Effects of mental demands on situation awareness during platooning
A driving simulator study

Heikoop, Daniël D.; de Winter, Joost C.F.; van Arem, Bart; Stanton, Neville A.

DOI
10.1016/j.trf.2018.04.015

Publication date
2018

Document Version
Final published version

Published in
Transportation Research Part F: Traffic Psychology and Behaviour

Citation (APA)
Heikoop, D. D., de Winter, J. C. F., van Arem, B., & Stanton, N. A. (2018). Effects of mental demands on situation awareness during platooning: A driving simulator study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *58*, 193-209. https://doi.org/10.1016/j.trf.2018.04.015

Transportation Research Part F 58 (2018) 193–209

Contents lists available at ScienceDirect

Transportation Research Part F

journal homepage: www.elsevier.com/locate/trf

Effects of mental demands on situation awareness during platooning: A driving simulator study

Daniël D. Heikoop [d,*], Joost C.F. de Winter [b], Bart van Arem [c], Neville A. Stanton [a]

[a] Transportation Research Group, Faculty of Engineering and the Environment, Boldrewood Innovation Campus, University of Southampton, Burgess Road, Southampton SO16 7QF, UK
[b] Department of BioMechanical Engineering, Delft University of Technology, The Netherlands
[c] Department of Transport & Planning, Delft University of Technology, The Netherlands
[d] Delft University of Technology, Faculty of Civil Engineering and Geosciences Transport & Planning, Stevinweg 1, 2628 CN Delft, The Netherlands

ARTICLE INFO

Article history:
Received 3 July 2017
Accepted 18 April 2018
Available online 22 June 2018

ABSTRACT

Previous research shows that drivers of automated vehicles are likely to engage in visually demanding tasks, causing impaired situation awareness. How mental task demands affect situation awareness is less clear. In a driving simulator experiment, 33 participants completed three 40-min runs in an automated platoon, each run with a different level of mental task demands. Results showed that high task demands (i.e., performing a 2-back task, a working memory task in which participants had to recall a letter, presented two letters ago) induced high self-reported mental demands (71% on the NASA Task Load Index), while participants reported low levels of self-reported task engagement (measured with the Dundee Stress State Questionnaire) in all three task conditions in comparison to the pre-task measurement. Participants' situation awareness, as measured using a think-out-loud protocol, was affected by mental task demands, with participants being more involved with the mental task itself (i.e., to remember letters) and less likely to comment on situational features (e.g., car, looking, overtaking) when task demands increased. Furthermore, our results shed light on temporal effects, with heart rate decreasing and self-constructed mental models of automation growing in complexity, with run number. It is concluded that mental task demands reduce situation awareness, and that not only type-of-task, but also time-on-task, should be considered in Human Factors research of automated driving.

1. Introduction

1.1. Emergence of automated platooning

Automated vehicles are being developed at a rapid pace, and systems are emerging that automate longitudinal and lateral control simultaneously. A specific concept that combines longitudinal and lateral automated driving is a platoon, a group of vehicles that drive closely together in a coordinated automated manner (Bergenhem et al., 2012; Ren & Green, 1994). Platooning of automated vehicles offers advantages compared to manual driving in terms of safety, road capacity, and fuel economy (e.g., Axelsson, 2017; Kunze et al., 2011; Larson, Liang, & Johansson, 2015).

---

* Corresponding author.
E-mail address: d.d.heikoop@tudelft.nl (D.D. Heikoop).

https://doi.org/10.1016/j.trf.2018.04.015
1369-8478/

1.2. The task of drivers in a platoon

Because a platoon may involve time headways as small as 0.3 s (Ploeg, Van de Wouw, & Nijmeijer, 2014), which is at the limit of human reaction time capabilities, it unreasonable to expect that platooning drivers take over control safely in case of emergency. Nonetheless, the possibility does exist that platooning drivers have to take over control, for example in case of hardware failure (De Waard, Van der Hulst, Hoedemaeker, & Brookhuis, 1999) or in case of a voluntary driver-initiated transition (e.g., exiting or leaving the platoon; see Levitan, Golembiewski, & Bloomfield, 1998; Nilsson, 2014). Accordingly, the question arises what happens to drivers' psychological readiness after having driven in a platoon for some time.

1.3. Previous research on situation awareness and automated driving

Previous driving simulator research has found that drivers of an automated car experience low levels of workload when having nothing to do (Cha, 2003; De Waard et al., 1999; Heikoop, De Winter, Van Arem, & Stanton, 2017; Young & Stanton, 2007). Heikoop et al. (2017) found that participants in a platoon were still able to remain attentive and detect the majority (95%) of irregularly occurring stimuli (red cars) during 40 min of driving, if tasked to do so (Heikoop et al., 2017). This indicated that participants are able to retain situation awareness despite low workload.

However, when participants were allowed to engage in secondary tasks, only about 40% of the targets were detected (Heikoop et al., 2017); many participants engaged in visually demanding tasks such as eating their lunch or using their phone, rather than attending to the roadway. Other research has also found that drivers of highly automated cars are likely to pick up visual tasks such as texting, reading, and watching a DVD (Llaneras, Salinger, & Green, 2013; Omae, Hashimoto, Sugamoto, & Shimizu, 2005), as well as mentally demanding tasks such as calling on a phone or listening to the radio (Carsten, Lai, Barnard, Jamson, & Merat, 2012; Kyriakidis, Happee, & De Winter, 2015). Although it is clear that visual demands impair situation awareness, it is less clear to what extent mental task demands (i.e., engaging in a mentally demanding secondary task) influence situation awareness in automated driving.

Previous research (Gold, Berisha, & Bengler, 2015; Louw, Madigan, Carsten, & Merat, 2016; Petermeijer, Cieler, & De Winter, 2017) has found that a visually demanding task (e.g., performing a SuRT task or looking at a video) has a stronger negative effect on drivers' take-over performance than a mentally demanding task (i.e., performing an N-back task). Several driving simulator studies have even found that mental demands induced by a verbal task yielded *improved* steering behaviour and lane keeping performance (Atchley, Chan, & Gregersen, 2014; Saxby, Matthews, & Neubauer, 2017; Verwey & Zaidel, 1999). These findings can be explained with the malleable resource theory (MART; Young & Stanton, 2002), as the added demands of using a cell phone could expand the available resource pools. How a mentally demanding secondary task influences driving performance may depend on its frequency and duration of use (Neubauer, Matthews, & Saxby, 2014), the relevance of its contents to the driving task (Saxby et al., 2017), and whether the secondary task is at all engaging (Bueno et al., 2016). Although mental demands in certain cases may improve driving performance and reaction times, it remains to be clarified whether mental demands are not harmful for higher levels of situation awareness. Indeed, previous research in manual driving suggests that mental demands (i.e., listening to auditory instructions from a navigation system) reduce level 2 (comprehension) and level 3 (anticipation) situation awareness, whereas visual demands (i.e., identifying a target symbol on a tablet display every 10 s) impair all three levels of situation awareness (Rogers, Zhang, Kaber, & Liang, 2011). Similarly, in an adaptive cruise control (ACC) study, the cognitive task of using the cell phone showed deleterious effects on drivers' level 3 situation awareness (Ma & Kaber, 2005).

1.4. Aim of this research

The aim of the present research was to investigate the impact of a mental secondary task on driver situation awareness during platooning. Participants performed three 40-min platooning runs in a simulator, and their situation awareness self-reported levels of workload, and associated physiological states were measured. We hypothesized that mental secondary task demands, as induced by a verbal N-back task, would have a negative effect on participants' situation awareness. We also probed participants' mental models (i.e., the participants' understanding of the working mechanisms) of the automation after each run. A mental model is an important concept that develops with driving experience (Beggiato & Krems, 2013) and which is considered to be a facilitator of situation awareness (e.g., Biester, 2008; Endsley, 1995; Sarter & Woods, 1991; see Heikoop, De Winter, Van Arem, & Stanton, 2016 for a review). Because car manuals are hardly read (Mehlenbacher, Wogalter, & Laughery, 2002), it appears realistic to investigate drivers' situation awareness and mental models without informing participants about the workings of the automated system and the environmental cues of relevance. Thus, in contrast to most other research using normative approaches by comparing to a ground truth (e.g., Situation Awareness Global Assessment Technique; Endsley, 1988), we used concurrent think-aloud protocols (Ericsson & Simon, 1980; Salmon, Lenne, Walker, Stanton, & Filtness, 2014) and self-reported concept maps (Revell & Stanton, 2012).

2. Methods

2.1. Participants

Thirty-three participants (19 male, 14 female) aged between 18 and 66 years (*M* = 31.0; *SD* = 13.0) with at least 1 year of driving experience (*M* = 12.5; *SD* = 13.1) participated in this experiment. All participants were recruited from the University of Southampton campus through an advertisement on the university internal webpage. Inclusion criteria for participants to partake in this experiment were that they had to hold a full driver's license, be native English speakers, have normal vision and good hearing, and be in a healthy condition. Participants received a monetary incentive of £20.

Of the participants who took part, 14 indicated to be students and/or researchers, 4 to be in a managerial position, 4 in a supporting or advisory position, 2 to be administrators, and 4 to have other types of professions. The remaining 5 participants had no profession or did not disclose one. Eleven participants indicated to drive daily, 7 participants reported 4–6 days a week, 6 reported 1–3 days a week, 5 reported once a month, 2 reported less than once a month, and 2 reported they never drove in the past 12 months. Those 2 also indicated to have 0 mileage over the last 12 months, while 8 drove 1–1000 miles, 6 drove 1,001–5000 miles, 12 drove 5,001–10,000 miles, and 5 drove 10,001–20,000 miles. No-one indicated to have driven more than 20,000 miles in the past 12 months.

The study was approved by the Ethics and Research Governance Online of the University of Southampton under submission ID number 18070, and all participants provided written informed consent.

2.2. Apparatus

The simulator and electrocardiography (ECG) equipment used for this experiment were identical to a previous study by Heikoop et al. (2017). The experiment was conducted in the Southampton University Driving Simulator (SUDS). The simulator consisted of a Jaguar XJ Saloon and ran on STISIM Drive 3 software. The simulation was presented on three front screens creating a 135-degree field-of-view, one back screen for a rear view image, and two side mirror displays.

The ECG measurements were performed with AD Instruments PowerLab26T, three MLA2505 biopotential electrodes, and LabChart 8 software. 'Normal to Normal' (NN) intervals were extracted by the LabChart 8 software using the standard human ECG mode. Ergoneers' Dikablis Professional head-mounted eye tracker with D-Lab software was used to capture eye movements.

2.3. Environment

The experiment entailed the same virtual environment as Heikoop et al. (2017). Specifically, the environment consisted of an eight-lane highway (four lanes in either direction) with mild curves and hills. Participants were transported automatically in a five-car platoon, with the third car being the participant's car. The time headway between cars was approximately 0.3 s. The longitudinal and lateral movements of all cars of the platoon, including the participant's car, were identical and fully automated. At the start of each run, the platoon accelerated to 120 km/h and maintained this speed for the entire run. The platoon made seven overtaking manoeuvres per run by means of a single lane change to the adjacent lane, and back.

2.4. Procedure

Upon arrival, participants received paper instructions explaining that they would be driving three 40-min runs on a highway in an automated platoon. Furthermore, information on the procedures of the experiment, condition-specific instructions (see Section 2.5), a consent form, a figure depicting electrode placement, a demographics questionnaire, and the short pre-task version of the Dundee Stress State Questionnaire (DSSQ; Matthews, Emo, & Funke, 2005) were provided. This pre-task DSSQ queried the participants' current stress state, whereas the post-task DSSQs queried the participants' stress state regarding the task they were performing in the preceding session.

Participants read the instructions and completed the questionnaires. In addition, the ECG electrodes were attached. The three electrodes were placed in a triangular configuration, with two electrodes placed below the far ends of the collar bones and one electrode over the xiphoid process (males), or one electrode at the top of the sternum and two electrodes below the ribs on both sides (females) (see e.g., Shaffer & Combatalade, 2013).

Once the forms were completed, the Quick Association Check (QuACk) was administered for measuring the participants' mental model of automated driving (Revell & Stanton, 2016). The QuACk method consists of three steps, namely (1) asking the participants about their prior experience with the technology, (2) asking them about their common use of said technology, and (3) asking them to create a pen-and-paper mental model of how they think the technology works. In order to conduct the experiment within a reasonable time frame, we applied only step 3 of the QuACk. Specifically, participants were provided with an A3 sheet of paper, a pen, and Post-It notes. They were instructed to create a concept map of how they think automated driving works by writing down concepts they thought were present in an automated driving system on the Post-It notes, placing the Post-It notes on the A3 sheet, and indicating with arrows drawn on the A3 sheet how they think these concepts link to each other. To minimize bias in the data collection of the mental models, participants were not assisted in creating ideas for concepts or links (Revell & Stanton, 2012). Furthermore, it was emphasized that there is no wrong or right answer.

As a final step in the preparation, participants were asked to wear the head-mounted eye tracker, after which it was calibrated. To indicate readiness to begin the experiment, participants pressed a handheld button, after which the first out of three runs was started. After each run, the participants received the post-task DSSQ, the NASA Task Load Index (TLX), and instructions for the next run. Once the questionnaires were completed, participants received the QuACk map back and were asked whether they want to add, remove or alter something based on the experience they had gained during the preceding run.

2.5. Conditions

The experiment consisted of three 40-min runs, one task condition per run in counterbalanced order. Prior to each run, participants were told that they had to monitor the road and intervene when a critical situation appeared. Furthermore, they were required to "think out loud" in 2-min intervals (i.e., 2 min of speaking followed by 2 min of silence, etc.), meaning that they had to say out loud whatever they were thinking of at that moment, regardless of its content. This resulted in ten 2-min think-out-loud periods per run, which were used to assess participants' situation awareness. Participants were alerted of the start and end of a 2-min think-out-loud interval by means of a pre-recorded voice saying "please resume protocol" and "please stop protocol".

Before each run, participants received paper instructions which differed per experimental condition:

(1) 'Low Task Demands' (LTD), in which no additional tasks were provided other than those mentioned above.
(2) 'Medium Task Demands' (MTD), in which participants were, next to the tasks in the LTD condition, *encouraged*, but not *required* to perform a 2-back task by repeating the consonant that was uttered 2 letters ago. The interval between two consonants was exactly 15 s and continued throughout the entire run.
(3) 'High Task Demands' (HTD), in which the participants were *required* to, next to all the basic tasks of the LTD condition, perform the 2-back task as in the MTD condition.

Note that, despite the fact that the participants were told to intervene when required, no intervention was possible throughout the experiment.

2.6. Dependent measures

The following dependent measures were calculated per run:

- DSSQ, a self-report measure of stress states. In this experiment, the short version of the DSSQ was used (see Helton, 2004; Matthews et al., 2005). To illustrate, the Engagement scale consisted of items such as "My attention was directed towards the task", the Distress scale consisted of items such as "I felt tense", and the Worry scale consisted of items such as "I felt concerned about the impression I am making". The resulting Engagement, Distress, and Worry scale scores ranged from 0 (min) to 32 (max; 8 items scored from 0 = Definitely false to 4 = Definitely true). The standardized change scores for the three scales were calculated as: (post-score–pre-score)/(standard deviation of the pre-score) (Helton, Warm, Matthews, Corcoran, & Dember, 2002).
- TLX, a self-report measure to assess workload (Hart & Staveland, 1988). The TLX is the most widely used measure of self-reported workload (see De Winter, 2014, for a review). Scores ranged from *very low* (0%) to *very high* (100%), except for the Performance item which ranged from *perfect* (0%) to *failure* (100%).
- Correct responses (%). The percentage of correct responses on the 2-back task (applies only to the MTD and HTD conditions).
- Heart rate (bpm).
- Heart Rate Variability. A time-domain (SDNN) and a frequency-domain (LF/HF ratio) measure were used. Both the SDNN and the LF/HF ratio were calculated from the NN intervals after a default artefact filter, using software by Vollmer (2015).
- Eye movements. Gaze spread (standard deviation of the gaze coordinates), dwell time (time focused on a particular area of interest [AOI]), and PERCLOS (percentage eye closure) were used to assess participants' attention levels to the road, environment, and driving task.
- Concepts written down by the participants were categorized into four stages of automation (1) Information Acquisition, (2) Information Analysis, (3) Decision Selection, and (4) Action Implementation (Parasuraman, Sheridan, & Wickens, 2000), with a fifth category ('Other') for non-applicable concepts. The number of concepts and links between concepts were compared between the three runs.

  The categorization of concepts into the four stages was performed by the first author. He obtained input from two Human Factors experts not involved in the present study, both of whom independently rated 173 selected concepts (a subsample, 50% the size of the main sample) from the experiment. These independent ratings were discussed and used by the first author to refine his categorization. Examples of categorized concepts are as follows: (1) "Condition sensor to look at road conditions" as Information Acquisition, (2) "Calculate best route – traffic – distance – delays etc." as Information Analysis,

(3) "Artificial Intelligence" as Decision Selection, (4) "Mechanical Output, i.e. braking, acceleration" as Action Implementation, and (5) "MOT tax and insurance" as Other.

The links between the concepts' stages were then counted, to create a 5 × 5 "To and From"-matrix. Links between concepts are an indicator of participants' understanding of the cause-effect relationships (Revell & Stanton, 2016). Additionally, the number of links and the number of concepts served as indicators of the complexity of the mental model (Johnson-Laird, 2001).

- Verbal protocol analysis. Uttered statements of the participants within the 2-min intervals were transcribed, and per condition (i.e., ten 2-min intervals) visualised by means of a semantic network created with Leximancer (Smith, 2003). The three semantic networks were analysed and compared to assess participants' situation awareness (see e.g., Grech, Horberry, & Smith, 2002; Salmon et al., 2014; for similar approaches).

In the present research the following settings were applied: First, only word-like concepts, such as 'cars' or 'looking', were identified (i.e., no name-like concepts, such as 'BMW' or 'John', were identified). Second, the 'context block' (i.e., a series of sentences that are assumed to have contextual coherence) was set to 'break at paragraph', with each paragraph containing the uttered statements during a 2-min interval. Third, word variants were merged. Fourth, because our analysis is concerned with colloquially spoken text, the 'prose test threshold' setting was set to 0. Fifth, "ehm"-concepts were disregarded manually from the thesaurus.

Within the Insight Dashboard (a quantitative analysis feature within Leximancer), the three different conditions (i.e., LTD, MTD, HTD) were compared regarding the concepts' *strength* (i.e., the probability that a text belongs to a certain condition, given that this concept is present in the text, meaning the probabilities for the three conditions add up to 100%) and *relative frequency* (i.e., how frequently the concept occurs in the text for that condition).

The resulting outputs were three topical networks (one per condition) as well as a single quadrant report showing the strength and relative frequency of the 30 most prominent concepts per condition. A topical network is a two-dimensional projection of the co-occurrence between concepts, created using a linear clustering algorithm.

2.7. Statistical analyses

Comparisons between the three conditions were performed with paired t tests. A Bonferroni correction was used to account for multiple comparisons. Thus, a result was considered significant when the p value was smaller than .05/3.

3. Results

3.1. Self-report questionnaires: DSSQ and TLX

The results of the DSSQ showed a substantial loss of engagement with respect to the pre-task score in all three conditions (i.e., scores below zero, see Fig. 1). Furthermore, the HTD condition yielded significantly higher distress than the LTD

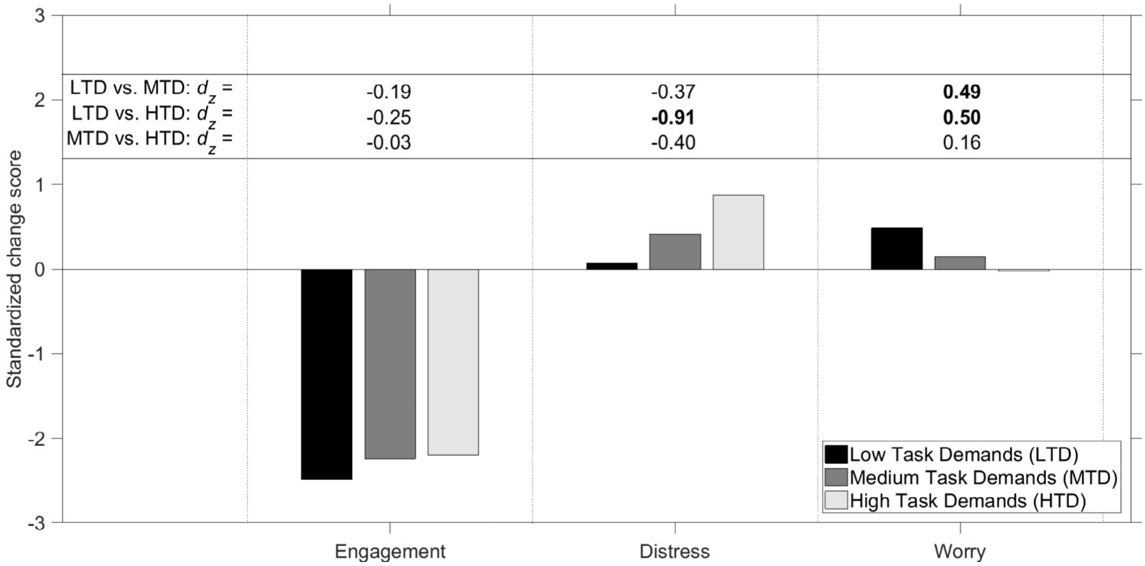

Fig. 1. Standardized change scores for the DSSQ for the three experimental conditions. LTD = Low Task Demands, MTD = Medium Task Demands, HTD = High Task Demands. For the pairwise comparisons, the Cohen's d_z effect size is shown in boldface if $p < .05/3$.

condition. The TLX showed significant differences between the three conditions, with relatively strong effects for Mental Demand, Performance, Effort, and Overall Workload (Fig. 2).

3.2. Performance on the 2-back task

The mean (*SD*) percentage correctly reported letters for the MTD and HTD conditions was 41.4% (22.2%) and 64.3% (20.9%), respectively. Pairwise comparison revealed a significant difference between the two conditions: $t(32) = -5.66$, $p < .001$. Furthermore, a decline in task performance over time occurred (Fig. 3).

3.3. Heart rate

Due to data recording errors, cardiovascular data were unavailable for 31 of 99 runs (i.e., 33 participants $*$ 3 runs). The analysis of heart rate and heart rate variability were performed for the available 68 runs.

The mean (*SD*) heart rate for the LTD, MTD, and HTD conditions were 74.6 bpm (10.3), 75.5 bpm (9.6), and 76.4 bpm (10.9), respectively. Pairwise comparisons showed that the three conditions were not significantly different: LTD vs. MTD: $d_z = -0.13$, $p = .576$; LTD vs. HTD: $d_z = -0.41$, $p = .098$; MTD vs. HTD: $d_z = 0.02$, $p = .947$.

A subsequent analysis on run number revealed clear differences between Run 1 ($M = 78.2$, $SD = 10.1$ bpm), Run 2 ($M = 75.2$, $SD = 10.3$ bpm), and Run 3 ($M = 72.8$ bpm, $SD = 9.6$ bpm). Pairwise comparisons showed significant differences: Run 1 vs. Run 2: $d_z = 1.05$, $p < .001$; Run 1 vs. Run 3: $d_z = 1.50$, $p < .001$; Run 2 vs. Run 3: $d_z = 0.76$, $p = .004$. The run effect of heart rate is shown in Fig. 4.

3.4. Heart rate variability

The mean (*SD*) SDNN for the LTD, MTD, and HTD conditions was 73.5 ms (34.9), 67.0 ms (29.0), and 64.5 ms (29.4), respectively. These effects were in the expected direction, with heart rate variability being lower for higher task demands (see Fig. 5), but pairwise comparisons showed no statistically significant differences (with Bonferroni correction) between conditions: LTD vs. MTD: $d_z = 0.11$, $p = .651$; LTD vs. HTD: $d_z = 0.62$, $p = .018$; MTD vs. HTD: $d_z = 0.21$, $p = .363$.

The mean (*SD*) LF/HF ratios for the LTD, MTD, and HTD conditions were 1.21 (0.27), 1.35 (0.35), and 1.38 (0.40), respectively. These differences were also in the expected direction with higher task demands corresponding to a higher ratio, but were not statistically significant: LTD vs. MTD: $d_z = -0.38$, $p = .128$; LTD vs. HTD: $d_z = -0.43$, $p = .083$; MTD vs. HTD: $d_z = -0.28$, $p = .231$.

3.5. Eye movements

A quality check of the eye tracker data revealed that for many participants there were drifts in the eye-gaze coordinates, presumably caused by slipping of the eye tracker on the participant's head. In addition, eye movement data were often noisy

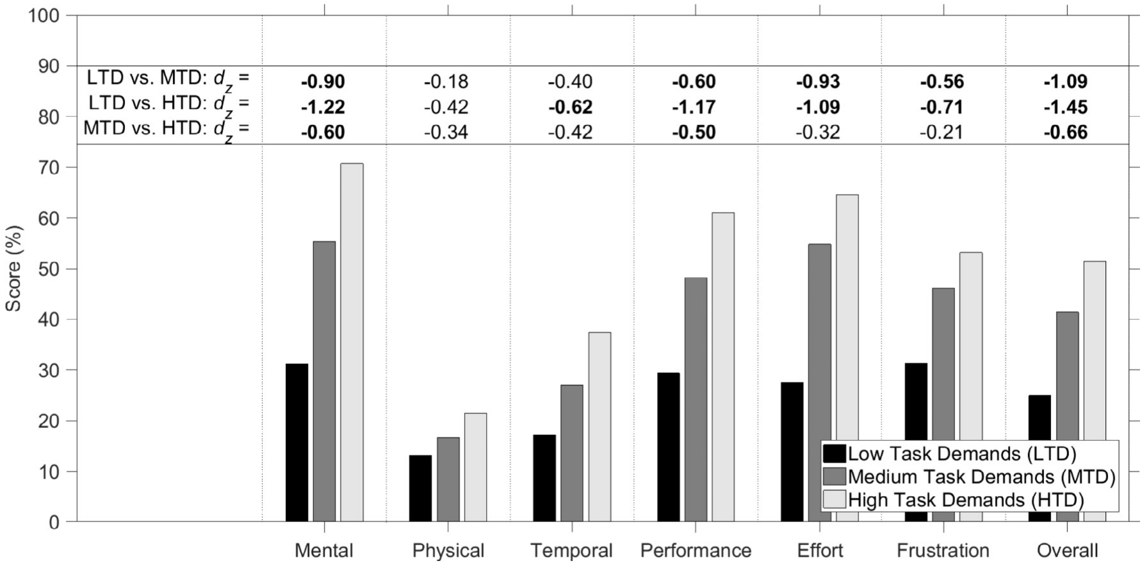

Fig. 2. Scores on the NASA Task Load Index for the three experimental conditions. The scores are expressed as a percentage and range from Very low (0%) to Very high (100%) for the Mental Demand, Physical Demand, Temporal Demand, Effort, and Frustration items, and from Perfect (0%) to Failure (100%) for the Performance item. For the pairwise comparisons, the Cohen's d_z effect size is shown in boldface if $p < .05/3$.

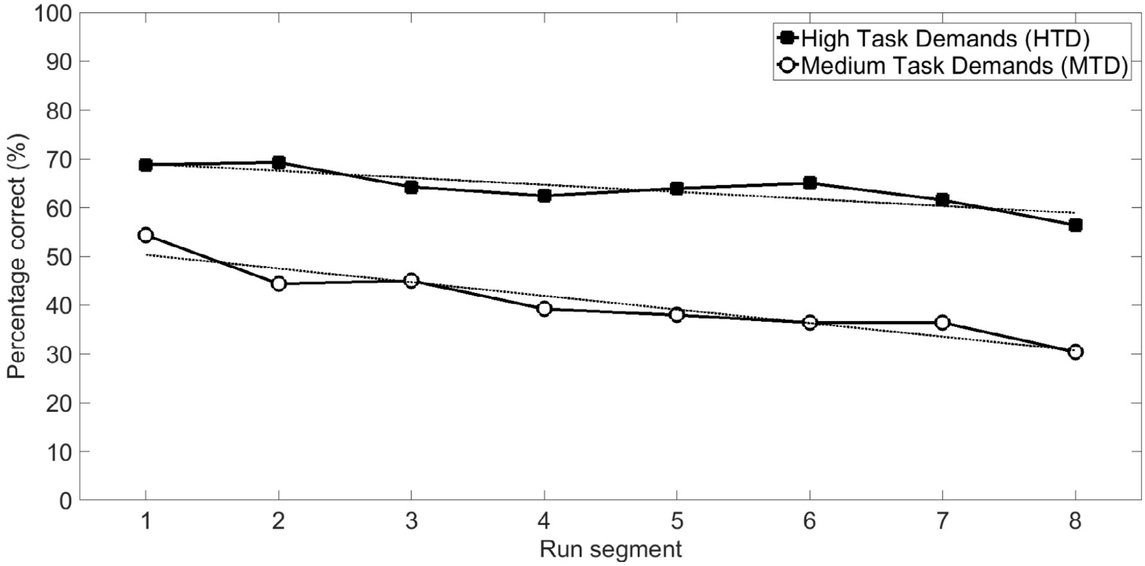

Fig. 3. Percentage of letters reported correctly in of the 2-back task per 5-min segment during the run. The dotted lines represent linear trend lines.

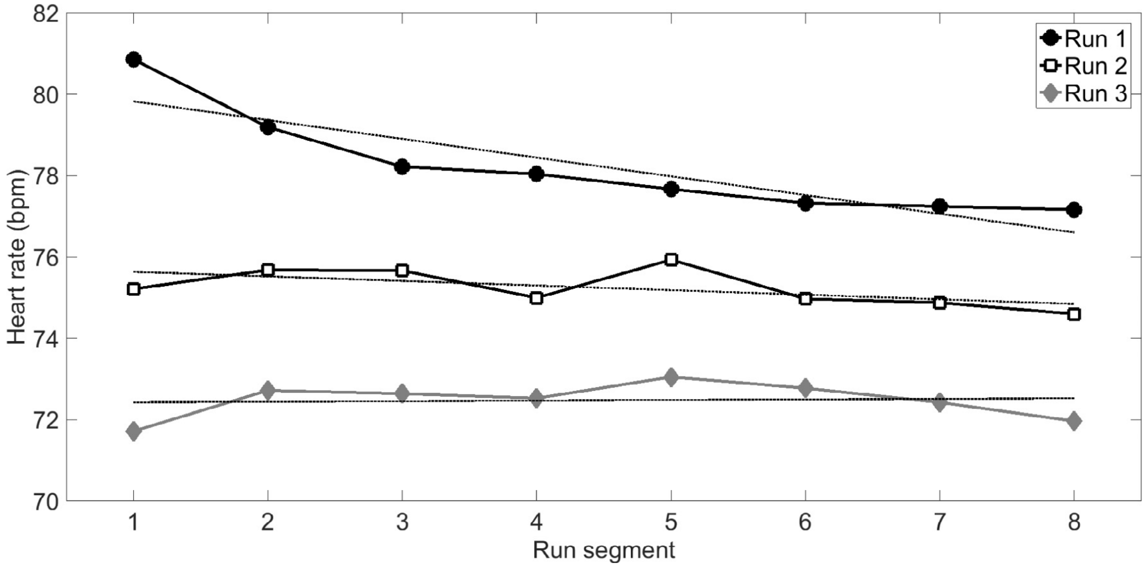

Fig. 4. Mean heart rate during Runs 1, 2, and 3 per 5-min segment. The dotted lines represent linear least-squares trend lines.

or unavailable. For these reasons, we refrained from quantitative analyses of measures such as dwell time, eye closure, or fixation duration between the three task conditions.

However, in more qualitative terms, a visual inspection of the raw data revealed that participants in all three task conditions predominantly focused on the road ahead, and occasionally glanced to the mirrors or dashboard. An illustration for one run of one participant (Run 2, MTD condition) is provided in Figs. 6 and 7. Fig. 6 shows that this participant focused on the road ahead for a large portion of the time (A), and sometimes glanced into the right mirror (B), the left mirror (C), the dashboard (D), or the rear-view mirror (E).

We performed an analysis of the horizontal gaze spread (standard deviation of the horizontal gaze coordinate) for 20 participants who did not exhibit excessive noise or missing values. The results showed no significant differences between the three conditions ($p > .2$ for the three combinations). In other words, the MTD and HTD conditions did not appear to cause evident visual tunnelling as compared to the LTD condition.

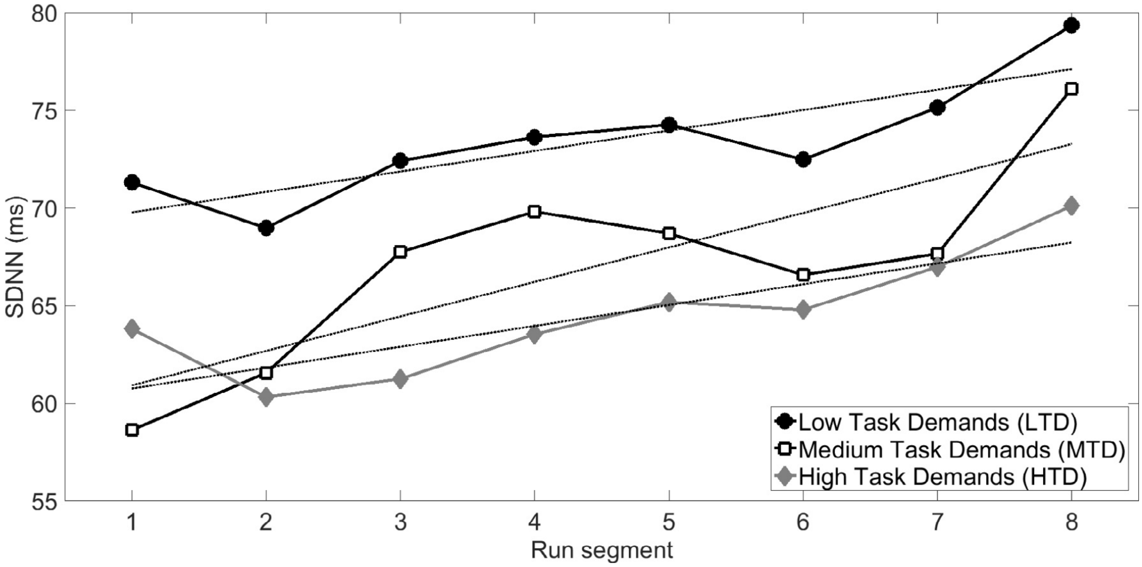

Fig. 5. Mean SDNN for each condition per 5-min segment. The dotted lines represent linear least-squares trend lines.

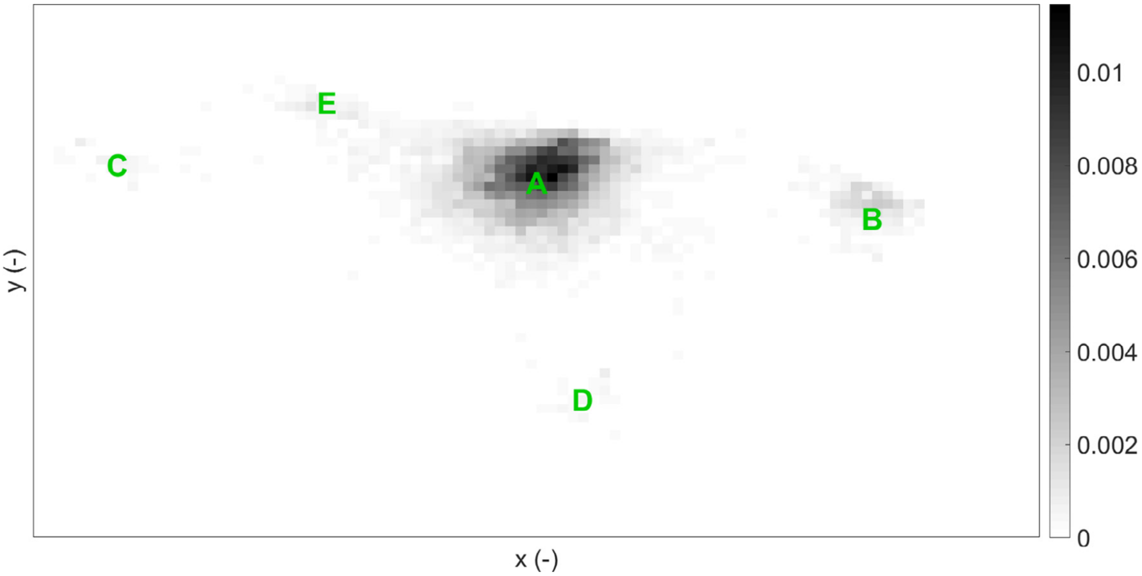

Fig. 6. Heatmap of x- and y-coordinates of eye gaze. The darkness of the pixel indicates how frequently the participant looked at this area (darker is more frequent). The total area of all pixels in the figure equals 1. The letters illustrate the approximate locations of the road ahead (A), the right mirror (B), the left mirror (C), the dashboard (D), and the rear-view mirror (E).

3.6. Mental models based on the Quick Association Check (QuACk)

An example of a mental model created with the QuACk method is provided in Fig. 8. In this case, the participant produced 11 concepts and 22 links between concepts.

Noteworthy is that none of the participants changed their mental model completely at any point during the experiment. From a possible 99 (33 participants $*$ 3 runs) times, participants changed (added or removed links/concepts, or altered the layout) their mental model 69 times.

The mean (*SD*) number of *concepts* in the participants' baseline mental model was 8.24 (3.39), and increased to 9.42 (3.39), 10.67 (3.35), 11.15 (3.62) after Runs 1, 2, and 3, respectively. Pairwise comparisons showed significant differences between all combinations: Baseline vs. Run 1: $d_z = -0.92$, $p < .001$; Baseline vs. Run 2: $d_z = -1.48$, $p < .001$; Baseline vs. Run 3: $d_z = -1.40$, $p < .001$; Run 1 vs. Run 2: $d_z = -1.20$, $p < .001$; Run 1 vs. Run 3: $d_z = -1.15$, $p < .001$; Run 2 vs. Run 3: $d_z = -0.52$, $p = .006$.

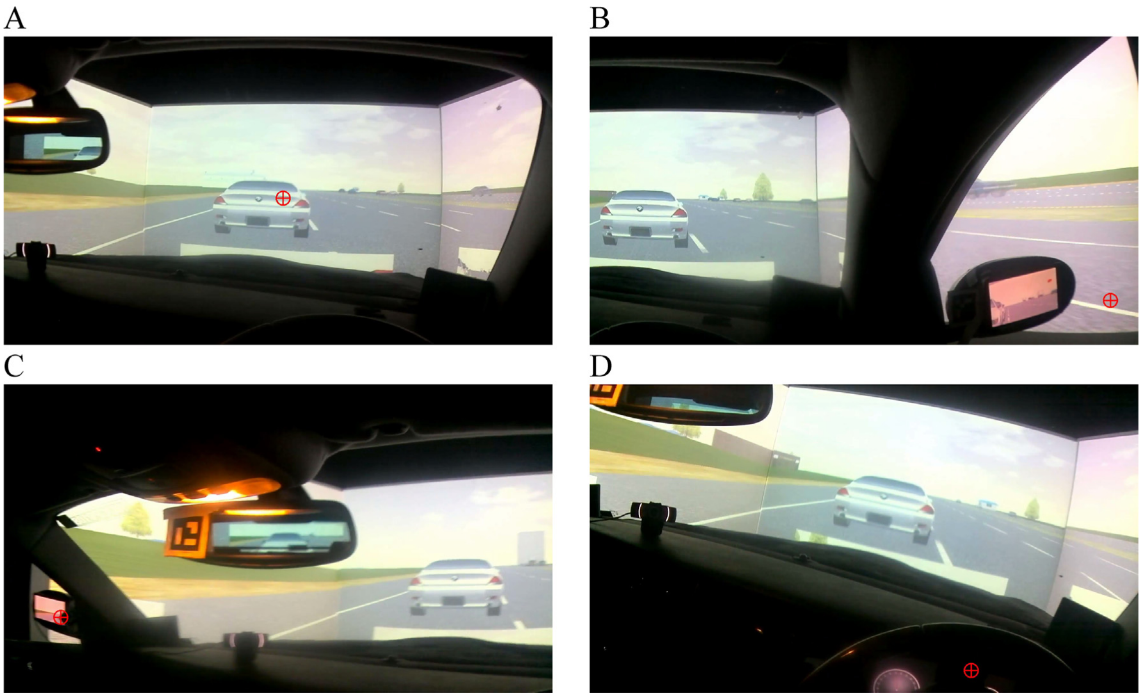

Fig. 7. Illustrative screenshots of the head-mounted forward-facing camera of the eye tracker. The red crosshair indicates the participant's momentary gaze. A = participant glances to the car ahead, B = participant glances right of the right mirror, C = participant glances into the left mirror, D = participant glances to the dashboard. The letters A, B, C, and D correspond to the letters in Fig. 6. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

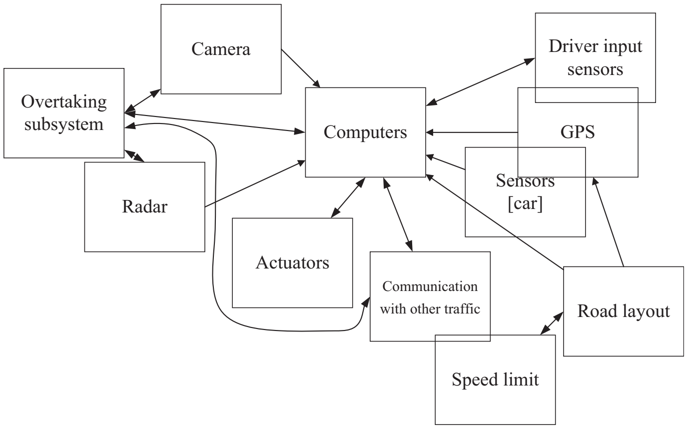

Fig. 8. An example of a participant's mental model after Run 3 (redrawn based on how the participant actually positioned the Post-It notes and arrows on the A3 sheet).

The mean (*SD*) number of *links* in the participants' baseline mental model was 11.45 (9.33), and rose to 13.42 (9.31), 16.21 (10.12), and 17.36 (10.49), in Runs 1, 2, and 3, respectively. Pairwise comparisons revealed significant differences between all combinations: Baseline vs. Run 1: $d_z = -0.90$, $p < .001$; Baseline vs. Run 2: $d_z = -1.13$, $p < .001$; Baseline vs. Run 3: $d_z = -1.18$; Run 1 vs. Run 2: $d_z = -0.76$, $p < .001$; Run 1 vs. Run 3: $d_z = -0.84$, $p < .001$; Run 2 vs. Run 3: $d_z = -0.51$, $p = .007$. It was further observed that the mean number of links and the mean number of concepts per participant were strongly correlated (Spearman's $\rho = .65$, $N = 33$).

Fig. 9 shows the results of the participants' mental models as categorized according to the stages of automation. The majority of the links between concepts involve Other concepts, whereas the least common links involve Information Analysis concepts. Furthermore, the drawn links were more often in agreement with the order of 'stages of automation' postulated by Parasuraman et al. (2000, black bars in Fig. 9) than in disagreement with that order (white bars in Fig. 9).

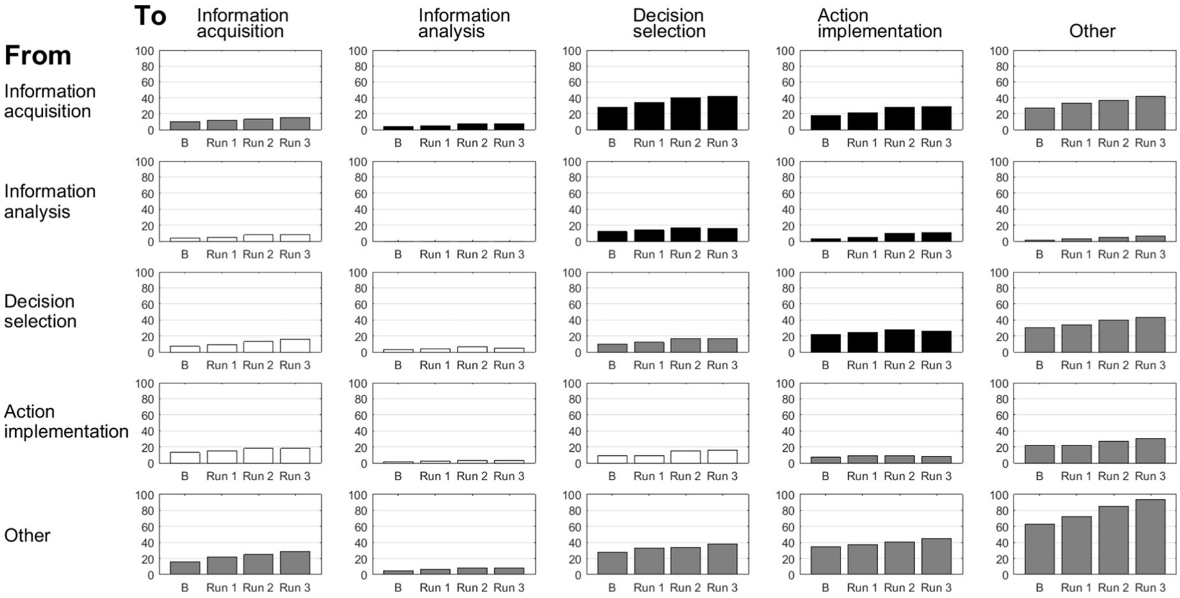

Fig. 9. Results of participants' mental models categorized into the four stages of automation (as defined by Parasuraman et al., 2000) for Baseline (B), Run 1, Run 2, and Run 3. Each bar represents the mean number of links per participant from a stage (rows) to another stage (columns). Bar graphs in black correspond to the order of the four stages (i.e., 1. Information acquisition → 2. Information analysis → 3. Decision selection → 4. Action implementation) as defined by Parasuraman et al. (2000). Bar graphs in white correspond to links that follow the opposite direction as the four-stage model by Parasuraman et al. Bar graphs in gray correspond to and from the Other concepts, and links to the same stage of automation.

3.7. Verbal protocol analysis

Of a total of 990 (33 participants × 3 runs × 10 intervals) possible 2-min intervals, 50 intervals were unavailable due to recording errors. Of the available intervals, 895 intervals contained relevant information (i.e., no untranscribable utterances, completely silent intervals, or merely containing a single word). The total number of ranked concepts for 895 intervals combined during the LTD, MTD, and HTD conditions was 2755, 2194, and 1640, respectively. The letters uttered by the participants as part of the 2-back task were not taken into account.

Fig. 10 shows the topical networks of the LTD, MTD, and HTD conditions, respectively. It can be seen that the concepts within the statements uttered by the participants were predominantly about the car in the LTD condition and predominantly about the letters (of the 2-back task) in the HTD condition.

These observations are supported by the quadrant report (Fig. 11), from which it is evident that in the MTD condition (blue), and particularly in the HTD condition (red), participants were occupied with trying to remember letters. The LTD condition (green) shows a relatively strong (towards the top) and frequent (towards the right) occurrence for situation and driving-task related concepts such as 'driving', 'road', 'car', 'overtaking', 'motorway', 'behind', 'front', 'lane', 'looking', whereas the strength and frequency of these concepts is comparatively low in the HTD condition. For example, in the LTD condition, the 'overtaking' concept had a strength of 48% and relative frequency of 3%. The corresponding strength and relative frequency for the MTD condition were 30% and 2%, respectively. For the HTD condition, the strength and relative frequency were 21% and 2%.

4. Discussion

4.1. Assessing the effects of mental demands

The present study aimed to assess driver's situation awareness as a function of mental demands during automated platooning. Additionally, the development of drivers' mental models of automated driving was investigated. Participants were transported in a simulated platoon and were requested to monitor the road and intervene whenever a critical situation occurred. In two of the three conditions, participants were either required (HTD) or requested (MTD) to perform a 2-back task by means of reporting the letter that was displayed two letters before by a pre-recorded voice through a speaker.

On a scale from *Very low* to *Very high*, participants in the present experiment reported Mental Demands of 55% (MTD) and 71% (HTD), compared to 26% (voluntary task) and 39% (visual detection task) in a previous platooning experiment by Heikoop et al. (2017). The fact that the 2-back task yielded a percentage of about 65% of correct answers is another indication that the 2-back task was indeed mentally demanding. Additionally, although the eye tracking data were not of high quality, it

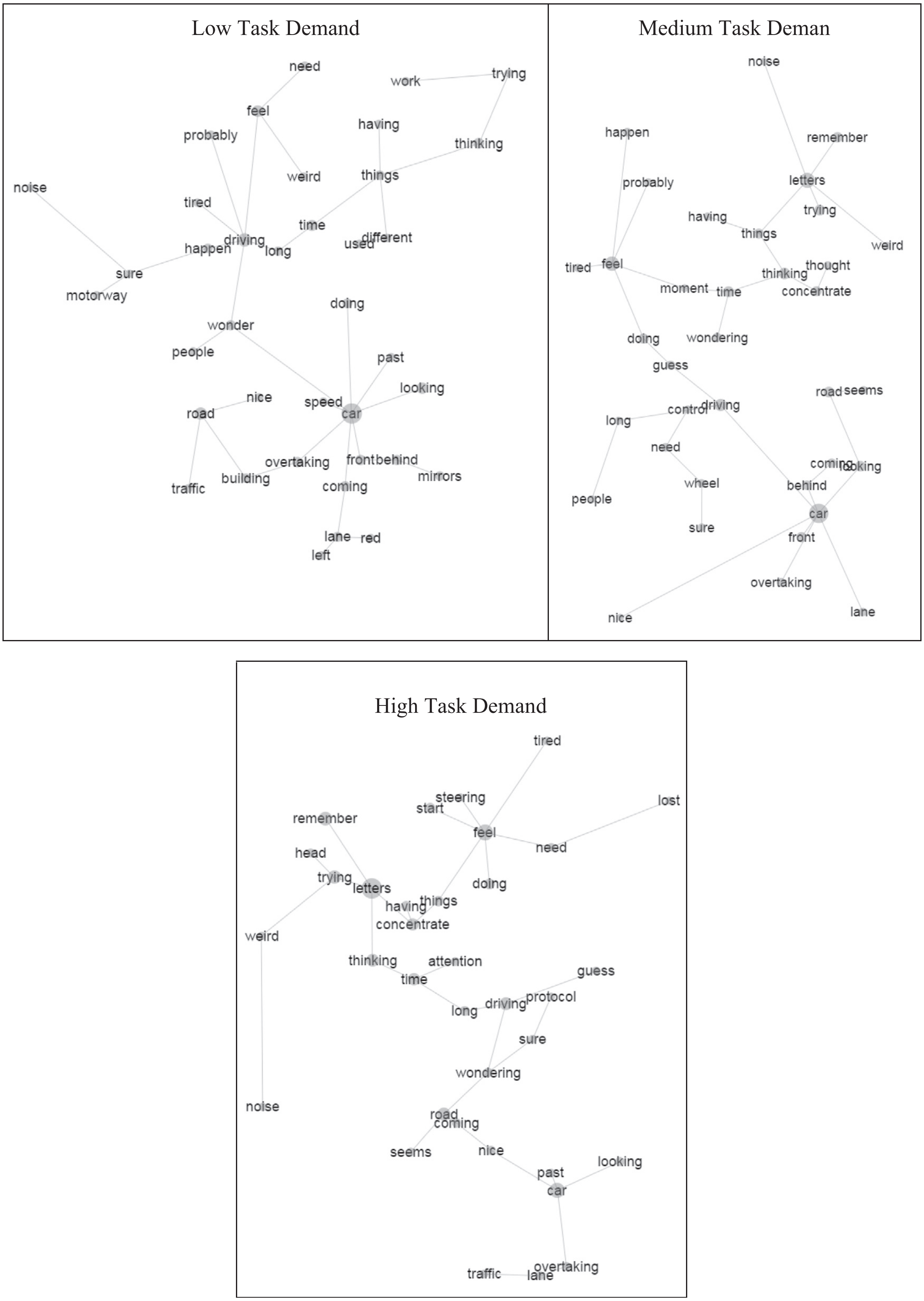

Fig. 10. Topical networks developed through Leximancer for the LTD, MTD and HTD conditions, respectively. The greater a concept node's diameter, the higher the relative frequency of the concept within the text. The links (grey lines) indicate concepts that are strongly connected.

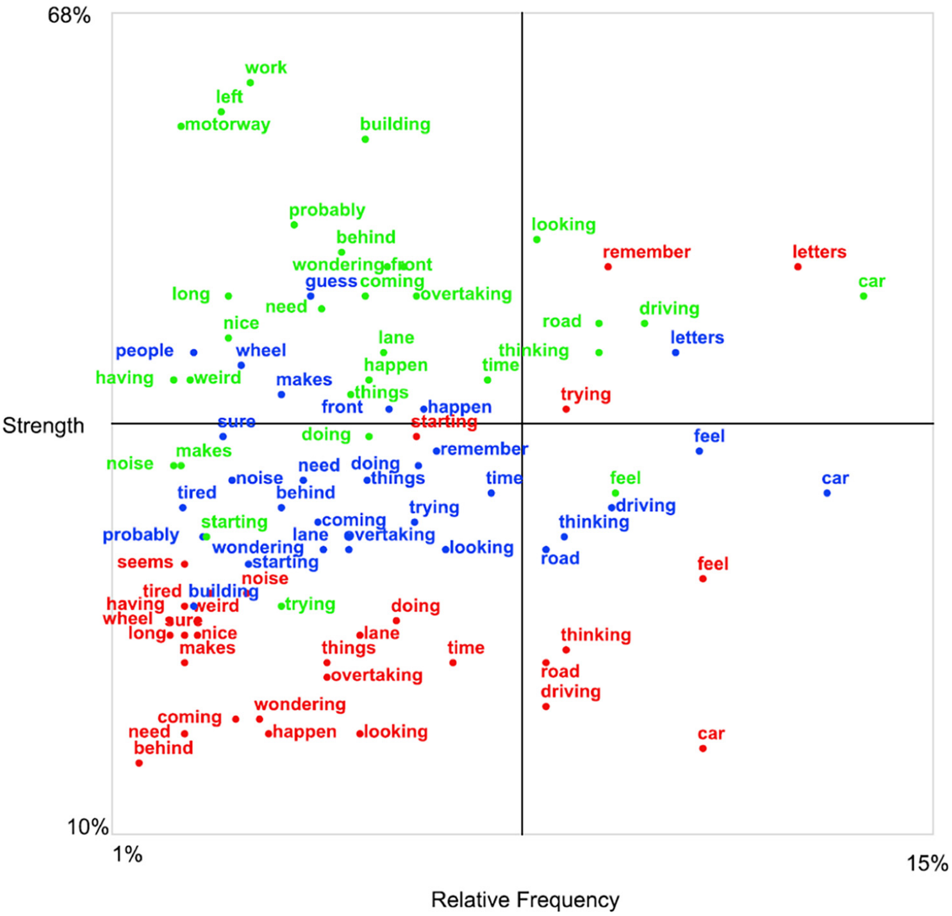

Fig. 11. Quadrant report of the verbal protocol analysis performed with Leximancer for each of the three conditions. Green = Low Task Demand; Blue = Medium Task Demand; Red = High Task Demand. The top 30 occurring concepts per condition are displayed and placed according to their relative frequency (x-axis) and strength (y-axis) in percentage. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

was clear that participants were attentive to the road in all three conditions (i.e., they were not predominantly engaging in visual secondary tasks). Furthermore, participants had a mean heart rate of 76.4 bpm in the HTD condition (a typical resting rate), and reported low levels of engagement as compared to the pre-task measurements.

Collectively, these findings illustrate that our experimental design was successful in eliciting mental demands: the mental task was not subjectively engaging, physiologically stressful, or visually distracting, yet was able to create three distinct levels of mental workload (as shown by the TLX).

4.2. Situation awareness

The verbal protocol analysis showed a clear effect of mental task demands on situation awareness. Statements regarding the remembering of letters were strong and frequent during the MTD and HTD conditions, and statements regarding the driving situation were strong and frequent in the LTD condition. Moreover, from Fig. 11 it can be seen that participants reported to be looking around in the LTD condition, whereas this was less evident in the MTD and HTD conditions, which could be indicative of the ‘look-but-failed-to-see’ phenomenon (Hills, 1980). Another finding is that with increased mental task demands the participants uttered fewer statements. A logical explanation is that the participants had to utter responses to the 2-back task; these 2-back responses were not taken into account in the verbal protocol analysis. A second explanation is that the participants were mentally occupied by the 2-back task, thereby not having enough resources left to establish their situation awareness and utter corresponding statements about their thoughts.

Our findings add to the literature in that mental demands impair situation awareness (Ma & Kaber, 2005; Rogers et al., 2011). Although drivers may be able to counter fatigue by performing a verbal task (Atchley et al., 2014), this does not imply they remain aware of the situation around them. In fact, our results indicate that having no additional task demands is best for maintaining situation awareness, as driving related statements such as looking and driving were most prevalent in the LTD condition. Accordingly, policy makers and designers of technology should be aware that the mere recommendation for drivers to engage in a verbal task, or not engage in visually demanding tasks (e.g., working, interacting with a smartphone) is insufficient to keep drivers situationally aware; mental demands alone also reduce situation awareness.

It should be noted that the observed effects were particularly strong for the ‘strength’ dimension of the verbal protocol analysis; effects were less clear for the ‘relative frequency’ dimension. This can be explained by the fact that participants in the LTD condition uttered substantially more words than in the MTD and HTD conditions. It is likely that the secondary task (that is, to report letters) interfered directly with the verbal protocol.

Also, one could wonder whether situation awareness on the level of looking and driving (i.e., level 1 situation awareness; Endsley, 1995) would be sufficient for taking over manual control. In a non-critical take-over situation, such as during exiting the platoon, this might suffice. However, in a critical or more complex situation, a higher level of situation awareness is important to act appropriately (cf. Radlmayr, Gold, Lorenz, Farid, & Bengler, 2014). Thus, it appears that multiple levels of situation awareness are necessary to be maintained by a driver of an automated vehicle. Therefore, it is suggested that at least some level of mental demand (e.g., the level of mental demand requested during the MTD condition) should be requested from the driver, in order for him/her to maintain a higher level of situation awareness. During the MTD condition, participants uttered statements regarding guessing and remembering next to statements regarding looking and driving, an indication of a balance between the three levels of situation awareness.

4.3. Mental models

A relatively novel approach was used for the assessment of mental models, namely by means of the Quick Association Check (QuACk; Revell & Stanton, 2016). Two noteworthy findings were obtained:

First, participants did appear to have a rudimentary understanding of how automation works as evidenced by the fact that links between concepts were more often in agreement with the order of the *four stages of automation* (as defined by Parasuraman et al., 2000) than in disagreement with it. Even so, participants produced highly different mental models. For example, some participants did not draw sensors or computers, but focused only on the vehicle’s basic components, such as the engine, tyres, and gears (classified as ‘Other’ in Fig. 9).

Second, during the course of the experiment, none of the participants overhauled their mental model completely. Participants appeared to stick to their original mental model and gradually added concepts and links, resulting in an increasingly complex mental model with run number. This may be explained by the fact that participants did not receive (dis)confirmatory information during the experimental runs: The automation always worked flawlessly, and hence participants may have had no incentive to alter their mental models, allowing for time to think about related concepts to be added. Our findings are different from Beggiato and Krems (2013), who found that non-experienced problems with the automation tend to disappear from participants’ mental models. The difference between our approach and that of Beggiato and Krems, however, is that in our case participants were not informed about problems that may occur with the automation; the mental models were entirely self-constructed.

Previous driving simulator research by Kazi, Stanton, Walker, and Young (2007) concurs that drivers tend to stick to their formed mental models of an automated driving system. In their experiment, participants were provided with a manual on the workings of an adaptive cruise control (ACC) system, as well as a list of features/functions of the ACC, and were subjected to either reliable, unreliable, or semi-reliable ACC over a ten-day period. The authors concluded that ‘‘conceptual models were consolidated over a short period of time, however they did not match that of designers’ model of Adaptive Cruise Control, thus better design solutions may be warranted.”

In summary, our results showed that mental models (operationalized via self-created concept maps) are *not* self-correcting, but rather become increasingly complex with time. This suggests that without prior information or training on automated driving systems, drivers could retain an inaccurate mental model (see also Kazi et al., 2007).

A limitation of our method is that the categorisation of concepts into four stages was often ambiguous. An example is ‘‘Computer”, which was classified as Decision Selection by the lead researcher, but which can also be plausibly classified as Information Analysis. Accordingly, the reproducibility of the results in Fig. 9 deserves further investigation. Another limitation is that the present study was conducted among a university population. It is likely that mental models of the general population, who may be less technology-oriented than the present university sample, may be less in agreement with Parasuraman et al.’s (2000) four stages of automation.

4.4. Time-on-task effects

This experiment showed that participants’ heart rate dropped during the course of the experiment. Furthermore, a declining trend in the percentage correct answers on the 2-back task occurred. These findings suggest that participants may have become fatigued and gradually lost their vigilance. Overall, the heart rate differed more between Run numbers than between the three task demands conditions.

4.5. Measurement issues

Although the heart rate variability measures showed effects in the expected direction (i.e., lower SDNN and higher LF/HF ratio with increasing mental demands), the effects were neither strong nor statistically significant. These observations indicate that physiological indexes are not as discriminative between mental workload conditions as self-reports. One of the issues is that heart rate itself strongly correlated with SDNN (Spearman $\rho = -0.46$ in the present experiment, $N = 29$) as well as with the LF/HF ratio ($\rho = 0.37$), which raises questions about the independency of these cardiovascular measures. Mehler, Reimer, and Wang (2011) previously found that heart rate itself was better able in detecting differences in both low and high workload scenarios than measures of heart rate variability.

The eye tracking data revealed several problems regarding quantitative analysis due to movement and slipping of the eye tracker during the experiment. Nevertheless, a qualitative analysis showed no significant differences between the three conditions on the account of visual tunnelling. We suggest that future research should encompass an ergonomic design of the eye tracker to avoid excessive slippage and movement of the eye tracker in order to improve data quality. Alternatively, a high-quality remote eye tracker rather than a head-mounted eye tracker may be considered.

4.6. Further research

The present study was concerned with drivers' psychological state; participants did not actually have to implement a response. Ultimately, safety is determined by behaviour, not by psychological state. Accordingly, we recommend that future research examines how drivers respond in safety-critical situations. If drivers behave unsafely when leaving a platoon, appropriate human-machine interfaces and training/instruction procedures may need to be developed to counteract this problem. Some previous research has already investigated driver behaviour after leaving a platoon. For example, studies showed increased driving speeds and decreased time headway during manual driving after having driven in a platoon (Brandenburg & Skottke, 2014; Levitan et al., 1998; Skottke, Debus, Wang, & Huestegge, 2014). Additionally, it has been found that manual drivers' headway and self-reported stress is affected when they drive next to a platoon (Gouy, Wiedemann, Stevens, Brunett, & Reed, 2014; Larburu, Sanchez, & Rodriguez, 2010).

The verbal protocol approach taken in this study could be further developed (e.g., by using a non-verbal mental task) in order to determine the different levels of situation awareness more precisely. Also concurrent psychophysiological measures could be used that are known to measure (levels of) situation awareness. For example, electroencephalography (EEG) could be used to relate fatigue patterns to situation awareness (e.g., French, Clarke, Pomeroy, Seymour, & Clark, 2007). Future research could also investigate which level of situation awareness is required or appropriate for different driving tasks or events. For example, future research could investigate what level of situation awareness is needed while transferring into and out of a platoon (i.e., intervening in a critical situation, or exiting the platoon, in mild or heavy traffic).

5. Conclusion

This experiment showed that mental demands of the 2-back task have a strong effect on driver's self-reported mental demands but not on their psychophysiological responses. Driver situation awareness (as analysed by a topical network through Leximancer) was impaired due to the additional mental demands.

Furthermore, clear time-on-task effects were seen in psychophysiological measures, secondary task performance (2-back performance), and the complexity of self-constructed mental models. This suggests that not only the type of task, but also time-on-task should have a role in future research on Human Factors in automated driving. Future research should concern on-road platooning experiments in which drivers have to resume manual control.

Acknowledgements

The authors are involved in the Marie Curie Initial Training Network (ITN): HFAuto (PITN-GA-2013-605817).

Appendix A.

Extensive Ranked Concept Lists for each condition. Only the concepts that make up the Topical Network of Fig. 10 are included in the list. Relevance is the percentage of occurrence of a concept relative to the most occurring concept. Therefore, the most occurring concept is 100%, regardless of its occurrence count.

Ranked Concept List LW			Ranked Concept List MW			Ranked Concept List HW		
Word-Like	Count	Relevance	Word-Like	Count	Relevance	Word-Like	Count	Relevance
Car	352	100%	Car	242	100%	Letters	165	100%
Driving	165	47%	Feel	161	67%	Feel	121	73%
Feel	150	43%	Letters	147	61%	Car	121	73%
Thinking	144	41%	Driving	119	49%	Remember	87	53%
Road	141	40%	Thinking	102	42%	Trying	78	47%
Looking	119	34%	Road	95	39%	Thinking	76	46%

Appendix A. (*continued*)

Ranked Concept List LW			Ranked Concept List MW			Ranked Concept List HW		
Word-Like	Count	Relevance	Word-Like	Count	Relevance	Word-Like	Count	Relevance
Time	97	28%	Time	76	31%	Road	70	42%
Overtaking	77	22%	Looking	70	29%	Driving	69	42%
Wonder	74	21%	Remember	67	28%	Time	49	30%
Front	70	20%	Doing	63	26%	Start	46	28%
Lane	69	20%	Trying	61	25%	Concentrate	43	26%
Doing	66	19%	Front	56	23%	Doing	41	25%
Building	65	18%	Things	52	21%	Lane	37	22%
Things	62	18%	Happen	52	21%	Looking	35	21%
Coming	60	17%	Lane	51	21%	Things	34	21%
Behind	60	17%	Concentrate	49	20%	Overtaking	32	19%
Happen	59	17%	Overtaking	49	20%	Wondering	26	16%
Mirrors	59	17%	Wondering	45	19%	Noise	26	16%
Need	56	16%	Guess	43	18%	Coming	24	15%
Probably	51	14%	Coming	41	17%	Weird	24	15%
Trying	44	12%	Need	41	17%	Tired	23	14%
Work	43	12%	Behind	40	17%	Nice	22	13%
Nice	41	12%	Wheel	34	14%	Head	22	13%
Left	40	11%	Noise	33	14%	Having	21	13%
Long	39	11%	Thought	32	13%	Long	21	13%
Used	36	10%	Sure	32	13%	Seems	21	13%
Weird	36	10%	Probably	30	12%	Traffic	21	13%
Motorway	35	10%	People	29	12%	Attention	20	12%
Having	34	10%	Tired	28	12%	Sure	20	12%
Noise	34	10%	Nice	28	12%	Need	20	12%
Tired	33	09%	Weird	27	11%	Lost	20	12%
Different	32	09%	Control	26	11%	Steering	18	11%
People	32	09%	Having	25	10%	Guess	17	10%
Speed	31	09%	Moment	25	10%	Past	15	09%
Traffic	31	09%	Long	23	10%	Protocol	12	07%
Sure	31	09%	Seems	22	09%			
Past	30	09%						
Red	30	09%						
Total	2755	100%		2194	80%		1640	60%

Appendix B: Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.4121/uuid:6312de8d-f053-4fda-a7fe-7998c5fa70f0.

References

Atchley, P., Chan, M., & Gregersen, S. (2014). A strategically timed verbal task improves performance and neurophysiological alertness during fatiguing drives. *Human Factors: The Journal of Human Factors and Ergonomics Society, 56*, 453–462. https://doi.org/10.1177/0018720813500305.

Axelsson, J. (2017). Safety in vehicle platooning: A systematic literature review. *IEEE Transactions on Intelligent Transportation Systems, 18*, 1033–1045. https://doi.org/10.1109/TITS.2016.2598873.

Beggiato, M., & Krems, J. F. (2013). The evolution of mental model, trust and acceptance of adaptive cruise control in relation to initial information. *Transportation Research Part F: Traffic Psychology and Behaviour, 18*, 47–57. https://doi.org/10.1016/j.trf.2012.12.006.

Bergenhem, C., Pettersson, H., Coelingh, E., Englund, C., Shladover, S., & Tsugawa, S. (2012). Overview of platooning systems. In: *Proceedings of the 19th intelligent transportation systems world congress*.

Biester, L. (2008). *Cooperative automation in automobiles* (Doctoral dissertation). Berlin, Germany: Humboldt-University.

Brandenburg, S., & Skottke, E.-M. (2014). Switching from manual to automated driving and reverse: Are drivers behaving more risky after highly automated driving? In: *Proceedings of the IEEE international conference on intelligent transportation systems (ITSC)* (pp. 2978–2983). Qindao, China.

Bueno, M., Dogan, E., Hadj Selem, F., Monacelli, E., Boverie, S., & Guillaume, A. (2016). How different mental workload levels affect the take-over control after automated driving. In: *IEEE 19th international conference on intelligent transportation systems (ITSC)*. http://doi.org/10.1109/ITSC.2016.7795886.

Carsten, O., Lai, F., Barnard, Y., Jamson, A. H., & Merat, N. (2012). Control task substitution in semi-automated driving: Does it matter what aspects are automated? *Special Section: Human Factors and Automation in Vehicles, 54*, 747–761. https://doi.org/10.1177/0018720812460246.

Cha, D. (2003). Driver workload comparison among road sections of automated highway systems. SAE Technical paper (No. 2003-01-0119). http://doi.org/10.4271/2003-01-0119.

De Waard, D., Van der Hulst, M., Hoedemaeker, M., & Brookhuis, K. A. (1999). Driver behavior in an emergency situation in the Automated Highway System. *Transportation Human Factors, 1*, 67–82. https://doi.org/10.1207/sthf0101_7.

De Winter, J. C. F. (2014). Controversy in human factors constructs and the explosive use of the NASA TLX: A measurement perspective. *Cognition, Technology & Work, 16*, 289–297. https://doi.org/10.1007/s10111-014-0275-1.

Endsley, M. R. (1988). Situation Awareness Global Assessment Technique (SAGAT). In: *Proceedings of the national aerospace and electronics conference* (pp. 789–795). New York: IEEE.

Endsley, M. R. (1995). Towards a theory of situation awareness in dynamic systems. *Human Factors, 37*, 32–64. https://doi.org/10.1518/001872095779049543.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*, 215–251. https://doi.org/10.1037/0033-295X.87.3.215.

French, H. T., Clarke, E., Pomeroy, D., Seymour, M., & Clark, C. R. (2007). Psycho-physiological measures of situation awareness. In M. J. Cook, J. M. Noyes, & Y. Masakowski (Eds.), *Decision Making in Complex Environments*. Aldershot: Ashgate Publishing Limited.

Gold, G., Berisha, I., & Bengler, K. (2015). Utilization of drivetime – Performing non-driving related tasks while driving highly automated. In: *Proceedings of the human factors and ergonomics society 59th annual meeting* (Vol. 59, pp. 1666–1670). http://doi.org/10.1177/1541931215591360.

Gouy, M., Wiedemann, K., Stevens, A., Brunett, G., & Reed, N. (2014). Driving next to automated vehicle platoons: How do short time headways influence non-platoon drivers' longitudinal control? *Transportation Research Part F: Traffic Psychology and Behaviour, 27*, 264–273. https://doi.org/10.1016/j.trf.2014.03.003.

Grech, M. R., Horberry, T., & Smith, A. (2002). Human error in maritime operations: Analyses of accident reports using the Leximancer tool. In: *Proceedings of the human factors and ergonomics society 46th annual meeting* (Vol. 46, pp. 1718–1721). http://doi.org/10.1177/154193120204601906.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Advances in Psychology, 52*, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9.

Heikoop, D. D., De Winter, J. C. F., Van Arem, B., & Stanton, N. A. (2016). Psychological constructs in driving automation: A consensus model and critical comment on construct proliferation. *Theoretical Issues in Ergonomics Science, 17*, 284–303. https://doi.org/10.1080/1463922X.2015.1101507.

Heikoop, D. D., De Winter, J. C. F., Van Arem, B., & Stanton, N. A. (2017). Effects of platooning on signal-detection performance: A driving simulator study. *Applied Ergonomics, 60*, 116–127. https://doi.org/10.1016/j.apergo.2016.10.016.

Helton, W. S. (2004). Validation of a short stress state questionnaire. In: *Proceedings of the human factors and ergonomics society 48th annual meeting* (Vol. 48, pp. 1238–1242). http://doi.org/10.1177/154193120404801107.

Helton, W. S., Warm, J. S., Matthews, G. S., Corcoran, K. J., & Dember, W. N. (2002). Further tests of an abbreviated vigilance task: Effect of signal salience and jet aircraft noise on performance and stress. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 46*, 1546–1550. https://doi.org/10.1177/154193120204601704.

Hills, B. L. (1980). Vision, visibility, and perception in driving. *Perception, 9*, 183–216. https://doi.org/10.1068/p090183.

Johnson-Laird, P. N. (2001). Mental models and deduction. *TRENDS in Cognitive Science, 5*, 434–442. https://doi.org/10.1016/S1364-6613(00)01751-4.

Kazi, T. A., Stanton, N. A., Walker, G. H., & Young, M. S. (2007). Designer driving: Drivers' conceptual models and level of trust in adaptive cruise control. *International Journal of Vehicle Design, 45*, 339–360. https://doi.org/10.1504/IJVD.2007.014909.

Kunze, R., Haberstroh, M. A., Hauck, E., Ramakers, R., Henning, K., & Jeschke, S. (2011). Automated truck platoons on motorways – A contribution to the safety on roads. In S. Jeschke, I. Isenhardt, & K. Henning (Eds.), *Automation, communication and cybernetics in science and engineering 2009/2010*. Berlin, Heidelberg: Springer.

Kyriakidis, M., Happee, R., & De Winter, J. C. F. (2015). Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transportation Research Part F: Traffic Psychology and Behaviour, 32*, 127–140. https://doi.org/10.1016/j.trf.2015.04.014.

Larburu, M., Sanchez, J., & Rodriguez, D. J. (2010). Safe road trains for environment: Human factors' aspects in dual mode transport systems. In: *Proceedings of the 17th world congress on intelligent transport systems. Busan, Korea.*

Larson, J., Liang, K. Y., & Johansson, K. H. (2015). A distributed framework for coordinated heavy-duty vehicle platooning. *IEEE Transactions on Intelligent Transportation Systems, 16*, 419–429. https://doi.org/10.1109/TITS.2014.2320133.

Levitan, L., Golembiewski, G., & Bloomfield, J. R. (1998). Human Factors issues for Automated Highway Systems. *ITS Journal – Intelligent Transportation Systems Journal: Technology, Planning, and Operations, 4*, 21–47. https://doi.org/10.1080/10248079808903735.

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 road. In: *Proceedings of the 7th international driving symposium on human factors in driver assessment, training, and vehicle design* (pp. 92–98). Bolton Landing, NY.

Louw, T., Madigan, R., Carsten, O., & Merat, N. (2016). Were they in the loop during automated driving? Links between visual attention and crash potential. *Injury Prevention*. http://doi.org/10.1136/injuryprev-2016-042155.

Ma, R., & Kaber, D. B. (2005). Situation awareness and workload in driving while using adaptive cruise control and a cell phone. *International Journal of Industrial Ergonomics, 35*, 939–953. https://doi.org/10.1016/j.ergon.2005.04.002.

Matthews, G. S., Emo, A. K., & Funke, G. J. (2005). A short version of the Dundee Stress State Questionnaire. *Paper presented at the twelfth meeting of the international society for the study of individual differences, Adelaide, Australia.*

Mehlenbacher, B., Wogalter, M. S., & Laughery, K. R. (2002). On the reading of product owners' manuals: Perceptions and product complexity. In: *Proceedings of the human factors and ergonomics society 46th annual meeting* (pp. 730–734). http://doi.org/10.1177/154193120204600610.

Mehler, B., Reimer, B., & Wang, Y. (2011). A comparison of heart rate and heart rate variability indices in distinguishing single-task driving and driving under secondary cognitive workload. In: *Proceedings of the sixth international driving symposium on human factors in driver assessment, training and vehicle design* (pp. 590–597). Lake Tahoe, CA.

Neubauer, C., Matthews, G. S., & Saxby, D. (2014). Fatigue in the automated vehicle: Do games and conversation distract or energize the driver? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58*, 2053–2057. https://doi.org/10.1177/1541931214581432.

Nilsson, J. (2014). *Safe transitions to manual driving from faulty automated driving system*. (Doctoral dissertation). Gothenburg, Sweden: Chalmers University of Technology.

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.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans, 30*, 286–297. https://doi.org/10.1109/3468.844354.

Petermeijer, S. M., Cieler, S., & De Winter, J. C. F. (2017). Comparing spatially static and dynamic vibrotactile take-over requests in the driver seat. *Accident Analysis and Prevention, 99*, 218–227. https://doi.org/10.1016/j.aap.2016.12.001.

Ploeg, J., Van de Wouw, N., & Nijmeijer, H. (2014). Lp string stability of cascaded systems: Application to vehicle platooning. *Proceedings of the IEEE Transactions on Control Systems Technology, 22*, 786–793. https://doi.org/10.1109/TCST.2013.2258346.

Radlmayr, J., Gold, C., Lorenz, L., Farid, M., & Bengler, K. (2014). How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58*, 2063–2067. https://doi.org/10.1177/1541931214581434.

Ren, W., & Green, D. (1994). Continuous platooning: A new evolutionary and operating concept for automated highway systems. *Proceedings of the American Control Conference, 11*(20), 21–25. https://doi.org/10.1109/ACC.1994.751685.

Revell, K. M. A., & Stanton, N. A. (2012). Models of models: Filtering and bias rings in depiction of knowledge structures and their implications for design. *Ergonomics, 55*, 1073–1092. https://doi.org/10.1080/00140139.2012.692818.

Revell, K. M. A., & Stanton, N. A. (2016). The Quick Association Check (QuACk): A resource-light, 'bias robust' method for exploring the relationship between mental models and behaviour patterns with home heating systems. *Theoretical Issues in Ergonomics Science, 17*, 554–587. https://doi.org/10.1080/1463922X.2016.1180439.

Rogers, M., Zhang, Y., Kaber, D., & Liang, Y. (2011). The effects of visual and cognitive distraction on driver situation awareness. In: D. Harris (Ed.), *Engineering psychology and cognitive ergonomics. EPCE 2011. Lecture notes in computer science* (pp. 6781). Berlin, Heidelberg: Springer.

Salmon, P. M., Lenne, M. G., Walker, G. H., Stanton, N. A., & Filtness, A. (2014). Exploring schema-driven differences in situation awareness between road users: An on-road study of driver, cyclist and motorcyclist situation awareness. *Ergonomics, 57*, 191–209. https://doi.org/10.1080/00140139.2013.867077.

Sarter, N. B., & Woods, D. (1991). Situation awareness: A critical but ill-defined phenomenon. *The International Journal of Aviation Psychology, 1*, 45–57. https://doi.org/10.1207/s15327108ijap0101_4.

Saxby, D. J., Matthews, G. S., & Neubauer, C. (2017). The relationship between cell phone use and management of driver fatigue: It's complicated. *Journal of Safety Research, 61*, 129–140. https://doi.org/10.1016/j.jsr.2017.02.016.

Shaffer, F., & Combatalade, D. C. (2013). Don't add or miss a beat: A guide to cleaner Heart Rate Variability recordings. *Biofeedback, 41*, 121–130. https://doi.org/10.5298/1081-5937-41.3.04.

Skottke, E.-M., Debus, G., Wang, L., & Huestegge, L. (2014). Carryover effects of highly automated convoy driving on subsequent manual driving performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 56*, 1272–1283. https://doi.org/10.1177/0018720814524594.

Smith, A. E. (2003). Automatic extraction of semantic networks from text using Leximancer. In: *Proceedings of the HLT-NAACL 2003 demonstrations* (pp. 23–24). Edmonton.

Verwey, W. B., & Zaidel, D. M. (1999). Preventing drowsiness accidents by an alertness maintenance device. *Accident Analysis and Prevention, 31*, 199–211. https://doi.org/10.1016/S0001-4575(98)00062-1.

Vollmer, M. (2015). A robust, simple and reliable measure of heart rate variability using relative RR intervals. *IEEE Conference on Computing in Cardiology Conference (CinC)*, 609–612. https://doi.org/10.1109/CIC.2015.7410984.

Young, M. S., & Stanton, N. A. (2002). Malleable attentional resources theory: A new explanation for the effects of mental underload on performance. *Human Factors: The Journal of Human Factors and Ergonomics Society, 44*, 365–375. https://doi.org/10.1518/0018720024497709.

Young, M. S., & Stanton, N. A. (2007). What's skill got to do with it? Vehicle automation and driver mental workload. *Ergonomics, 50*, 1324–1339. https://doi.org/10.1080/00140130701318855.