Van der Heiden, R.M.A., Kenemans, J.L., Donker, S.F., Janssen, C.P. (in press 2021) The effect of cognitive load on auditory susceptibility during automated driving. *Human Factors* DOI: 10.1177/0018720821998850

14

The effect of cognitive load on auditory susceptibility during automated driving

Remo M.A. van der Heiden*; J. Leon Kenemans; Stella F. Donker; Christian P. Janssen

* Corresponding author:
Utrecht University
Department of Experimental Psychology & Helmholtz Institute
Utrecht, The Netherlands
r.m.a.vanderheiden@uu.nl

Manuscript type: Research Article

Word count: 6478

Acknowledgements
Remo van der Heiden was supported by the Dutch Traffic Authority (Rijkswaterstaat). Christian Janssen was supported by a Marie Sklodowska-Curie fellowship of the European Commission (H2020-MSCA-IF-2015, grant agreement 705010, Detect and React). The funders had no role in study design, data collection, analysis, decision to publish, or manuscript preparation. We would like to thank Nina Haukes for her assistance with the data collection. Preliminary results of this work were presented at the Auto-UI 2019 conference during the work-in-progress track (Janssen, van der Heiden, Donker, Kenemans, 2019).

Keywords: Distractions and interruptions; Dual task; Mental workload; Autonomous driving; Cognitive neuroscience.

Abstract

Objective: We experimentally test the effect of cognitive load on auditory susceptibility during automated driving.

Background: In automated vehicles, auditory alerts are frequently used to request human intervention. To ensure safe operation human drivers need to be susceptible to auditory information. Previous work found reduced susceptibility during manual driving and in a lesser amount during automated driving. However, in practice, drivers also perform non-driving tasks during automated driving, of which the associated cognitive load may further reduce susceptibility to auditory information. We therefore study the effect of cognitive load during automated driving on auditory susceptibility.

Method: 24 participants were driven in a simulated automated car. Concurrently, they performed a task with two levels of cognitive load: *repeat* a noun or *generate* a verb that expresses the use of this noun. Every noun was followed by a probe stimulus to elicit a neurophysiological response: the frontal P3, which is a known indicator for the level of auditory susceptibility.

Results: The frontal P3 was significantly lower during automated driving with cognitive load compared to without. The difficulty level of the cognitive task (repeat or generate) showed no effect.

Conclusion: Engaging in other tasks during automated driving decreases auditory susceptibility as indicated by a reduced frontal P3.

Application: Non-driving task can create additional cognitive load. Our study shows that performing such tasks during automated driving reduces the susceptibility for auditory alerts. This can inform designers of semi-automated vehicles (SAE levels 3 and 4), where human intervention might be needed.

Précis: Being susceptible to auditory information is important for safe operation of (semi-)automated vehicles. Using EEG measurements in a driving simulator experiment, we test the effect of cognitive load on auditory susceptibility. We show that engaging in other tasks during automated driving decreases auditory susceptibility of the brain.

Introduction

Automation in everyday life is rapidly increasing. Although automation can take away tasks from the human, there are many forms of automation that involve both the human and the system (e.g., Dekker & Woods, 2002; Parasuraman & Riley, 1997; Parasuraman, Sheridan, & Wickens, 2000; Sheridan & Verplank, 1978). Such shared control systems require the human operator to be informed of the system state. In the past these tasks were typically left to skilled, well-trained, professional users such as airplane pilots and control room monitors. However, today more and more automation finds its way to consumer products which are operated by non-professional users who lack extensive training (Janssen, Donker, Brumby, & Kun, 2019). Therefore, intuitive design of these systems becomes even more important.

The domain of automated driving is one of the fields that has seen an increasing amount of automation. The Society of Automotive Engineers distinguishes six levels of automation in vehicles (SAE International, 2018). These levels differ in tasks that are performed by the driver (human) and tasks that are performed by the vehicle (machine). At SAE levels 3 and 4, the automated vehicle is expected to be able to drive for prolonged time without human intervention (within specific operational design domains). However, at times the human might be required (SAE level 3) or requested without obligation (SAE level 4) to assist the automation. Although the way in which the car alerts the driver about this assistance can vary between systems, a likely candidate are auditory signals, as these are omnidirectional, already widely applied in cars, and have relatively fast response time across multiple studies of SAE level 2 cars (Zhang, De Winter, Varott, Happee, & Martens, 2019).

As humans are expected to continue to play a role in many forms of (semi-) automated driving (Noy, Shinar, & Horrey, 2018), it is important to understand how well the human brain processes auditory alerts in general. Is this general ability for example reduced under automated driving conditions? And how is this general ability to process auditory alerts impacted when someone is performing a non-driving task while the automated vehicle is driving without human intervention? We investigate those questions in this paper using a technique from neuroscience, which is described next.

Frontal P3 (fP3) as a measure of susceptibility

In this manuscript, we refer to the brain's general ability to process alerts as *susceptibility*. The online Oxford advanced learner's dictionary (2020) defines susceptibility as: *"the state of being very likely to be influenced, harmed or affected by something"*. Our definition is consistent with this broad definition, but more specific:

susceptibility refers to the extent to which an observer is in a mode that allows for detection of external signals to such a degree that an adequate behavioral response can be based on the detection (cf. Kenemans, 2015).

To assess auditory susceptibility, we use the auditory novelty oddball paradigm (for a review see Polich, 2007), consisting of a stream of at least identical standard tones, mixed with (semi-) unique novels. Concurrent brain activity recording (EEG ERP: Electroencephalogram Event-Related Potential) can then be used to quantify the novel-probe-elicited cortical activation (corrected for the standard-elicited activation). The most prominent feature of this novelty-oddball response is the so-called frontal P3 (fP3) response in the ERP: a positive peak over frontal regions (e.g., electrode FCz) around 300 ms after stimulus onset (Allison & Polich, 2008; Squires, Squires, & Hillyard, 1975; Ullsperger, Freude, & Erdmann, 2001), indicating an increase in susceptibility to the stimulus.

The fP3 is a relatively generic response (Friedman, Cycowicz, & Gaeta, 2001; Polich, 2007; Kenemans, 2015; Wessel and Aron, 2013), elicited by any sufficiently salient event. In the current study these are auditory novels, but the salient event can also be visual, emotionally laden, or occasional auditory or visual countermanding signals (see Kenemans, 2015 for examples). In relation to our aforementioned definition of susceptibility, note also that the fP3 as an evolving process has been associated with a direct consequence for behavior, in the sense of behavioral interrupt, or a transient general slowing of the motor system (Kenemans, 2015).

The fP3 has therefore been widely used to index susceptibility in a variety of conditions and tasks, including driving (Van der Heiden et al., 2018; Wester, Böcker, Volkerts, Verster, & Kenemans, 2008), mental fatigue during driving (Massar et al., 2010), manual tracking (Scheer, Bülthoff, & Chuang, 2016, 2018), games (e.g., Allison & Polich, 2008; Miller, Rietschel, McDonald, & Hatfield, 2011), arithmetic (e.g., Ullsperger, Freude, & Erdmann, 2001), and during cognitive tasks without visual or manual components (Van der Heiden, Janssen, Donker, & Kenemans, 2020). Susceptibility can also be reduced in other ways that are not tied to a task, such as alcohol (Wester et al., 2010) and passive fatigue (Massar et al., 2010). In other words, the fP3 response is a probe to the more general susceptibility of the brain to external signals. We therefore prefer susceptibility over other, closely related, terms such as inattentional deafness (which is tied to auditory stimuli; e.g., Scheer, Bülthoff, & Chuang, 2018) or attentional reorienting (Corbetta, Patel, & Shulman, 2008; Corbetta & Shulman, 2002; Schröger & Wolff, 1998) and workload (for a review see Murphy, Spence, & Dalton, 2017) (which are tied to even more specific mechanisms). Other perspectives have focused more on potential predictors of reduced susceptibility, such as the EEG alpha-rhythm power (O'Connell, Dockree, Robertson, Bellgrove, Foxe, & Kelly, 2009), known to greatly increase across hours of monotonous driving (e.g., Schmidt et al., 2009).

For the domain of driving, previous work found a reduction in fP3 response (i.e., indicating a reduction in susceptibility to novel stimuli) under driving and automated driving conditions (Van der Heiden et al., 2018; Wester et al., 2008) when compared to a stationary (non-driving) baseline. It has not been explored how performing additional tasks during automated driving (e.g., a telephone call) affects auditory susceptibility. In-vehicle non-driving tasks can take many forms and their variety is expected to increase with higher levels of automation (e.g., Banks et al., 2018; Carsten et al., 2012; Llaneras et al., 2013; Pfleging et al., 2016). To be able to measure the effects of performing additional tasks during (automated) driving on auditory susceptibility we need to induce cognitive load in a systematic way.

To this end, we use the verb task (Abdullaev & Posner, 1998; Petersen, Fox, Posner, Mintun, & Raichle, 1989; Snyder, Abdullaev, Posner, & Raichle, 1995). In this task, participants hear nouns, and either need to *repeat* the noun (e.g., apple - apple), or *generate* a verb that is related to the noun (e.g., apple - eat). The generate task is known to induce cognitive load (Abdullaev & Posner, 1998; Snyder et al., 1995), which can interfere with dual-task performance (cf. Iqbal, Ju, & Horvitz, 2010; Kunar, Carter, Cohen, & Horowitz, 2008; Strayer & Johnston, 2001; Van der Heiden et al., 2019), and increase activity in the frontal cortex when compared with the easier repeat task (Abdullaev & Posner, 1998; Bijl et al., 2007). Furthermore, the generate task reduces auditory susceptibility in non-driving conditions (Van der Heiden et al., 2020). This makes it a good candidate to assess how susceptibility changes when automated driving is combined with another (cognitive load inducing) task, which was the aim of the current work.

We included both a *generate* and a *repeat* condition to obtain better insight in the mechanism by which the additional task (on top of automated driving) reduces susceptibility: Is it the mere production of a vocal response, or more specifically active search within the semantic network (only in generate)?

Study aim and hypotheses

We test how induced additional cognitive load influences general susceptibility to auditory stimuli while people are driven by an automated vehicle. We hypothesize that fP3 is reduced (i.e., indicating a reduced susceptibility to auditory stimuli) when:

1. cognitive load is added during automated driving (using either the repeat or the generate task) compared to stationary and automated driving without additional tasks (cf. Abdullaev & Posner, 1998; Snyder et al., 1995).
2. automated driving is combined with generating a verb compared to automated driving while repeating a noun, as the generating task is hypothesized to create more cognitive load (due to active search within the semantic network; Abdullaev & Posner, 1998; Snyder et al., 1995).
3. driving in automated conditions compared to stationary (cf. Van der Heiden et al., 2018).

Method

Participants

We conducted a power-test in G*power 3.1.9.4. With effect size (d) 0.71 (difference stationary and automated in Van der Heiden et al., 2018), alpha-level of .0125 (the level used in pairwise comparison), and power of 0.8, we required at least 22 participants.

24 participants (21 F; 3 M) were recruited through on-campus flyers, word of mouth, and advertisement on the participant pool website of the university. Participants were 23 years old on average (ages 18 to 55, *SD* = 7.2 years of age). All participants indicated to have normal or corrected to normal vision. All participants were novel to the experiment and did not participate in similar experiments. Participants had a driver’s license for 4.3 years on average (SD = 5.9 years; one participant had no driver’s license, range for others was 0.5-30 years).

This research complied with the tenets of the Declaration of Helsinki and was approved by the Institutional Review Board at Faculty of Social and Behavioral Sciences of Utrecht University (FETC16-042). Informed consent was obtained from each participant. Participants were compensated with either €12 or course credits for their time.

Materials

Driving simulator

A medium fidelity fixed base driving simulator, based on an original Green Dino three screen setup, was used. The setup (see Figure 1) included three 40-inch screens and surround sound. OpenDS 4.5 (www.opends.eu) was used as simulator software. The driving environment consisted of a three-lane highway that followed the trajectory of two semi circles, with a radius of 1135.9 m (one clockwise, one counterclockwise). The automated car drove in the middle lane of the highway at 80 km/h. There were no other cars in the driver's lane, but cars occasionally drove in the other lanes (left 87 km/h and right 73 km/h).

A direct matching to SAE levels is not representative due to the relatively simple driving scenario (with e.g. no sudden events) and such a comparison was also not provided to participants. Our scenario is closest to SAE level 4 (SAE International, 2018), in that the driver was not asked for any driving related action (i.e., there were no transitions of control). However, unlike the requirements in SAE level 4, our participants were instructed to sit still and look at the road. Therefore, our results should not be tied to specific SAE levels (as that would require further testing), but rather as an indication of human general susceptibility to sounds during prolonged periods where a driver is being driven by a car and is performing other tasks (in our case: generating verbs or repeating nouns). A driving simulator was used as previous results with fP3 ERP studies in simulated manual driving seem to replicate well in on the road driving (Wester, 2009). In the stationary condition the car stayed stationary at the start location with the engine idle. The other cars, however, still occasionally drove in the other lanes.

Figure 1. Driving simulator setup with participant wearing 64 electrode EEG cap.

Presentation of auditory stimuli

Two types of auditory stimuli were used in this experiment: oddball probe stimuli and verb task stimuli. All stimuli were presented using Presentation (Neurobehavioral Systems) at 75 dB trough Earlink earphones.

Oddball probe

We used a two-stimulus novelty oddball probe (Van der Heiden et al., 2020). In 75% of cases, the stimuli consisted of a standard sound: a 1000 Hz pure tone of 400 ms. In 25% of cases, the stimuli consisted of novel sounds: environmental sounds such as a dog barking or a human sneezing, that were taken from a database by Fabiani and Friedman (1995). The database consisted of 100 unique sounds that were between 159 ms and 399 ms in duration.

Verb generation and noun repetition task stimuli

Nouns were presented for a verb “generate” task (responding to a noun by saying a related verb) or a noun “repeat” task (repeating the noun), see design. Previous work suggests that the generate task (compared to the repeat task) induces more cognitive load (Abdullaev & Posner, 1998; Snyder et al., 1995), stronger dual-task interference (cf. Iqbal, Ju, & Horvitz, 2010; Kunar, Carter, Cohen, & Horowitz, 2008; Strayer & Johnston, 2001; Van der Heiden et al., 2019), and increased activity in the frontal cortex (Abdullaev & Posner, 1998; Bijl et al., 2007). As our aim is to study how the fP3 response changes under automated driving as a function of additional load, we included both a noun repetition and a verb generation version of the task.

For the materials, a set of 96 spoken nouns was used in the verb generation and noun repetition task. In the *verb generate* task (Abdullaev and Posner, 1998), participants were instructed to generate a verb that fitted with the noun they heard. For example: hammer → pound. In the *noun repeat* task, participants were instructed to repeat the exact noun they heard (i.e., hammer → hammer).

Since our participants were Dutch, we used a Dutch translation by Van der Heiden et al. (2020) of spoken nouns based on an English set used by Abdullaev and Posner (1998). For the current study, we only used 96 nouns of the 144 words used by Van der Heiden et al. (2020), as each block had 32 words (see design), so the total number of words had to be a multiple. The selected 96 words had the fewest errors on trials where participants had to repeat the words in Van der Heiden et al. (2020).

As described in more detail in Van der Heiden et al. (2020), word selection focused on using words that are familiar to Dutch speakers, and which could be presented in a short time interval. Only Dutch words that had one or two syllables were used. Per word, a WAV sound file was generated using text-to-speech website www.texttospeech.io with default settings of the text-to-speech algorithm (Dutch female, volume 1, rate 1, pitch 1). Nouns of which presentation took longer than 500 ms were removed. For the remaining words, the tempo was adjusted per word, such that each noun had a playback time of exactly 400 ms.

Design

To assess the effect of cognitive load that is added on top of an automated driving condition we used a single factor within-subjects design with 4 levels: Stationary, Automated, Automated + repeat, and Automated + generate. This allowed us to assess the effect of cognitive load as it comes on top of that of automated driving relative to stationary. Within each block, participants heard both standard tones and novel sounds. The fP3 response is calculated as a difference wave in the event-related potential between standards and novels (see section on signal recording).

Testing blocks

There were 12 experimental blocks, each about 3 minutes long. Each experimental condition (e.g., Stationary, Automated, Automated + repeat, and Automated + generate) was used in 3 blocks. Per set of 4 blocks, all conditions were used. Within that set, the order was varied between participants. For the first four blocks, the order was counterbalanced across participants. For the remaining two sets of four blocks, orders were shuffled such that participants were offered with different orders than before. For example, the first set that participant 1 experienced was: automated without extra task (A), automated + generate (AG), stationary (S), automated + repeat (AR). Subsequently, the order of the second and third block were respectively S, AG, A, AR and S, AR, AG, A.

Within each experimental block, 80 oddball probes were presented. In blocks where automation was combined with verb generation (AG) or noun repetition (AR), there were three types of stimuli: nouns (for the generate or repeat task; each stimulus exactly 400 ms), standards, and novels. To test the effect that the cognitive process associated with verb generation (AG) or noun repetition (AR) had on fP3 response, we carefully balanced when these stimuli were presented in the AG and AR blocks. Specifically, per block, 16 nouns were played immediately preceding a standard oddball probe, 16 immediately preceding a novel oddball, and 48 standards were played without a prior noun presentation. If a probe followed a noun presentation, the next probe was presented 4400 ms after the onset of the preceding oddball stimulus to prevent interference from speech production. On all other trials (where no noun was played, including trials of the S and A blocks), the interval between the onset of two probe stimuli was 2000 ms (cf. Van der Heiden et al., 2018; Wester et al., 2008).

For the word task, 96 different nouns were used. To vary these between blocks, we made six sets of 32 nouns, three sets for the generate task (containing all 96 unique words, shuffled), and three for the repeat task (again with all 96 words). The order of words within a set was randomized for each participant. In effect, each word was used twice per participant: once in the generate task, and once in the repeat task.

Procedure

Participants received verbal and written information about the experiment and then provided written consent. Next, for the intelligibility test, all nouns were played to the participant, who was tasked to repeat each noun after playback. To validate that all nouns were intelligible, the experimenter in the meantime made notes of nouns that were incorrectly replied to.

The experimenter then applied the EEG electrodes. Participants were then told that they should not hold the steering wheel because the car would drive on its own and manual input would not be needed. A practice block was started where participants performed the verb generation task for 1 minute, while they were also driven by the automated vehicle and the oddball probes were used. The participant then performed the 12 experimental blocks, with a few minutes rest after every four blocks. After the experiment, participants were asked to fill out a questionnaire on demographics and general feedback. The total experiment lasted just under two hours.

Signal recording

EEG setup

EEG was recorded using a BioSemi ActiveTwo system with 64 active Ag-AgCl electrodes positioned following the international 10/10 system (Sharabrough, 1991), and the standard BioSemi CMS/DRL on-line reference, at a sample rate of 2048 Hz. Two electrodes were placed on mastoids, for later re-referencing to average mastoids. Four ocular electrodes were applied to enable offline ocular-artifact control with horizontal and vertical electrooculography (HEOG and VEOG). After measuring the head circumference, a matching EEG cap was applied. Conductive gel was applied and the corresponding electrodes were plugged in.

Signal analysis was done in BrainVision Analyzer 2.1 (Brain Products GmbH, München, Germany), following similar procedures as in earlier work (Van der Heiden et al., 2018; Van der Heiden et al., 2020; Wester et al., 2008). We first down-sampled the data to 256 Hz (after anti-alias filter). Data were then re-referenced to average mastoids signal. A high-pass filter of 0.16 Hz, a low-pass filter of 30 Hz, and a notch filter of 50 Hz were applied. We then created segments for each of the four conditions for both standard and novel probes starting 1000 ms before and ending 1500 ms after oddball probe onset. Before calculating the ERPs, we applied the Gratton & Coles ocular correction to compensate for eye movement during the recorded segments (Gratton, Coles, & Donchin, 1983). Artifacts in individual channels were rejected by the following criteria in an epoch: maximum voltage step > 120 μV/ms within 200 ms before or after events; maximum difference > 100 μV within 200 ms; minimum activity < 0.5 μV within 100 ms. Finally, grand averages were created for each of the conditions. Our analysis focuses on a *difference wave*, which was obtained by subtracting the ERP in response to standard tones from the ERP in response to novel sounds.

To determine the time interval at which the fP3 peak occurred at electrode location FCz, we used a collapsed localizer. The interval 285-335 ms after stimulus onset was found to best represent the fP3 peak area when the ERPs for all four conditions were collapsed. We took the average value in the fP3 interval for statistical peak analysis.

Speech response time

To check our cognitive-load inducing task manipulation, we measured speech response time. Based on earlier literature, we would expect that response times are faster when participants merely repeat a noun, compared to when they need to generate a verb (e.g., Iqbal, Ju, & Horvitz, 2010; Van der Heiden et al., 2019). However, we would expect that there is no difference whether a noun was preceded by a standard tone or a novel sound. We used a microphone, connected to the auxiliary input of the BioSemi. We used an average level (i.e., calculated using a moving average) of 1000 μV over 15 samples as threshold for speech production. As speech response time we took the interval starting at noun offset (oddball probe onset) and ending at the start of speech production. We excluded the first four participants from this analysis as no microphone was present during that time. We did not record the content of what participants said.

Statistical analysis

For statistical analysis, we use R statistics (R Core Team, 2014), with an alpha level of .05. Partial eta-squared is used for effect sizes. For fP3 results we analyze the difference wave (novel-standard, expressed in μV) using a one-way (omnibus) ANOVA with four levels: stationary, automated, automated + repeat, and automated + generate. For pairwise comparisons, we used planned contrasts with four levels, to compare effects in the order that was expected, namely that extra tasks increase load and reduce fP3. Specifically, whether: (1) automated was lower than stationary, (2) automated + repeat was lower than automated, (3) automated + generate was lower than automated, and (4) automated + generate was lower than automated + repeat. To control for the family-wise error, our criterion for calling a difference significant was alpha / 4 (i.e., .05 / 4 = .0125).

For speech-response time (expressed in ms) we use a 2 (Oddball probe: Standard or Novel) x 2 (Cognitive load inducing task: repeat or generate) ANOVA.

Results

frontal P3

For each of the four conditions (i.e., Stationary, Automated, Automated + repeat, and Automated + generate), we calculated the difference wave of fP3 ERP at electrode FCz (i.e., difference between response to the novel probe and standard probe). Figure 2 shows the fP3 peak, the area of which the mean value was used for statistical analysis is indicated with dashed lines. There was a main effect of condition on the mean fP3 peak activation, *F*(3,69) = 16.1, *p* < .001, η_p^2= 0.58. Subsequently, we performed four pairwise comparisons to test which conditions differed from each other. Tests were done in order for the conditions where we predicted the highest fP3 value (stationary) to where we expected the smallest fP3 value (automated with generate). fP3 was highest during single task stationary (*M* = 11.5 µV, *SD* = 6.1 µV). Pairwise comparisons between condition revealed that stationary did not differ significantly from single task automated (*M* = 9.9 µV, *SD* = 4.4 µV, *p* = .049). fP3 in the Automated + repeat condition (*M* = 5.1 µV, *SD* = 4.9 µV) was significantly lower than Automated (*p* < .001). Automated + repeat did not differ significantly from Automated + generate (*M* = 5.6 µV, *SD* = 2.9 µV, *p* = .57). Automated + generate did also differ significantly from Automated (*p* < .001). That is, our results suggest that performing a concurrent task under automated driving conditions reduces fP3 response and associated auditory susceptibility. Figure 3 shows for various time intervals how electrical activity is distributed across the scalp as a difference between the response to the novel compared to the standard. The figure illustrates that the fP3 response is indeed the highest in the frontal area of the brain, near electrode FCz that we analyzed. Moreover, it shows how cognitive load influences this activity.

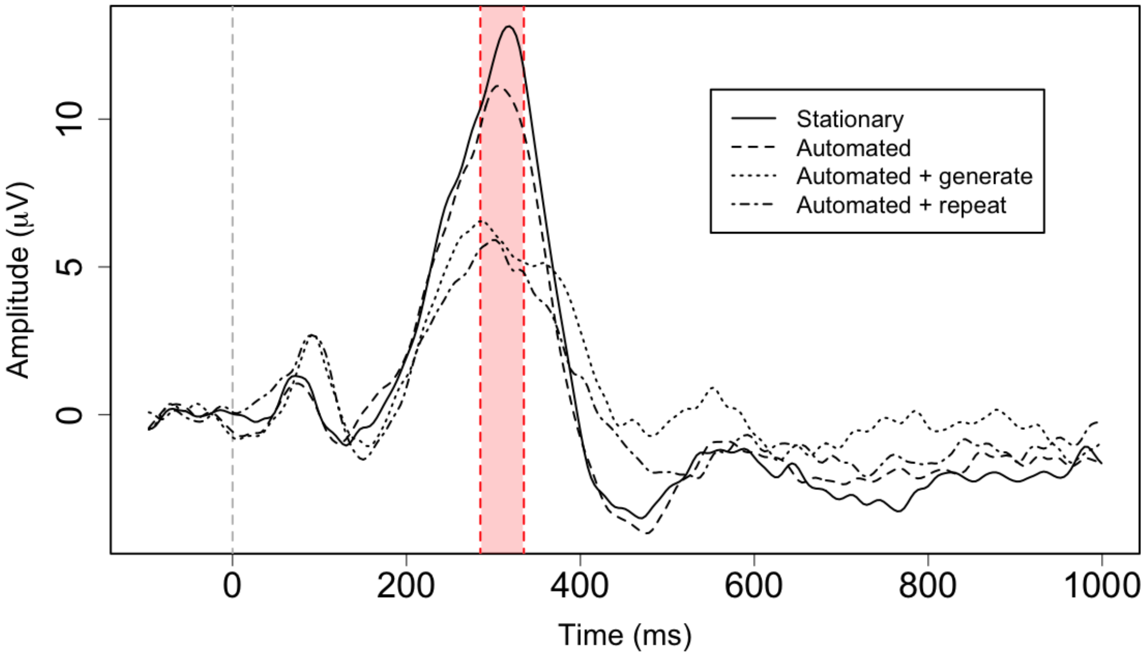

Figure 2. Event related potential of the four conditions (Stationary, Automated, Automated + generate, Automated + repeat). Vertical lines show onset of oddball stimulus (time point 0 ms), noun stimulus (onset at -400 ms in gray), and fP3 peak area used for statistical analysis (285-335 ms).

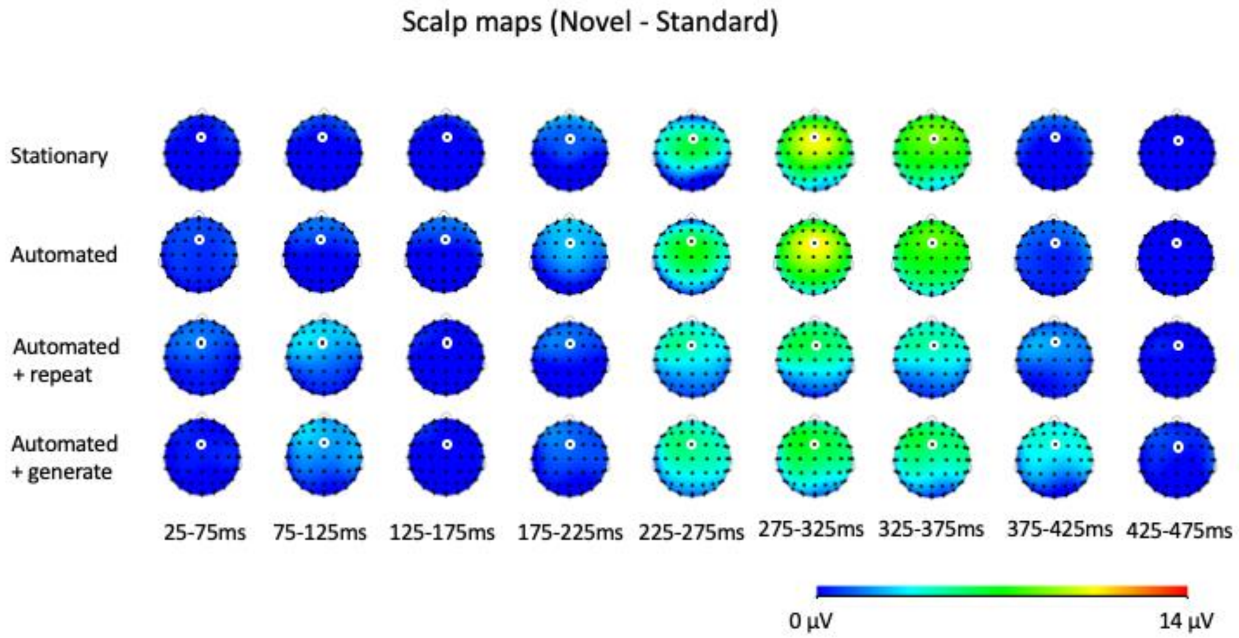

Figure 3. Scalp maps for various 50 ms time intervals from 25 ms after oddball probe onset to 475 ms after oddball probe onset. Average mastoid is used as reference value.

Speech response time

Figure 4 shows the average speech activation level for the different conditions over time, as measured from the point of noun offset and oddball probe onset. As the green line shows that there is no consistent background noise, we dropped all word absent trials for statistical analysis.

A 2 (Oddball probe: Standard or Novel) x 2 (Cognitive load inducing task: repeat or generate) ANOVA showed that there was no main effect of oddball probe $F(1,19) = 3.24$, $p = .09$, $\eta_p^2 = 0.15$. There was a main effect of cognitive load inducing task $F(1,19) = 174.1$, $p < .001$, $\eta_p^2 = 0.90$. Speech response time was higher under the generate condition (*Mdn* = 680 ms) compared to the verb generation time (*Mdn* = 287 ms). There was no significant interaction effect, $F(1,19) = 0.13$, $p = .72$, $\eta_p^2 = 0.007$

In other words, our manipulation of cognitive load succeeded: responses take longer in the generate condition compared to the repeat condition (cf. Iqbal et al., 2010; Van der Heiden et al., 2019). There was no effect of the type of oddball stimulus (standard or novel).

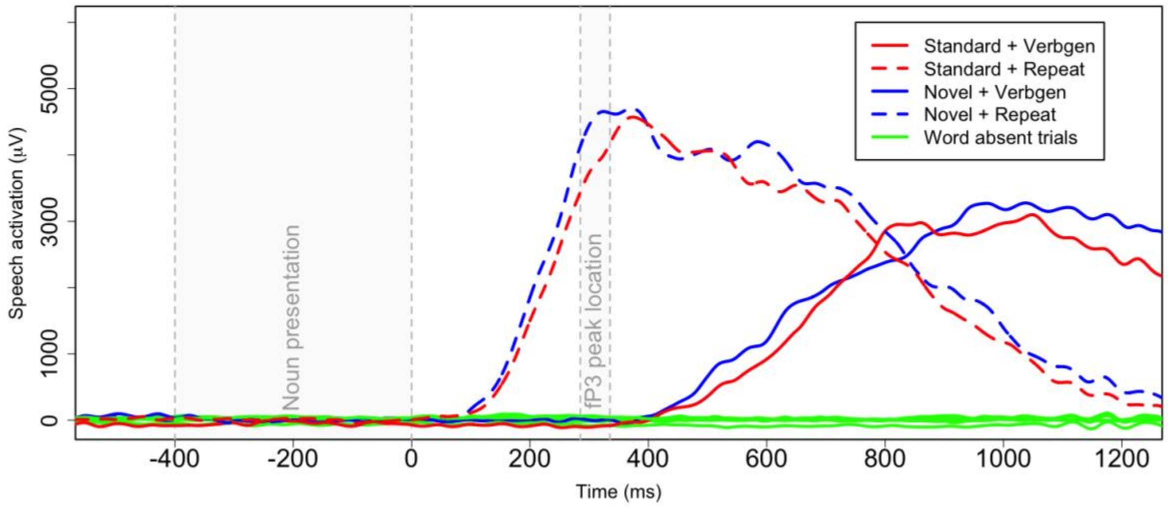

Figure 4. Average speech activation level for different conditions, no speech activation is expected when word presentation is absent. Dashed lines show activation for the repeat condition, solid lines show activation for verb generation condition. Red lines show task combined with a standard tone, blue lines show task combined with a novel sound. Note that time point 0 corresponds to noun offset and probe onset. The grey areas indicated when in the trial a noun was presented, and when fP3 peak activation was analyzed in the ERP data (Figure 2).

Comparison to manual driving and single-task verb generation

This study found that the fP3 peak is reduced when a cognitive load inducing task is performed during automated driving conditions. For a wider context, we compared our results to those from two previous studies in our lab that were run by the same team, with the same EEG set-up and comparable stimuli (Van der Heiden et al., 2018; 2020). Figure 5 shows bar diagrams of the average fP3 amplitude of the novel-standard difference wave as observed in this study and as observed in previous studies.

Brief description of previous studies' methodology

Van der Heiden et al. (2018) manipulated within-subjects whether participants were in a stationary control (watching a screenshot of a road), being driven by an automated vehicle, or driving manually. The driving task was performed in a low-fidelity simulator (Logitech steering wheel and pedals, 1 screen), the scenario was a trajectory that looped between driving on a regular road, merging onto a highway with other traffic, and unmerging back to the regular road. For the oddball stimuli, the 2018 study used a three-stimulus novelty oddball paradigm, containing standard tones (80% of stimuli; same stimuli as here), novel sounds (10% of stimuli; same stimuli as here), and deviant tones (10% of stimuli; 1100 Hz tones). Apart from the driving manipulations, between subjects the authors manipulated whether participants had to press a button when hearing a deviant tone (active condition), or not (passive condition).

Van der Heiden et al. (2020) presented frequent oddball stimuli using a 2-stimulus oddball experiment (without deviant; as done here), where 80% of oddball stimuli were standards, and 20% were novels (same stimuli as here). Within each block, some oddball stimuli were not preceded by a noun (baseline control), other oddball stimuli were preceded by a noun with an offset of 0 ms, 200 ms, or 400 ms. Participants always had to respond to a noun by generating a verb. In the 2020 study, no repeat condition was used, and no driving condition was used.

Comparison of results

In all three studies (Van der Heiden et al., 2018; 2020; current study), the fP3 response (and associated susceptibility to novel stimuli) is highest in the baseline conditions (in Van der Heiden et al., 2018: stationary), with amplitude values around 10-12 μV. The exception is the passive condition of Van der Heiden et al. (2018), which has a slightly lower peak value (main effect of active/passive).

In both Van der Heiden et al. (2018) and the current study, the condition where there is automated driving without another task lowers the mean fP3, which was significant in the 2018 study but not here (here: *p*-value of .049, with alpha at .0125). Interestingly, manual driving (Van der Heiden et al., 2018) and generating verbs without another task (Van der Heiden et al., 2020: 0, 200, and 400 ms conditions) both strongly reduce the fP3 amplitude.

In other words, it seems like a floor effect occurs in three situations: manual driving (Van der Heiden et al., 2018), generating verbs (Van der Heiden et al., 2020), or combining automated driving with repeating or generating (current study). Another perspective is that the introduction of any concurrent task, irrespective of difficulty and the specific processing demands (either manual driving, repeating words, or generating words), induces costs of such concurrence (Kok, 2001).

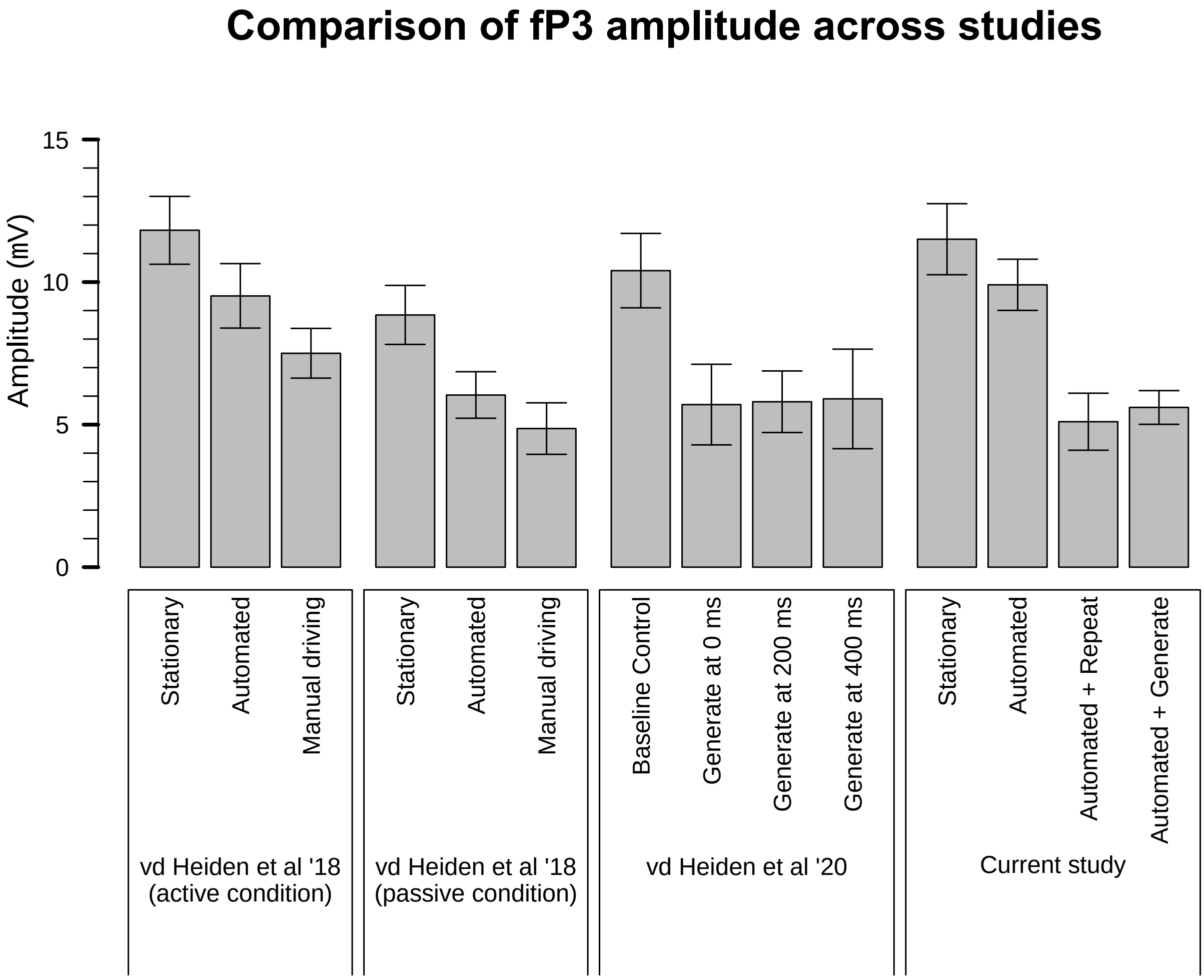

Figure 5. Comparison of amplitudes of fP3 response between three studies: Van der Heiden et al. (2018), Van der Heiden et al. (2020) and the current study. See text for details.

General Discussion

This study found that the fP3 peak is reduced when drivers are performing an additional (cognitive load inducing) task under automated driving conditions. Previous research on the verb task suggests that the generate condition should lead to more cognitive load compared to the repeat condition (cf. Iqbal, Ju, & Horvitz, 2010; Kunar et al., 2008; Strayer & Johnston, 2001; Van der Heiden et al., 2019). We therefore expected that possibly fP3 response would be lower in the generate (while automated driving) condition compared to the repeat (while automated driving) condition. In contrast to our expectations and previous research, our study did not find a difference between the generate and repeat conditions on fP3 peak. This is unlikely to reflect cognitive load induced by response production; whereas this could hold for repeat, overt responses and therefore preparatory response production processes were much later in generate, and very probably too late to affect the production of the fP3. Rather, the lack of differential fP3 could reflect equal cognitive load in repeat and generate, but induced by response production in the former and by semantic search (preceding response production) in the latter.

For the difference between stationary (no task) and automated driving (without additional load inducing task), the pattern was in the expected direction where fP3 response is highest in the stationary condition (cf. Van der Heiden et al., 2018). However, we did not statistically replicate the finding that automated driving by itself (i.e. without the addition of a secondary task) causes lower auditory susceptibility, as indicated by a decrease in the fP3 peak, compared to being stationary (Van der Heiden et al., 2018). It is conceivable that this difference was less clear in the current study because the context of the verb-generation task induces a general relevance of all auditory stimulation. In a similar vein, the reduction of fP3 when driving compared to when stationary has been reported to disappear when the sequence of probes contains additional stimuli that have to be responded to behaviorally (Wester et al., 2008; Van der Heiden et al., 2018 active condition – see also Figure 5).

In the present study we did discover that performing an additional cognitive task during automated driving reduces susceptibility. This is a relevant finding, given people’s tendency to perform other non-driving tasks in semi-automated driving settings (e.g., Banks et al., 2018; Carsten et al., 2012; Dunn, Dingus, Soccolich, 2019; Llaneras et al., 2013), and the likelihood that auditory signals will be part of alerts in (semi-) automated vehicles to require (SAE level 3) or request (SAE level 4) human assistance. Another way of interpreting these results (cf. Figure 5), is that replacing a human task (e.g., driving) through automation frees cognitive resources of the human that allow for higher susceptibility to unexpected resources (i.e., fP3 is higher in automated compared to manual driving conditions). However, in practice drivers might perform additional tasks (e.g., out of boredom; Dunn et al., 2019). In an irony of automation (Bainbridge, 1983), our results suggest that automating a task could then (through drivers’ engagement in additional tasks) decrease (instead of increase) human susceptibility.

An alternative view is inspired by our analysis of speech data, which revealed a median voice-onset latency of 287 ms during repeat, relative to probe onset (see Figure 4). This indicates that a considerable amount of voice response was produced while information was still being sampled from the probe stimulus, or immediately after that. This may have induced a form of (backward) masking that reduced the difference between novel- and standard fP3, perhaps to an extent comparable to that in the generate condition (in which median voice-onset latencies were much later, i.e., 680 ms). Further work is needed to see if, and how strongly, the repeat and generate conditions can be differentiated. Or, more generally, how different levels of cognitive load affect fP3 response and associated susceptibility under automated driving conditions.

Our comparison of fP3 magnitude with those observed in previous studies (see Figure 5) suggests a floor effect in fP3 response in three situations: manual driving (Van der Heiden et al., 2018; see also Wester et al., 2008), generating verbs (Van der Heiden et al., 2020), or combining automated driving with repeating or generating (current study). Although automated driving by itself does not necessarily bring susceptibility to the lowest levels, as soon as another task is combined with it (be it some manual driving as in Van der Heiden, 2018, or a cognitive task), susceptibility is reduced.

Having a low level of susceptibility might be problematic during manual driving as the associated brain process is interpreted to reflect the process of orienting to novel stimuli and the susceptibility to new information (Friedman, Cycowicz, & Gaeta, 2001; Polich, 2007; Kenemans, 2015). So, for example, the ability to orient (and subsequently respond) to an unexpected alert or sound in the driving environment such as a dog running after a ball. A reduced susceptibility is probably even more problematic under automated driving conditions in SAE level 3, where the driver might be engaged in a non-driving task while automation is controlling the vehicle, but where the vehicle can demand human assistance at any time. Our work suggests that under such conditions, humans might have a general reduced susceptibility to alerts. As their prolonged work on a non-driving task might have limited their situational awareness of the driving environment, their ability to act might be reduced.

Although reduced susceptibility may not always lead to failed detection, in an ideal scenario (where alerts are critical), susceptibility should be high. System designers should take this reduced susceptibility into account, and develop strategies to overcome this, for example, by using multi-modal alerts or pre-alerts (Borojeni, Weber, Heuten, & Boll, 2018; Van der Heiden, Janssen, & Iqbal, 2017).

A comparable approach to issues of cognitive load and susceptibility during process control has been offered by Strayer and colleagues (e.g., 2013; 2015). In their EEG-based analysis the focus is on a P3 response over posterior cortical regions (also known as the ‘P3b’ response), which is normally elicited by events that are both relatively rare and task-relevant (e.g., targets for a behavioral response such as an emergency brake). The presently used fP3 (sometimes also referred to as ‘P3a’) is typically elicited by (highly) salient novels without any demand for an overt response. In this way it provides a continuous, yet unobtrusive measure for the susceptibility to potentially critical events that are outside the focus of direct task-associated attention. This is relevant in the context of automated driving, where drivers might occasionally focus on other tasks (e.g., writing an e-mail, handling a phone call) while the automation is handling most of the driving task. In addition, methodologically, the fP3 (or P3a) and P3b seem to differ in their ability to be captured under dynamic driving conditions. Whereas effects observed for the P3b under simulated manual driving did not always replicate under driving conditions in an instrumented vehicle (see Strayer et al., 2013; 2015), for the fP3 (or P3a), previous studies did replicate effects between simulated driving and on-the-road driving (see Wester, 2009, chapters 5 and 6).

Limitations & future work

Although in the current study both conditions in which a cognitive load inducing task is present (i.e., automated + generate and automated + repeat) showed a reduction in fP3 response compared to automated driving and to stationary, we did not find a difference between the two cognitive load inducing task conditions. This might be due to the timing of our probe; as outlined above this may have induced masking effects in the repeat condition. One way to avoid this, is to apply a delayed-response setting in which voice onsets during repeat are forced to occur much later, although admittedly this could induce undesired working memory load. Another option is to use longer intervals between noun and probe. Our previous study (Van der Heiden et al., 2020) showed that this does not affect fP3 during generating verbs, but this may be expected to not hold for repeating nouns (after the voice response fP3 may well recover to a single-task level).

The point in time that we measure is a limitation of our work in general. We probed susceptibility at a fixed interval: 0 ms after presentation of the noun stimulus. This interval was chosen as previous work that involved only the generate task found that extending the interval between stimulus and probe to 200 or 400 ms (i.e. in contrast to directly after) does not influence the level of measured susceptibility (Van der Heiden et al., 2020). Future work could also look into the effect over longer time spans, such as 1 s after stimulus offset. It is an open question whether susceptibility is fully restored after the oral response to the verb task (i.e., whether it is a phasic response process), or whether some level of reduced susceptibility remains (i.e., a tonic process).

A limitation of our set-up, in which the generate and repeat task trials are always succeeded by an oddball probe, is that the noun might function as a cue for an oddball probe, and thereby affect fP3 response. This way, the oddball stimulus is more predictable. Moreover, at that time, listening to an auditory sound is behaviorally relevant (because a response to the noun is needed). Previous work suggests that actively engaging in an auditory task at random times (i.e., occasionally pressing a button in response to a specific tone) can increase auditory susceptibility in general (Van der Heiden et al., 2018). Therefore, if anything, having a predictable probe might have resulted in relatively higher fP3 activation. If the effect of the cue would be controlled, then even lower levels of fP3 activation might be found in the repeat and generate conditions.

Implications for practice

Our results show that cognitive load can reduce general susceptibility to alerts. Therefore, it is important for safety-critical systems to take into account the possibility of delayed or absent response from the human operator due to such reduced susceptibility. In the case of automated driving, safety critical alerts such as handover of control requests might therefore build in resilient mechanisms, such as multi-modal alerts, or using earlier “pre-alerts” to forewarn a driver about an upcoming transition of control (Borojeni, Weber, Heuten, & Boll, 2018; Van der Heiden, Janssen, & Iqbal, 2017). Future work can look in more detail into the qualities of specific alarm types for different in-car applications.

Key points

An oddball probe was used to elicit an fP3 ERP to measure the effect of a cognitive load inducing task during automated driving.

We found that the fP3 is reduced when performing a task that induces cognitive load, either due to load induction by response production, or due to masking in one condition and load induction by semantic search in the other.

The results of this study can be used to inform designers of safety critical systems.

References

Abdullaev, Y. G., & Posner, M. I. (1998). Event-related brain potential imaging of semantic encoding during processing single words. *Neuroimage*, 7(1), 1-13.

Allison, B. Z., & Polich, J. (2008). Workload assessment of computer gaming using a single-stimulus event-related potential paradigm. *Biological psychology*, 77(3), 277-283. https://doi.org/10.1016/j.biopsycho.2007.10.014

Bainbridge, L. (1983). Ironies of automation. In *Analysis, design and evaluation of man–machine systems* (pp. 129-135). Pergamon.

Banks, V. A., Eriksson, A., O'Donoghue, J., & Stanton, N. A. (2018). Is partially automated driving a bad idea? Observations from an on-road study*. Applied Ergonomics*, 68, 138–145. https://doi.org/10.1016/j.apergo.2017.11.010

Bijl, S., de Bruin, E. A., Böcker, K. E., Kenemans, J. L., & Verbaten, M. N. (2007). Effects of chronic drinking on verb generation: an event related potential study. *Human Psychopharmacology: Clinical and Experimental*, 22(3), 157-166. https://doi.org/10.1002/hup.835

Borojeni, S. S., Weber, L., Heuten, W., & Boll, S. (2018, September). From reading to driving: priming mobile users for take-over situations in highly automated driving. *In Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services* (p. 14). ACM.

Burns, P. C., Parkes, A., Burton, S., Smith, R. K., & Burch, D. (2002). *How Dangerous is Driving with a Mobile Phone?: Benchmarking the Impairment to Alcohol* (Vol. 547). TRL.

Caird, J. K., Willness, C. R., Steel, P., & Scialfa, C. (2008). A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis & Prevention*, 40(4), 1282-1293.

Carsten, O., Lai, F. C. H., Barnard, Y., Jamson, a. H., & Merat, N. (2012). Control Task Substitution in Semiautomated Driving: Does It Matter What Aspects Are Automated? *Human Factors*, 54(5), 747–761. https://doi.org/10.1177/0018720812460246

Dekker, S. W., & Woods, D. D. (2002). MABA-MABA or abracadabra? Progress on human–automation co-ordination. *Cognition, Technology & Work*, 4(4), 240-244.

Dunn, N., Dingus, T., & Soccolich, S. (2019) *Understanding the Impact of Technology: Do Advanced Driver Assistance and Semi-Automated Vehicle Systems Lead to Improper Driving Behavior?* Washington, DC: AAA Foundation.

Fabiani, M., & Friedman, D. (1995). Changes in brain activity patterns in aging: the novelty oddball. *Psychophysiology*, 32(6), 579-594.

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews*, 25(4), 355-373. https://doi.org/10.1016/S0149-7634(01)00019-7

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and clinical neurophysiology*, 55(4), 468-484.

Hancock, P. A., Simmons, L., Hashemi, L., Howarth, H., & Ranney, T. (1999). The effects of in-vehicle distraction on driver response during a crucial driving maneuver. *Transportation Human Factors*, 1(4), 295-309.

Horrey, W. J., & Wickens, C. D. (2006). Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human factors*, 48(1), 196-205.

Iqbal, S. T., Ju, Y. C., & Horvitz, E. (2010, April). Cars, calls, and cognition: Investigating driving and divided attention. *In Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1281-1290). ACM.

Janssen, C. P., Donker, S. F., Brumby, D. P., & Kun, A. L. (2019). History and future of human-automation interaction. *International Journal of Human-Computer Studies.*

Janssen, C. P., van der Heiden, R.M., Donker, S. F., & Kenemans, J. L. (2019). Measuring susceptibility to alerts while encountering mental workload. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings* (pp. 415-420). New York, NY: ACM.

Kenemans, J. L. (2015). Specific proactive and generic reactive inhibition. *Neuroscience & Biobehavioral Reviews*, 56, 115-126. https://doi.org/10.1016/0013-4694(75)90263-1

Kunar, M. A., Carter, R., Cohen, M., & Horowitz, T. S. (2008). Telephone conversation impairs sustained visual attention via a central bottleneck. *Psychonomic bulletin & review*, 15(6), 1135-1140.

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 Seventh International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, 92–98.

Massar, S. A., Wester, A. E., Volkerts, E. R., & Kenemans, J. L. (2010). Manipulation specific effects of mental fatigue: evidence from novelty processing and simulated driving. *Psychophysiology*, *47*(6), 1119-1126.

Miller, M. W., Rietschel, J. C., McDonald, C. G., & Hatfield, B. D. (2011). A novel approach to the physiological measurement of mental workload. *International Journal of Psychophysiology*, 80(1), 75-78.

Murphy, S., Spence, C., & Dalton, P. (2017). Auditory perceptual load: A review. *Hearing Research*, *352*, 40-48.

Noy, I. Y., Shinar, D., & Horrey, W. J. (2018). Automated driving: Safety blind spots. *Safety Science*, *102*, 68–78. http://doi.org/10.1016/j.ssci.2017.07.018

O'Connell, R. G., Dockree, P. M., Robertson, I. H., Bellgrove, M. A., Foxe, J. J., & Kelly, S. P. (2009). Uncovering the neural signature of lapsing attention: electrophysiological signals predict errors up to 20 s before they occur. *Journal of Neuroscience*, 29(26), 8604-8611.

Oxford advanced learner's dictionary (2020) Definition of susceptibility. Accessed on 14 August 2020 at *https://www.oxfordlearnersdictionaries.com/definition/english/susceptibility*

Parasuraman, R., & Riley, V. (1997). Humans and automation: *Use, misuse, disuse, abuse. Human factors,* 39(2), 230-253.

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(3), 286-297.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1989). Positron emission tomographic studies of the processing of single words. *Journal of cognitive neuroscience,* 1(2), 153-170.

Pfleging, B., Rang, M., & Broy, N. (2016). Investigating user needs for non-driving-related activities during automated driving In: *Proceedings of the international conference on mobile and ubiquitous multimedia*, pp. 91–99. New York, NY: ACM.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology*, *118*(10), 2128-2148. https://doi.org/10.1016/j.clinph.2007.04.019

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

SAE International. (2018). j3016: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems. Retrieved from https://www.sae.org/standards/content/j3016_201806/

Scheer, M., Bülthoff, H. H., & Chuang, L. L. (2016). Steering demands diminish the early-P3, late-P3 and RON components of the event-related potential of task-irrelevant environmental sounds. *Frontiers in human neuroscience*, 10, 73. https://doi.org/10.3389/fnhum.2016.00073

Scheer, M., Bülthoff, H. H., & Chuang, L. L. (2018). Auditory Task Irrelevance: A Basis for Inattentional Deafness. *Human factors*, 60(3), 428-440. https://doi.org/10.1177%2F0018720818760919

Schmidt, E. A., Schrauf, M., Simon, M., Fritzsche, M., Buchner, A., & Kincses, W. E. (2009). Drivers' misjudgement of vigilance state during prolonged monotonous daytime driving. *Accident Analysis & Prevention*, *41*(5), 1087-1093.

Sharbrough, F. (1991). American Electroencephalographic Society guidelines for standard electrode position nomenclature. *J clin Neurophysiol*, *8*, 200-202.

Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators.* Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab.

Snyder, A. Z., Abdullaev, Y. G., Posner, M. I., & Raichle, M. E. (1995). Scalp electrical potentials reflect regional cerebral blood flow responses during processing of written words. *Proceedings of the National Academy of Sciences*, 92(5), 1689-1693.

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and clinical neurophysiology*, 38(4), 387-401.

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological science*, 12(6), 462-466.

Strayer, D. L., Cooper, J. M., Turrill, J., Coleman, J., Medeiros-Ward, N., & Biondi, F. (2013). *Measuring cognitive distraction in the automobile*. Washington, DC: AAAFoundation. https://aaafoundation.org/wp-content/uploads/2018/01/MeasuringCognitiveDistractionsReport.pdf

Strayer, D. L., Turrill, J., Cooper, J. M., Coleman, J. R., Medeiros-Ward, N., & Biondi, F. (2015). Assessing Cognitive Distraction in the Automobile. *Human Factors*, 57(8), 1300–1324. https://doi.org/10.1177/0018720815575149

Ullsperger, P., Freude, G., & Erdmann, U. (2001). Auditory probe sensitivity to mental workload changes–an event-related potential study. *International Journal of Psychophysiology,* 40(3), 201-209.

Van der Heiden, R. M., Iqbal, S. T., & Janssen, C. P. (2017). Priming drivers before handover in semi-autonomous cars. *In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 392-404). ACM.

Van der Heiden, R. M., Janssen, C. P., Donker, S. F., Hardeman, L. E., Mans, K., & Kenemans, J. L. (2018). Susceptibility to audio signals during autonomous driving. PloS one, 13(8), e0201963. https://doi.org/10.1371/journal.pone.0201963

Van der Heiden, R. M., Janssen, C. P., Donker, S. F., & Merkx, C. L. (2019). Visual in-car warnings: How fast do drivers respond?. *Transportation research part F: traffic psychology and behaviour*, *65*, 748-759.

Van der Heiden, R. M., Janssen, C. P., Donker, S. F., & Kenemans, J. L. (2020). The influence of cognitive load on susceptibility to audio. Acta psychologica, 205, 103058.

Wessel, J.R. and A.R. Aron, Unexpected Events nduce Motor Slowing via a Brain Mechanism for Action-Stopping with Global Suppressive Effects. *The Journal of Neuroscience*, 2013. 33(47): p. 18481-18491.

Wester AE. Attention Deficit and Impulsivity: Driving, Drugs and Electrophysiology. 2009. Utrecht: Utrecht University. PhD thesis.

Wester, A. E., Böcker, K. B. E., Volkerts, E. R., Verster, J. C., & Kenemans, J. L. (2008). Event-related potentials and secondary task performance during simulated driving. Accident Analysis & Prevention, 40(1), 1-7. https://doi.org/10.1016/j.aap.2007.02.014

Wester, A. E., Verster, J. C., Volkerts, E. R., Böcker, K. B., & Kenemans, J. L. (2010). Effects of alcohol on attention orienting and dual-task performance during simulated driving: An event-related potential study. *Journal of psychopharmacology*, 24(9), 1333-1348.

Zhang, B., De Winter, J., Varotto, S., Happee, R., & Martens, M. (2019). Determinants of take-over time from automated driving: A meta-analysis of 129 studies. *Transportation Research Part F: Traffic Psychology and Behaviour,* 64(5), 285– 307.

Biographies

Remo M.A. van der Heiden obtained his PhD in experimental psychology at Utrecht University (2020). He received his Master's degree in applied cognitive psychology from Utrecht University in 2015.

J. Leon Kenemans is a full professor of biopsychology and human psychopharmacology at Utrecht University. He received his PhD in psychophysiology from Utrecht University in 1990.

Stella F. Donker is an associate professor of experimental psychology at Utrecht University. She received her PhD in movement sciences from the University of Groningen in 2002.

Christian P. Janssen is an assistant professor of experimental psychology at Utrecht University. He received his PhD in human-computer interaction from UCL in 2012.