands to reason whether professional drivers exhibit similar behaviour in a real vehicle.

Instead of factors that pertain to drivers' motivation due to the simulator setting, we suggest that the present results may be partially explained by a lack of drivers' self-awareness regarding their monitoring behaviour when performing a particularly engaging non-driving related activity. For example, time perception is known to be malleable by task characteristics (e.g., Hart et al., 1979), potentially skewing subjects' perception of the time spent with one or the other task in a dual-task scenario.

Secondly, drivers may exhibit an incomplete understanding of what constitutes necessary monitoring performance, e.g., unrealistic beliefs about failure frequencies or detection ability. Such intuitions are hard to gather from instructions but are typically acquired through interactive experience and in particularly consequences of one’s action or inaction. In the present experiment, consequences (e.g., crashes) of failing to monitor properly were not presented purposely for other reasons, but it is also expected that an absence of performance feedback regarding the monitoring task is realistic. Technical advancements will gradually decrease failure rates of automated systems. In a well-working L2 vehicle, opportunities for a reinforcement of proper monitoring behaviour will thus become rarer. Ideally, such reinforcement is provided in terms of positive reinforcement, e.g. a system failure that is compensated for by a successful intervention by the driver. With decreasing failure rates, however, potential failures may unfortunately even lead to more fatal outcomes because drivers are encountering them unprepared.

Unfortunately, past research has shown that drivers are more likely to take up non-driving related activities while driving (monitoring) an automated vehicle, presumably simply to combat boredom (e.g., Omae et al., 2005, see also review by Cunningham & Regan, 2017). Taken together, these observations may be of relevance for the designers of automated vehicles and vehicle HMIs. For example, designers may strive to implement in-vehicle systems that offer and encourage safe non-driving related activities (i.e., auditory-verbal activities) or that facilitate transforming unsafe activities into safe activities, for example, by offering services to integrate a driver’s mobile devices into the vehicle’s infotainment system (e.g., by “pairing” a device, see NHTSA, 2013). Another possibility is the introduction of on-line driver monitoring and warning systems (e.g., Llaneras et al., 2017). These systems could for example be employed to raise drivers’ awareness concerning their monitoring duties and appropriateness of their glance behaviour. Such approaches may be of particular relevance in vehicles that offer multiple automation levels (i.e., *L2* and *L3*), to assist drivers in adapting to the respective automation requirements.

Acknowledgments

This research was supported by the German Federal Ministry of Economics and Technology (Bundesministerium für Wirtschaft und Energie, BMWi) through the TANGO project (www.projekt-tango-trucks.com).

References

Campbell, J.L., Brown, J.L., Graving, J. S., Richard, C.M., Lichty, M.G., Bacon, L.P., Morgan, J.F., Li, H., Williams, D.N., & Sanquist, T. (2018). *Human Factors Design Guidance for Level 2 and Level 3 Automated Driving Concepts* (Report DOT HS 812 555). Columbus, Ohio, USA: Battelle Memorial Institute.

Cunningham, M.L. & Regan, M.A. (2017). Driver distraction and inattention in the realm of automated driving. *IET Intelligent Transport Systems*, *12*, 407-413.

Davies, D.R. & Parasuraman, R. (1982). *The psychology of vigilance*. London: Academic Press.

Hart, S.G., McPherson, D., & Loomis, L. (1978). Time estimation as a secondary task to measure workload: summary of research (N79-15634). In *14th Annual Conference on Manual Control* (pp. 693-712). Mountain View, California: NASA Ames Research Center.

Llaneras, R.E., Salinger, J., & 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 7th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (pp. 92-98). Iowa City, Iowa: University of Iowa.

Llaneras, R.E., Cannon, B.R., & Green, C.A. (2017). Strategies to Assist Drivers in Remaining Attentive While Under Partially Automated Driving: Verification of Human--Machine Interface Concepts. *Transportation Research Record*, *2663*, 20-26.

NHTSA (2013). Visual-Manual NHTSA Driver Distraction Guidelines for Portable and Aftermarket Devices, Docket No. NHTSA-2013-0137. *Federal Register*, 81, 87656-87683.

NTSB (2017). *Collision between a Car Operating with Automated Vehicle Control Systems and a Tractor-Semitrailer Truck near Williston, Florida May 7, 2016* (Report NTSB/HAR-17/02). Washington, D.C., USA: NTSB.

Omae, M., Hashimoto, N., Sugamoto, T., & Shimizu, H. (2005). Measurement of driver's reaction time to failure of steering controller during automatic driving. *Review of automotive engineering*, *26*, 213-215.

Pauzié, A. (2008) A method to assess the driver mental workload: The driving activity load index (DALI). *IET Intelligent Transport Systems, 2,* 315-322.

Petermann-Stock, I., Hackenberg, L., Muhr, T., & Mergl, C. (2013). Wie lange braucht der Fahrer? Eine Analyse zu Übernahmezeiten aus verschiedenen Nebentätigkeiten während einer hochautomatisierten Staufahrt. *6. Tagung Fahrerassistenzsysteme*. Munich, Germany: TU Munich, Lehrstuhl für Fahrzeugtechnik (FTM).

SAE (2016). *Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems* (Report J3016). Warrendale, Pennsylvania, USA: SAE International.