Acta Psychologica 205 (2020) 103058

Contents lists available at ScienceDirect

Acta Psychologica

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

The influence of cognitive load on susceptibility to audio

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

Utrecht University, Department of Experimental Psychology, Utrecht, the Netherlands
Helmholtz Institute, Utrecht, the Netherlands

ARTICLE INFO

Keywords:
P300
Frontal P3
Cognitive load
Susceptibility
Verb generation task
Oddball

ABSTRACT

In this study we evaluate how cognitive load affects susceptibility to auditory signals. Previous research has used the frontal P3 (fP3) event related potential response to auditory novel stimuli as an index for susceptibility to auditory signals. This work demonstrated that tasks that induce cognitive load such as visual and manual tasks, reduced susceptibility. It is however unknown whether cognitive load without visual or manual components also reduces susceptibility. To investigate this, we induced cognitive load by means of the verb generation task, in which participants need to think about a verb that matches a noun. The susceptibility to auditory signals was measured by recording the event related potential in response to a successively presented oddball probe stimulus at 3 different inter-stimulus intervals, 0 ms, 200 ms or 400 ms after the offset of the noun from the verb generation task. An additional control baseline condition, in which oddball response was probed without a verb generation task, was also included. Results show that the cognitive load associated with the verb task reduces fP3 response (and associated auditory signal susceptibility) compared to baseline, independent of presentation interval. This suggests that not only visual and motor processing, but also cognitive load without visual or manual components, can reduce susceptibility to auditory signals and alerts.

1. Introduction

1.1. Cognitive load

Cognitive load induced by a task can have a negative impact on performance of other tasks. For example, in everyday life, the cognitive load induced by contributing to a phone conversation can degrade driving performance (e.g., Brookhuis, de Vries, & De Waard, 1991; Horrey & Wickens, 2006). Beyond the task at hand, cognitive load can also affect long-term behavior such as learning. Cognitive load has therefore been considered for the design of various environments and interventions. Examples concern how cognitive (over-)load can be minimized during driving (e.g., Green, 2008) and how cognitive load theory can be used to optimize educational interventions and the design of educational resources (e.g., Castro-Alonso, Ayres, & Sweller, 2019; Sweller, van Merriënboer, & Paas, 2019).

Given this importance of cognitive load in many domains, it has therefore also been studied in detail in the cognitive (neuro-) science literature. Perhaps the best-known task is the psychological refractory period task (e.g., Anderson, Taatgen, & Byrne, 2005; Howes, Lewis, & Vera, 2009; Meyer & Kieras, 1997; Pashler, 1994; Schumacher et al., 1999). The consistent finding in these studies is that processing one task (e.g., a visual-manual task) can impact the processing time of another task (e.g., auditory-vocal task), and that the impact depends on how far apart the presentation of stimuli is.

The psychological refractory period task is an example of a dual-task experiment. Dual-task experiments require a participant to process and respond to two tasks. Many fruitful dual-task experiments have investigated cognitive load or the impact of cognitive load manipulations on task performance (e.g., Engström, Markkula, Victor, & Merat, 2017; Lavie, 1995, 2005; Wickens, 2002, 2008). However, a complication in such setups is that the exact response pattern and the nature of cognitive load are then dependent on the characteristics of both tasks. For example, whether both tasks are presented in the same or different modalities (e.g., Wickens, 2002, 2008), what difficulty level each of the tasks has (e.g., Lavie, 1995, 2005), and how practiced one is at the task (e.g., Engström et al., 2017). Given the complex interactions of these factors and the need to understand the context of testing, dual-task experiments are limited in their generalizability to measure the cognitive load of a single task.

1.2. Measuring auditory susceptibility

As an alternative, the load of one task might also be measured

* Corresponding author at: Utrecht University, Department of Experimental Psychology, Utrecht, the Netherlands.
E-mail address: r.m.a.vanderheiden@uu.nl (R.M.A. van der Heiden).

https://doi.org/10.1016/j.actpsy.2020.103058
Received 31 October 2019; Received in revised form 27 February 2020; Accepted 6 March 2020

through a more passive probe, for which no physical response is required. A promising approach is the use of an auditory novelty oddball paradigm, which has previously been successfully used to measure workload (e.g., Allison & Polich, 2008; Miller, Rietschel, McDonald, & Hatfield, 2011; Ullsperger, Freude, & Erdmann, 2001). In the novelty oddball probe technique, participants are presented with frequent stimuli, called standards, which are interspersed with rare stimuli, or novels (Friedman, Cycowicz, & Gaeta, 2001; Polich, 2007). For example, a common auditory version of the oddball task which we also use in our study plays 1000 Hz tones to the participant as standards, and occasionally intersperses these with unique environmental sounds such as coughs, sneezes, and bark sounds (Fabiani & Friedman, 1995). The consistent finding in the literature is that presentation of a novel stimulus creates a positive peak in electrical activity at the frontal central region of the brain, which peaks around 300 ms after stimulus/probe onset. This resulting peak is also known as the frontal P3 or fP3.

The common interpretation in the literature is that the fP3 peak is related to susceptibility towards new information, and to a process of orienting to novel stimuli (Friedman et al., 2001; Polich, 2007; for a review see: Kenemans, 2015). Moreover, previous research found that the fP3 peak is reduced when people perform visual-manual tasks, such as driving or manual tracking (e.g., Scheer, Bülthoff, & Chuang, 2016, 2018; Van der Heiden et al., 2018; Wester, Böcker, Volkerts, Verster, & Kenemans, 2008), which is interpreted as reflecting a reduction in susceptibility for unexpected but potentially relevant information (cf. Friedman et al., 2001; Kenemans, 2015; Polich, 2007).

The aforementioned studies demonstrated the effect of a concurrent task on susceptibility to novel auditory stimuli. However, these studies involved tasks that had visual stimuli or that required manual action (Scheer et al., 2016, 2018; Van der Heiden et al., 2018; Wester et al., 2008). Moreover, even previous oddball (fP3) studies that manipulated cognitive load all required visual stimuli and/or manual action (e.g., Allison & Polich, 2008; Miller et al., 2011; Ullsperger et al., 2001). It is therefore unclear whether the observed lower susceptibility to novel stimuli only occurs when visual or manual processing is involved. Would this same reduction in novelty processing also be present when the task does not involve manual or visual interaction? We address this question in the current study.

Since many tasks that people perform in daily life can create cognitive load, being able to measure the effect of cognitive load in the absence of visual and manual input/tasks on performance allows for a better general understanding of mental processing induced performance degradation. In particular, it is relevant to understand how such load affects susceptibility to sounds. For example, while driving to the supermarket, one might be thinking about what ingredients are needed to cook a meal (i.e., inducing cognitive load), which in turn might influence the ability to notice an unexpected signal, such as a car horn or blind spot warning, and to then react appropriately. Understanding correct perception of and reaction to unexpected signals is even more urgent today, given the rise of automated systems in which humans and (semi-automated) machines perform tasks together (Janssen, Donker, Brumby, & Kun, 2019), and in which humans form a crucial backup to respond to system alerts (Janssen, Iqbal, Kun, & Donker, 2019).

1.3. *Current study*

To investigate whether cognitive load in the absence of visual and manual input/tasks also reduces susceptibility to novel processing, we will induce cognitive load using the verb generation task (Abdullaev & Posner, 1998; Bijl, de Bruin, Böcker, Kenemans, & Verbaten, 2007; Snyder, Abdullaev, Posner, & Raichle, 1995). In this task, participants hear or see a noun, and in response either need to *generate* an appropriate verb for that word (e.g., apple-bite), or *repeat* the noun (e.g., apple-apple). Generating verbs in response to a noun (compared to repeating nouns) has been associated with higher cognitive load (Abdullaev & Posner, 1998; Snyder et al., 1995), with increased frontal activity in the cortex (Abdullaev & Posner, 1998; Bijl et al., 2007), and with increased dual-task interference (e.g., Iqbal, Ju, & Horvitz, 2010; Kunar, Carter, Cohen, & Horowitz, 2008; Strayer & Johnston, 2001; Van der Heiden et al., 2018). The common interpretation in the literature is therefore that the generate condition of the verb task induces relative high cognitive load. We will therefore use the auditory presented generate version of the task to induce cognitive load.

To assess the induced cognitive load while one devises a verb in the verb task, we present an oddball stimulus as a probe and record the subsequent brain activity with an electroencephalogram (EEG). By averaging multiple response measurements to an event related potential (ERP), a high signal to noise ratio can be achieved (e.g., Glatz, Krupenia, Bülthoff, & Chuang, 2018; Squires, Squires, & Hillyard, 1975; Van der Heiden et al., 2018; Wester et al., 2008). Verb generation is a mental process that involves different brain areas over time (Abdullaev & Posner, 1998; Bijl et al., 2007). For a written-spoken version of the verb-task, Abdullaev and Posner (1998) propose that the noun stimulus first triggers an attention-related network in the anterior cingulate region, followed by activation related to semantic processing in the left prefrontal cortex, followed by activation in Wernicke's area much later. It is yet unclear at which of these stages the most cognitive load is occurring. Therefore, we manipulated the timing between presentation of the noun and presentation of the probe, akin to manipulations in psychological refractory period task studies (e.g., Anderson et al., 2005; Howes et al., 2009; Meyer & Kieras, 1997; Pashler, 1994; Schumacher et al., 1999). This can reveal how different stages of the verb task (Abdullaev & Posner, 1998; Bijl et al., 2007) and their associated load change over time.

To summarize, previous research has observed a reduced susceptibility to auditory novel stimuli, which has been interpreted as reduced ability to allocate mental resources to novel stimuli in the environment (Kenemans, 2015). Our research question is whether cognitive load in the absence of visual stimulation and/or required manual action, as created through the verb generation task (Abdullaev & Posner, 1998), also reduces susceptibility towards auditory stimuli. We test this by recording the novel-elicited response following a verb generation task stimulus. As verb generation is a mental process that goes through different stages (Abdullaev & Posner, 1998; Bijl et al., 2007) with already early activity around electrode Fz (Abdullaev & Posner, 1998), we test susceptibility at various inter-stimulus intervals: 0, 200, and 400 ms after noun offset. We compare this to a baseline condition with no noun present. If cognitive load in the absence of visual stimulation and/or required manual action also reduces susceptibility, then we expect to find a fP3 response in the verb generation condition which is reduced compared to the baseline. In line with the psychological refractory period literature, we would also expect to find a higher interference at briefer inter-stimulus intervals.

2. Method

2.1. *Participants*

13 participants (8 M; 5 F) were recruited through social media using opportunity sampling (age 19 to 28 years of age, M = 23, SD = 2.6 years of age). All speakers were fluent in Dutch, 12 were native speakers, the one non-native participant spoke Dutch for 16 years. All participants reported normal hearing. The experiment was approved by the ethics committee of the Faculty of Social and Behavioral Sciences of Utrecht University (approval number FETC16-042). All participants gave written informed consent prior to the experiment. Participants were compensated €12 for their time.

2.2. *Materials*

2.2.1. *Presentation of auditory stimuli*

We used two types of auditory stimuli: an oddball probe and a verb generation task. All auditory stimuli were presented using Presentation (Neurobehavioural Systems) to participants through Earlink earphones at a constant level of 75 dB. Oddball stimuli were presented binaurally, word stimuli for the verb generation task only monaural through the earphone in the right ear.

2.2.2. Oddball probe

We used a 2-stimulus novelty oddball probe (Polich, 2007) in which 80% of the stimuli consisted of a 1000 Hz tone which was presented for 400 ms (labelled: standards). For the remaining 20% of the stimuli (labelled: novels) we presented novel sounds that were obtained from a database by Fabiani and Friedman (1995). This database consists of 100 unique environmental sounds, such as for example a person sneezing or a dog barking. The duration of the novels varied from 159 ms to 399 ms. All novels and standards were presented at 75 dB.

2.2.3. Verb generation task

In the verb generation task, participants heard a noun in a subset of the trials, and were asked to respond verbally with a verb that would fit the noun they heard (cf. Abdullaev & Posner, 1998). For example, when participants heard “apple” they could reply with “eat” or any other verb that they associated with apple.

As our participants were Dutch speakers, we created a Dutch word list. First, we translated the set of English nouns from Abdullaev and Posner (1998) to Dutch. Next, we removed all words that, as assessed by a research assistant and the researchers, were ambiguous in Dutch. As we wanted the words to be presented in a specific short interval, we also removed words that took long to pronounce, as assessed by having more than two syllables. We then converted the words to wave sound files using text-to-speech website www.texttospeech.io. We threw away words of which presentation took longer than 500 ms using default settings of the text-to-speech algorithm (Dutch female, volume 1, rate 1, pitch 1). The final set consisted of 144 unique words. Next, we determined the imaginability of each word, based on the scores in Van Loon-Vervoorn (1985). As the words differed in their imaginability score, we created 12 sets of 24 words each, such that for each set of words the mean and distribution of imaginability scores were comparable.

The next step was to play words at a constant length, while retaining intelligibility. Such constant length was needed to get the inter-stimulus timings between the verb task and the oddball probe exact. We conducted a small pilot on ourselves in which we changed the playback speed of all verb generation audio files, using Audacity (audacity.sourceforge.net). We eventually settled on a length of exactly 400 ms. This was the minimal length needed to have clear understandable word presentation, and also the same length as a standard tone in the oddball probe. For some words, this meant that they were played slightly faster or slower than the default text-to-speech algorithm. For example, the Dutch word “cel” (English: “cell”), originally took 450 ms to pronounce, which was brought back to 400 ms.

When participants performed the verb generation task in the experiment, our intent was for the word to be easily understandable, yet fast to present. Our intention was not to detect a specific response to a specific word. Therefore, we did not keep track of which words participants replied. We also encouraged participants to respond with the first word that came to mind, as we were only interested in the cognitive load inducing characteristics of the task, not in the actual words that participants came up with.

2.2.4. Intelligibility

To make sure participants were able to understand the words, we tested intelligibility of the words through a separate “repeat” task. In this task, all 144 words were played to the participants with 1600 ms between trials (i.e., offset word A and onset of word B). The task for the participants was to repeat the words. They were instructed to not reply in cases where they did not understand a word. Each participant received the words in the same order. During this stage, the experimenter made notes of mistakes and skipped words.

On average the words were correctly understood in 91% of the cases. The lowest participant score was 85% of words correct, and the highest 94%. There were no words consistently misinterpreted by all participants, rather the words that participants misinterpreted varied between participants.

2.2.5. Verb response reaction time

To be able to register the verbal response reaction time we placed two electrodes on the left cheek of the participants. Based on pilot tests, we set 40 μv for the average of both electrodes as a threshold for the detection of jaw movements, and associated speech. We then tested at what interval (relative to word presentation) jaw movements were detected, and whether such response times differed between experimental conditions.

2.3. Design

To measure the effects of the stage of word processing and associated cognitive interference on susceptibility to the oddball novel stimuli, we varied Inter-stimulus interval (ISI) between word offset (of the verb task) and onset of the oddball probes. We used a single factor within-subject design. *Inter-stimulus interval* was used as the factor with 4 levels: 0 ms, 200 ms, 400 ms, and Baseline. Fig. 1 shows presentation times visually.

2.3.1. Testing blocks

In total there were 12 experiment blocks. In each block, 80 oddball stimuli/trials (64 standards, 16 novels) were presented and 24 nouns were presented. Nouns for the verb generation task were followed by a standard tone of the oddball probe in half the cases (12 trials) and by a novel sound in the other half. One third of these combinations (i.e., 4 standards, 4 oddballs) was used with an ISI of 0 ms, one third with an ISI of 200 ms, and one third with an ISI of 400 ms. The position of these trials within the block was random. The remainder of the trials of a block was filled with presentation of 52 standard tones.

When all measurements of the 12 blocks are combined, each ISI condition had 48 words followed by a standard tone, and 48 words followed by a novel sound. For the baseline condition we also used 48 standard tones and 48 novel responses (which were not preceded by a noun from the verb task).

The interval between two subsequent standard tones was 2000 ms (cf. Van der Heiden et al., 2018; Wester et al., 2008). For trials where a noun was presented, the interval between noun and oddball was set by condition. The subsequent stimulus (either a noun or a standard tone) was presented 4400 ms after the onset of the noun. This longer interval was chosen to prevent any interference from generating a verb response.

2.4. Procedure

Participants first received general instructions and an overview of the experiment. They were then asked to read and sign the informed consent form. Subsequently, participants were asked to be seated in a driving simulator and adjust the seat to accommodate their feet and arms in an upright yet comfortable position.[1] Then intelligibility of the words was tested. We then applied 4 ocular electrodes to the face and 2 electrodes at left and right mastoid. A cap with 64 electrodes was placed on the participant's head and the signal from each electrode was

[1] For this specific study, a driving simulator setup is not needed. However, some of our other interests are in testing human performance under driving and automated driving conditions (e.g., Van der Heiden et al., 2018). To make our results comparable to future driving studies, we used this similar set-up.

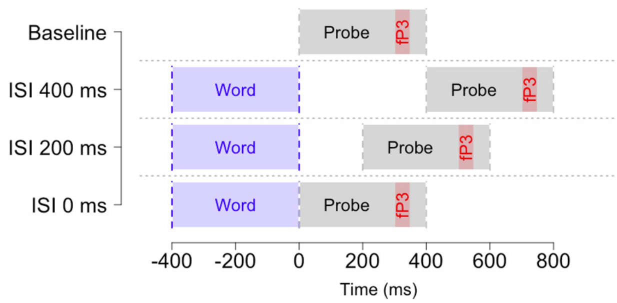

Fig. 1. Schematic representation of (inter-stimulus interval, or ISI) timings of word (noun from verb generation task) and oddball probe presentation. fP3 response was measured (305–355 ms after probe onset).

inspected to have an offset between 25 and −25 mV. Where needed, slight adjustments were made (e.g., adding extra conductive gel).

After receiving instructions, a practice block was used to allow participants to get accustomed with the verb generation and oddball probe. Twenty trials were played to the participant consisting of 16 standards and 4 novels, 6 of these stimuli were preceded by a word (3 standards, 3 novels). For these 6 stimuli, each ISI from the 3 possibilities (0, 200, 400 ms) was played in combination with a standard and with a novel in random order.

The main experiment consisted of 12 blocks with a short break after 6 blocks. When all blocks were finished the participants were asked to fill out a short questionnaire on demographics, experience and feedback. The total length of the experiment was 2 h.

2.5. Signal recording and analysis

EEG recordings were made using a BioSemi ActiveTwo system at 2048 Hz with 64 active Ag-AgCl electrodes, positioned following 10/10 system (Chatrian, Lettich, & Nelson, 1985). The average of left and right mastoids was used as reference. Horizontal and vertical EOG electrodes were placed on both outer canthi of the eyes, as well as above and below the right eye in line with the pupil.

EEG signal analysis was performed with Brain Vision Analyzer 2 software (Brain Products GmbH, München, Germany). First, data were re-referenced to the average of left and right mastoid signal. Then, three filters were applied: a 50 Hz notch filter to compensate for noise from the mains, a 0.16 Hz high-pass filter with a slope of 24 dB/oct, and a 30 Hz low-pass filter with a slope of 24 dB/oct to remove frequency bands that were not of our interest (cf. Van der Heiden et al., 2018; Wester et al., 2008).

Next, ERP segments of 2500 ms were distilled starting 1000 ms before the oddball stimulus onset, and ending 1500 ms after. To correct for eye movements and blinks, an ocular correction was applied to every segment using the Gratton, Coles and Donchin method (Gratton, Coles, & Donchin, 1983). For all conditions, data were baseline corrected over the 100 ms interval preceding probe presentation. Subsequently, artifacts in individual channels were rejected. Criteria for rejection of an epoch were at least one of the following conditions: 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 (standard settings in Brainvison Analyzer 2.1).

Finally, grand averages were calculated per probe stimulus type (standard and novel) per condition (Baseline, 0 ms, 200 ms, and 400 ms). Given words at certain ISI, potential ERP artifacts induced by these word stimuli should be equal or novel-elicited and standard-elicited ERPs. Therefore, to annihilate these artifacts *difference waves* were calculated by subtracting the average ERP for the standard tones from the average ERPs for the novels for each condition. Note that according to this logic, only the difference waves can have a meaningful interpretation, not the separate ERPs to novels and standards, respectively. Our analysis focused on the amplitude of the fP3 peak in these difference waves. We used a collapsed localizer that included all conditions to determine the location of the fP3 peak at electrode position FCz for this study (Luck & Gaspelin, 2017). This localizer showed two peaks, one at 303–353 ms and one at 307–357 ms. Since these were two relatively close peaks we used the difference wave between 305 ms and 355 ms after probe onset as the window for the fP3 peak.

Speech reaction time was calculated from the average response across the two electrodes on the cheek. Raw data were filtered by a 50 Hz notch filter to compensate for noise from the mains, a 10 Hz high-pass filter with a slope of 24 dB/oct, and a 60 Hz low-pass filter with a slope of 24 dB/oct. As a threshold to identify speech production we used 40 μv increase from baseline. For each participant in each condition we took the average speech reaction time from the trials that met the criterion.

2.5.1. Statistical analysis

For statistical analysis, we used a repeated measures ANOVA to analyze the effect of inter-stimulus interval (0, 200, 400 ms and baseline) on fP3 response and speech reaction time. We used an alpha level of 0.05 for significance. Where needed, this was followed by holm-corrected pairwise comparisons.

3. Results

3.1. fP3

Fig. 2 shows the difference waves (novel probe minus standard probe) for each *inter-stimulus interval*. Vertical lines indicate position of probe onset (0 ms) and fP3 peak. For statistical analysis we analyzed the mean fP3 value in the interval.

305–355 ms after probe onset. The barplot in Fig. 3 shows this mean peak activation for each condition.

There was a main effect of *inter-stimulus interval* on fP3 amplitude, $F(3,36) = 4.86$, $p = .006$, $\eta_p^2 = 0.28$. A holm-corrected post-hoc test showed that the baseline condition ($M = 10.4$ μv, $SD = 4.7$) had a significantly higher fP3 amplitude compared to all the other conditions in which the oddball presentation was preceded by a noun (all p's $< .05$). The amplitudes for the three different conditions where a noun was presented before the oddball did not differ significantly from each other: 0 ms ISI ($M = 5.7$ μv, $SD = 5.1$μv), 200 ms ISI ($M = 5.8$ μv, $SD = 3.9$ μv), and 400 ms ISI ($M = 5.9$ μv, $SD = 6.3$ μv). In other words: when an oddball probe was presented after presentation of a noun (i.e., when cognitive load was induced), the fP3 response was reduced compared to baseline (no cognitive load), irrespective of *when* the oddball was presented (0, 200, 400 ms ISI).

Fig. 4 shows activation across the scalp over time. It makes clear

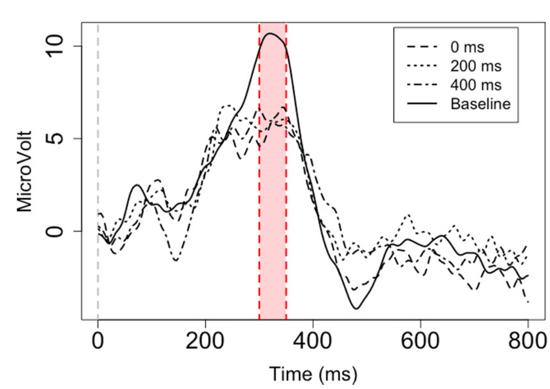

Fig. 2. Event related potential of 3 ISI conditions and baseline. Vertical lines show onset of oddball stimulus (0 ms) and fP3 peak area used for statistical analysis (305–355 ms).

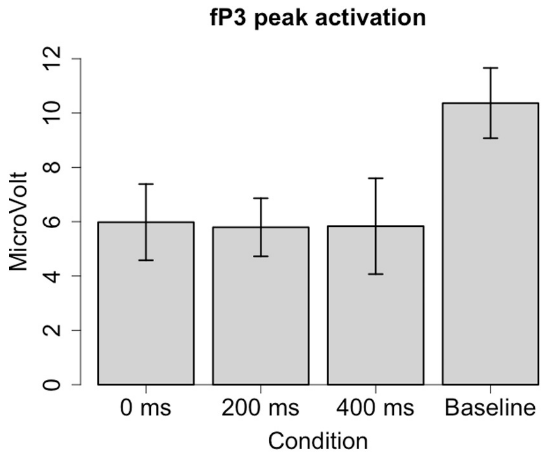

Fig. 3. fP3 peak activation in 3 ISI conditions and baseline. Error bars show standard error of the mean.

that the difference in activity between novel and standard is indeed most pronounced in frontal areas of the brain, during the interval of our analysis, and that the activity is reduced in the conditions with cognitive load (0, 200, 400 ms condition) compared to the baseline condition.

3.2. Speech reaction time

We analyzed the interval between word presentation onset and onset of speaking, using two electrodes on the cheek (see Materials). This measure of speech reaction time was analyzed to ensure that there were no differences in response time between verbs that were followed by a novel or by a standard oddball probe. We also analyzed whether this differed between inter-stimulus interval conditions. Results of a 2 (probe stimulus condition: novel or standard) × 3 (inter-stimulus interval: 0, 200, 400 ms) ANOVA showed that there was no significant difference between probe stimulus conditions (i.e., standard or novel), $F(1,12) = 0.056$, $p > .1$. However, there was a main effect of inter-stimulus interval, $F(2,24) = 30.68$, $p < .001$. A holm-corrected post-hoc test showed that reaction time differed between all three ISI conditions (all p's $< .005$). The reaction time was fastest for the 0 ms condition ($M = 1300$ ms, $SD = 162$ ms), followed by the 200 ms condition ($M = 1440$ ms, $SD = 180$ ms), followed by the 400 ms condition ($M = 1520$ ms, $SD = 181$ ms). This suggests some influence of ISI on the speech production process. There was no interaction effect between probe stimulus condition and inter-stimulus interval, $F(2,24) = 1.657$, $p > .1$. Further note that the lack of an effect of probe-stimulus condition on reaction time supports the idea that word processing was equal for novels versus standards, and therefore the ERP correlates of word processing can be annihilated by the novel-standard subtraction.

4. Discussion

We have tested whether cognitive load in the absence of visual stimulation and/or required manual action influences auditory susceptibility. Our experiment combined a cognitive load inducing task (verb generation task) with an auditory oddball novelty paradigm as probe task for susceptibility. Results show that auditory susceptibility to novel sounds is indeed reduced under conditions of cognitive load, independent of *when* the susceptibility is probed within an interval range of 0–400 ms after presentation of the load inducing stimulus.

Our results are consistent with previous work, which also found that concurrent task performance reduces susceptibility as measured through the novel response (e.g., Scheer et al., 2016, 2018; Van der Heiden et al., 2018; Wester et al., 2008). The result is also consistent with previous work that suggested that cognitive load can be measured with the oddball paradigm and fP3 measurements (Allison & Polich, 2008; Miller et al., 2011; Ullsperger et al., 2001). However, all these previous studies used tasks that involved visual stimuli and/or manual responses. As our work does not involve a visual or manual component, to the best of our knowledge it is the first to demonstrate that reduction in fP3 response can also occur without visual stimulation and/or required manual action, thereby strengthening the support for the claim that fP3 response measures the cognitive load associated with a task, and not merely visual- or manual engagement.

4.1. Limitations and future work

It is an open question whether the fP3 response can only be used to detect the presence or absence of cognitive workload (i.e., as a binary measure, as used in this paper), or whether it can also be used to detect

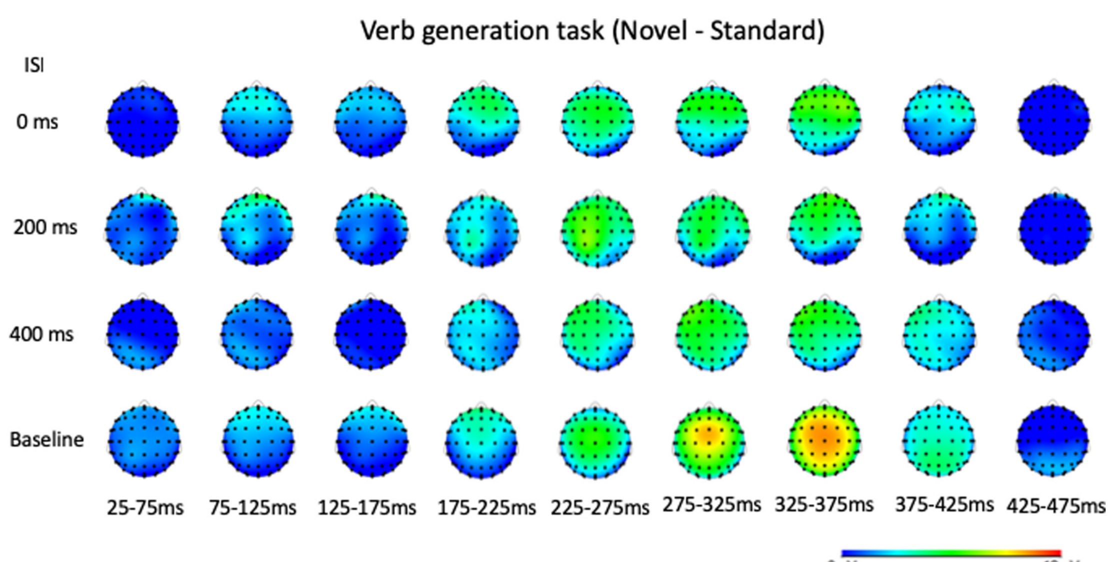

Fig. 4. Top view of electrode positions and activation over time for ISI conditions and baseline. For each inter-stimulus interval (rows), in 50 ms epochs (columns) the mean difference activation is expressed as a heatmap, with blue being close to zero (i.e. no difference) and red being 12 μv or more difference between the standard and novel oddball probed activation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

various levels of workload. Although we have not tested this here, previous work has shown that fP3 response level can be used to differentiate conditions that differ on other aspects of demand such as being stationary, driving manually, or being driven (Van der Heiden et al., 2018) or difficulty level of a visual-manual control task (Scheer et al., 2016).

Although our study has shown that cognitive load reduces the fP3 response, it is less clear what cognitive process underlies this reduction. Within our tested interval (0–400 ms after word onset), there was no distinction between fP3 response levels. However, previous work does suggest that generating a word consists of multiple stages (Abdullaev & Posner, 1998; Bijl et al., 2007; Snyder et al., 1995). Further research is needed to distinguish the cognitive load inducing effects of each stage. For the moment we conclude that when a novel probe is presented anywhere between 400 ms and 800 ms after noun onset, the fP3 that it elicits is reduced relative to a no-verb task conditions irrespective of when exactly between 400 and 800 the probe was presented. This is consistent with a model in which verb-generation processing induces capacity limits (for novel susceptibility) starting at least as early as 700 ms after verb onset (fP3 latency in the 0-ms ISI condition), and lasting at least as long as until 1200 ms after verb onset.

The current fP3 assessment shows the influence of cognitive load on auditory susceptibility. However, no performance measures were applied in relation to the novel-oddball stimuli. Therefore, the ramifications of reduced auditory susceptibility for adequate behavioral reactions to perient, possibly relevant auditory stimuli outside the primary-task context, remains a matter of speculation. An interesting option would be to use a three-stimulus oddball set-up including a repeated (e.g., 10% of the trials) deviant pitch, in which participants need to react to the deviant tone (see e.g., Scheer et al., 2016, 2018; Van der Heiden et al., 2018; Wester et al., 2008). These previous studies have found some indications for longer deviant-detection latencies under driving than under stationary conditions. In addition, adding a deviant-related task as such increased fP3 amplitudes in driving conditions to the level of stationary conditions. It would be interesting to find out whether these same dependencies hold when a task like verb generation is added.

It is unclear whether the induced load of the verb task differs between (semantically) incorrect replied verbs and correct replies. Our study did not collect the content of the generated words, and could therefore not analyze this. Future studies could look into the effects of correctness of the replied verb on (experienced) cognitive load and susceptibility to auditory stimuli.

Another avenue of future work can be the effect of modality of stimulus presentation and response. We measured susceptibility through use of auditory tasks with verbal response only. We chose an auditory version of the verb generation task since it is known to create central interference (Kunar et al., 2008; Strayer & Johnston, 2001) without visual distraction. It is however unknown how auditory susceptibility is affected by the use of other modalities, or what the effect of a multimodal task is on auditory susceptibility. Previous work has used visual presentation of noun stimuli (e.g., Abdullaev & Posner, 1998; Bijl et al., 2007). Future work can explore what the impact is of different modalities on experienced load and susceptibility.

4.2. Conclusion

Cognitive load without concurrent visual or manual interaction reduces auditory susceptibility, as assessed using a fP3 response. The time between load inducing stimulus and auditory susceptibility probe did not influence this reduction, within the interval we measured (0–400 ms). This suggests that the cognitive process of generating a word provides a comparable level of cognitive load during this interval. This is consistent with a model in which verb-generation processing induces capacity limits (for novel susceptibility) starting at least as early as 700 ms after noun onset (fP3 latency in the 0-ms ISI condition), and lasting at least as long as until 1200 ms after noun onset.

CRediT authorship contribution statement

Remo M.A. van der Heiden:Conceptualization, Methodology, Software, Writing - original draft, Writing - review & editing, Investigation, Formal analysis, Data curation, Visualization, Project administration.**Christian P. Janssen**:Conceptualization, Methodology, Supervision, Writing - review & editing, Funding acquisition.**Stella F. Donker**:Conceptualization, Methodology, Supervision, Writing - review & editing.**J. Leon Kenemans**:Conceptualization, Methodology, Supervision, Writing - review & editing.

Acknowledgements

Interim results of this work have been presented as a poster at Auto-UI 2019 (Janssen, Van der Heiden, Donker, & Kenemans, 2019). 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.

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. https://doi.org/10.1006/nimg.1997.0309.

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.

Anderson, J. R., Taatgen, N. A., & Byrne, M. D. (2005). Learning to achieve perfect time sharing: Architectural implications of Hazeltine, Teague, and Ivry (2002). *Journal of Experimental Psychology: Human Perception and Performance, 31*, 749–761. https://doi.org/10.1037/0096-1523.31.4.749.

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.

Brookhuis, K. A., de Vries, G., & De Waard, D. (1991). The effects of mobile telephoning on driving performance. *Accident Analysis & Prevention, 23*(4), 309–316. https://doi.org/10.1016/0001-4575(91)90008-S.

Castro-Alonso, J. C., Ayres, P., & Sweller, J. (2019). Instructional visualizations, cognitive load theory, and visuospatial processing. *Visuospatial processing for education in health and natural sciences* (pp. 111–143). Cham: Springer.

Chatrian, G. E., Lettich, E., & Nelson, P. L. (1985). Ten percent electrode system for topographic studies of spontaneous and evoked EEG activities. *The American Journal of EEG Technology, 25*(2), 83–92. https://doi.org/10.1080/00029238.1985.11080163.

Engström, J., Markkula, G., Victor, T. W., & Merat, N. (2017). Effects of cognitive load on driving performance: The cognitive control hypothesis. *Human Factors, 59*(5), 734–764 (doi:10.1177%2F0018720817690639).

Fabiani, M., & Friedman, D. (1995). Changes in brain activity patterns in aging: The novelty oddball. *Psychophysiology, 32*(6), 579–594. https://doi.org/10.1111/j.1469-8986.1995.tb01234.x.

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.

Glatz, C., Krupenia, S. S., Bülthoff, H. H., & Chuang, L. L. (2018). Use the right sound for the right job: Verbal commands and auditory icons for a task-management system favor different information processes in the brain. *Proceedings of the SIGCHI conference on human factors in computing systems*New York, NY: ACM. https://doi.org/10.1145/3173574.3174046 paper 472.

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. https://doi.org/10.1016/0013-4694(83)90135-9.

Green, P. (2008). *Driver interface/HMI standards to minimize driver distraction/overload (no. 2008-21-0002). SAE technical paper*.

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 (doi:10.1518%2F001872006776412135).

Howes, A., Lewis, R. L., & Vera, A. (2009). Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action. *Psychological Review*, 717–751. https://doi.org/10.1037/a0017187.

Iqbal, S. T., Ju, Y. C., & Horvitz, E. (2010). Cars, calls, and cognition: Investigating driving and divided attention. *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1281–1290). New York, NY: ACM. https://doi.org/10.1145/1753326.1753518.

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, 131*, 99–107. https://doi.org/10.1016/j.ijhcs.2019.05.006.

Janssen, C. P., Iqbal, S. T., Kun, A. L., & Donker, S. F. (2019). Interrupted by my car? Implications of interruption and interleaving research for automated vehicles. *International Journal of Human-Computer Studies, 130*, 221–233. https://doi.org/10.1016/j.ijhcs.2019.07.004.

Janssen, C. P., Van der Heiden, R. M. A., Donker, S. F., & Kenemans, J. L. (2019). Measuring susceptibility to alerts while encountering mental workload. *Adjunct proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications*New York, NY: ACM. https://doi.org/10.1145/3349263.3351524.

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. https://doi.org/10.3758/PBR.15.6.1135.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Learning and Memory, 21*(3), 451–468. https://doi.org/10.1037/0096-1523.21.3.451.

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences, 9*(2), 75–82. https://doi.org/10.1016/j.tics.2004.12.004.

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology, 54*(1), 146–157. https://doi.org/10.1111/psyp.12639.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance. 2. Accounts of psychological refractory-period phenomena. *Psychological Review, 104*(4), 749–791.

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. https://doi.org/10.1016/j.ijpsycho.2011.02.003.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116*(2), 220.

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.

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 (doi:10.1177%2F0018720818760919).

Schumacher, E. H., Lauber, E. J., Glass, J. M., Zurbriggen, E., Gmeindl, L., Kieras, D. E., & Meyer, D. E. (1999). Concurrent response-selection processes in dual-task performance: Evidence for adaptive executive control of task scheduling. *Journal of Experimental Psychology: Human Perception and Performance, 25*(3), 791–814.

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. https://doi.org/10.1073/pnas.92.5.1689.

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 (doi:10.1111%2F1467-9280.00386).

Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review,* 1–32.

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. https://doi.org/10.1016/S0167-8760(00)00188-4.

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 Loon-Vervoorn, W. A. (1985). *Voorstelbaarheidswaarden van Nederlandse woorden: 4600 substantieven, 1000 verba en 500 adjectieven.* Swets & Zeitlinger.

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.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*(2), 159–177. https://doi.org/10.1080/14639220210123806.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors, 50*(3), 449–455. https://doi.org/10.1518/001872008X288394.