Janssen, C. P., Schutte, I., & Kenemans, J. L. (2024). Does level of cognitive load affect susceptibility?. *International Journal of Psychophysiology*, 112443.
https://doi.org/10.1016/j.ijpsycho.2024.112443

Does level of cognitive load affect susceptibility?

Christian P. Janssen[1]*, Iris Schutte[1,2], J. Leon Kenemans[1]

[1]: Utrecht University, Experimental Psychology and Helmholtz Institute

[2]: Twente University, Faculty of Science and Technology and Library, ICT-Services & Archive

* Corresponding author:

Dr. Christian P. Janssen

Experimental Psychology & Helmholtz Institute

Utrecht University

Heidelberglaan 1

room H0.52

3584 CS Utrecht

The Netherlands

c.p.janssen@uu.nl

+31 (0)30 253 4582

iris.schutte2706@gmail.com

j.l.kenemans@uu.nl

Abstract

We compared how different levels of cognitive load affect frontal P3 (fP3) Event-Related Potential (ERP) to novel sounds. Previous studies demonstrated the predictive value of the probe-elicited frontal P3 (fP3) ERP for subsequent detection failures. They also demonstrated how fP3 is reduced when performing visual and/or manual and/or cognitively demanding tasks. These results are consistent with fP3 indexing orienting to novels or, more neutrally: susceptibility. Here, we tested how fP3 is affected by a threefold variation of cognitive load induced by the verb (generation) task. Participants heard a noun and either listened to it, repeated it, or generated a semantically related verb. These conditions were manipulated between groups. One group (N = 16) experienced the listen and repeat condition; the other group (N = 16) experienced the listen and generate condition. When fP3 was probed 0 or 200 ms after noun offset, it was reduced (relative to no noun) only while repeating or generating, not while listening. An additional probe-elicited ERP was identified as novelty-related negativity, and its contaminating influence on fP3 estimation accounted for by a novel vector-filter procedure. We conclude that cognitive load does not affect fP3-indexed susceptibility. Instead, fP3-indexed susceptibility is affected by presentation of the stimulus, with the most pronounced effect in conditions where a vocal response is needed (i.e., repeat or generate, but not listen), independent of the complexity of the response.

Keywords: novel P3, attention, verb task, Event Related Potential, Susceptibility

1. Introduction

There are many situations in which humans need to respond to unexpected situations which are triggered by auditory stimuli: the cry of a baby, a kitchen timer going off, an incoming e-mail alert, a shout of a friend to watch out for a bicycle crossing the road, or an in-car alert to take over control of a semi-automated vehicle. In all these cases, an auditory event triggers a process to briefly interrupt what you were doing, to orient to something that relates to the sound (baby, sauce pan, e-mail, bicycle, car), and to decide whether and how to take appropriate action.

Interrupting and orienting are aspects of susceptibility to such auditory stimuli. Susceptibility is used here to denote the extent to which an observer or process controller is in a mode which allows for detection of external signals to such a degree that an adequate behavioral response can be based on the detection (see also Van Der Heiden et al., 2018; van der Heiden et al., 2020; Van der Heiden et al., 2022). It is important that susceptibility can be assessed without overt-responding demands.

A successful paradigm in this regard has been the auditory novelty oddball paradigm (Friedman, Cycowicz, & Gaeta, 2001; Kenemans, 2015; Polich, 2007). In this paradigm, participants hear a series of tones, which are interspersed with occasionally presented unique, novel environmental sounds that elicit a pronounced Event-Related Potential (ERP) in the EEG, referred to as frontal P3 or fP3, in ERP measurements of frontal brain regions (electrode FCz). The fP3 had been associated with reorienting of attention, but perhaps more convincingly with generalized inhibition in response to an unexpected but potentially relevant event (Kenemans, 2015; Wessel & Aron, 2013). For example, in the latter study the ERP fP3 signature of outright stopping was found to be activated by unexpected novels as well.

The usefulness of fP3 as a measure of susceptibility is also underscored to the extent that fP3 is increased when the novel stimuli can be used as cues for behaviorally relevant subsequent (e.g., visual) stimuli (Escera, Alho, Schröger, and Winkler, 2000; Escera, Alho, Winkler, & Näätänen, 1998). Even more pertinent, O'Connell et al. (2009) reported that missed, relative to detected, targets in a visual detection task were preceded by reduced fP3s to preceding stimuli. The fP3 is also reduced when people are engaged in other visual or manual tasks, such as visual-manual tracking (e.g., Scheer, Bülthoff, & Chuang, 2016; 2018) and manual and automated driving (Figalová, Bieg, Reiser, Liu, Baumann, Chuang, Pollatos, 2024; Figalová, Pichen, Chandrayan, Pollatos, Chuang, Baumann, 2023; van der Heiden, 2020; van der Heiden et al., 2018; van der Heiden, Kenemans, Donker, & Janssen, 2022; Wester, Böcker, Volkerts, Verster, & Kenemans, 2008).

Of note, a sizable literature exists about a related, yet different, frontal positive wave named P3a. The P3a is elicited by modest changes in basic features (e.g., pitch) (e.g., Muller-Gass & Schröger, 2007; Sabri, Liebenthal, Waldron, Medler, & Binder, 2006; Volosin & Horváth, 2020; Wiens, van Berlekkom, Szychowska, & Eklund, 2019) whereas the fP3 is elicited by unique novel environmental stimuli. It is questionable whether the 'deviance-elicited' P3a can be identified with the novelty-elicited fP3. First, when P3a en fP3 are compared in single samples, the latter is 4 to 6 times larger in magnitude (Wester, Verster, Volkerts, Böcker, Kenemans, 2010). Second, fP3s show robust Fz/ FCz maximum distributions (e.g., Figalová et al., 2023, 2024; Scheer et al., 2016, 2018; van der Heiden et al., 2022; Wester et al., 2008), whereas this is less clear for P3a, where the distribution may be more centro-parietally focused (e.g., Volosin & Horváth, 2020), and/ or reflect contributions from temporal regions.

Most of the preceding studies investigated susceptibility to sounds (using fP3) during performance of a visual or manual task, rather than investigating susceptibility during cognitive and auditory tasks. In our study, we therefore test how cognitive load (induced by a verb task, introduced later) affects susceptibility to novel sounds.

1.1 Inducing cognitive load using a verb task

One study that has studied how susceptibility (fP3 response) is affected by cognitive load is by Van der Heiden and colleagues (2020). They examined how susceptibility is influenced by an auditory cognitive task: the verb task. In the verb task, participants hear a noun, and need to respond vocally with a semantically related verb (e.g., “bed” -> “sleep”, “water” -> “drink”) (Petersen, Fox, Posner, Mintun, & Raichle, 1988). The process of generating a word involves various brain regions over time, including frontal regions (Abdullaev & Posner, 1998; Snyder, Abdullaev, Posner, & Raichle, 1995). Generating a verb under dual-task conditions creates dual-task interference (Iqbal, Ju, & Horvitz, 2010; Kunar, Carter, Cohen, & Horowitz, 2008; Strayer & Johnston, 2001; van der Heiden, Janssen, Donker, & Merkx, 2019), and is therefore thought to induce cognitive load.

Van der Heiden and colleagues (2020) found that susceptibility to novel stimuli (assessed using fP3) is reduced during the verb generation process. They probed susceptibility 0, 200, and 400 ms after noun offset and consistently found a reduction in fP3. Similarly, a reduction was found when the generation task was combined with automated driving, and when participants had to repeat the noun under automated driving conditions (van der Heiden et al., 2022). However, the repeat condition has not been tested without driving (Van der Heiden et al., 2022), and neither of the Van der Heiden studies (2020, 2022) has included a simpler “listen only” condition to separate any effects that come from preparing a word versus merely listening to a word. We therefore also include a listen condition in the current study, and we do not combine the verb task with other tasks (apart from probing for susceptibility). In parallel with this submission, preliminary results of Figalová et al. (2023) were presented where a driving task was combined with a visual N-back task and an oddball paradigm. Similar to Van der Heiden et al. (2022), it was found that fP3 response was reduced under conditions where driving was combined with a high cognitive load task (2-Back task). However, there was no control for conditions without driving and with various levels of cognitive load.

More generally, tasks and activities can differ in the amount of cognitive load that they produce. Therefore, in the present study, we test how susceptibility varies under different cognitive load conditions, again using the verb task. Specifically, participants hear a noun and either need to: *listen* to the noun (without vocal response), *repeat* the noun, or *generate* a semantically related verb. Dual-task paradigms suggest that these tasks create increasing levels of cognitive load and increasing dual-task interference (Kunar et al., 2008; Strayer & Johnston, 2001). Does this gradual increase in cognitive load also result in a gradual reduction of auditory susceptibility?

At the time these data were collected it was not clear yet what the effect was of the temporal interval between the (offset of the) presentation of the noun and the (onset of the) presentation of the sound-probe (standard or novel) (this was later tested and reported in Van der Heiden et al., 2020). We therefore maintained two intervals (0 ms and 200 ms, also used in Van der Heiden et al., 2020) within each version of the verb task (listen, repeat, generate). The intervals were chosen such that they were close to the word presentation, but far away from the moment at which a participant spoke and mouth movements might interfere with EEG measurement.

Finally, we also included a control condition in these trials, which probed susceptibility without a preceding noun. Based on preceding work, we hypothesized the following pattern of results:

- Similar to Van der Heiden and colleagues (2020), we expected the fP3 response to be highest under control conditions where a probe was not preceded by a noun. This is also consistent with the larger body of fP3 literature in which a strong fP3 peak is expected in response to a novel stimulus (Friedman et al., 2001; Kenemans, 2015; Polich, 2007).
- If susceptibility is affected by cognitive load level (which has not yet been tested before in contrast with a 'listen' condition), then we expected little or no reduction in fP3 amplitude in the listen condition, independent of the moment of probing. This is consistent with theories that the listen task generates little cognitive load (Kunar et al., 2008; Strayer & Johnston, 2001).
- The strongest reduction in fP3 is expected under the condition with highest cognitive load: the generate condition (also cf. Figalová et al., 2023, that high cognitive load reduces fP3). Some reduction is expected in the repeat condition (cf. Van der Heiden et al., 2022), as some cognitive load may be involved when transforming a stimulus into a response, however constrained the response may be (Kunar et al., 2008; Strayer & Johnston, 2001).
- At the time of data collection, we had not yet a hypothesis about which time interval any effect of fP3 reduction might be pronounced most. However, based on later work by Van der Heiden et al. (2020, using a generate condition), we expected that the effects of the thinking process associated with generating verbs and repeating nouns reduces fP3 amplitude independent of probe timing (0 or 200 ms).

2. Method

2.1 Participants

Thirty-two participants[1] participated in the study (20 female; 12 male), with a mean age of 24.1 years (*SD* = 4.2 years, range 19-36 years). They were split into two groups (see design) of 16 participants each. The listen-repeat group consisted of 10 female and 6 male participants with a mean age of 24.9 years (*SD* = 2.5 years, range 20-30 years). The listen-generate group consisted of 10 female and 6 male participants with a mean age of 23.9 years (*SD* = 5.4 years, range 19-36 years).

Participants were recruited via advertisement at Utrecht University and via a recruitment website. All were native Dutch speakers and reported to have normal hearing. A control test on word intelligibility was conducted to objectively test participants' ability to hear and repeat words. This test confirmed that none of the participants had structural hearing loss and that there were no structural problems with auditory stimuli either (see Supplementary materials S1).

The experiment was approved by the ethics committee of the Faculty of Social and Behavioral Sciences of Utrecht University (approval number FETC16-042) and conducted in accordance with the declaration of Helsinki. All data were collected between October and December 2019. Participants received 6 euro per hour for participation.

[1] G power does not allow an accurate power analysis for the type of between-within design applied in our study, so we estimated our sample size based on previous studies and their ability to detect effects: 18 for a between-subjects design in (van der Heiden et al., 2018), and 13 (van der Heiden et al., 2020) and 24 (van der Heiden et al., 2022) for studies with a within-subject manipulation of respectively timing of verbs and generate/repeat. We ended up with 32 as a conservative choice, that allowed for sufficient counter balancing.

2.2 Task

Participants performed an auditory verb (generation) task (Petersen et al., 1988), while also being probed for susceptibility using an auditory novelty oddball task (Friedman et al., 2001; Kenemans, 2015; Polich, 2007). All sounds were presented while subjects looked at a white fixation cross on a black computer screen. Presentation software (Neurobehavioural Systems) was used for stimulus presentation and sounds were played binaurally at approximately 75 dB through Earlink earphones. The peak levels of all sound and word stimuli were equalized.

Verb task

For the verb task, participants heard a series of nouns, and either had to:

- *Listen*: listen to the noun without any response
- *Repeat*: vocally repeat the noun (e.g., “bed” -> “bed”)
- *Generate*: generate and vocalize a verb that is semantically related to the noun (e.g., “bed” -> “sleep”)

For the stimuli, we used 120 Dutch nouns from a set of 144 words that was developed by Van der Heiden and colleagues (2020), based on the English list of Abdullaev and Posner (1998). The 120 selected words had the best audibility scores (i.e., ability to correctly repeat) in Van der Heiden’s study and had the highest imaginability scores (van Loon-Vervoorn, 1985). Each noun was pronounced by a text-to-speech program (similar to van der Heiden et al., 2020). The duration of the nouns was set such that it was close to the length of the standards and novels (see below), while still being audible as a noun. In the end, this resulted in words that had a length of on average 398 ms (SD = 0.5 ms; range 396.6-400 ms).

Sound probes

For the sound probes, stimuli consisted either of standard tones (1000 Hz pure tones) or novel stimuli (environmental sounds like sneezing or chirping of a bird from a database by Fabiani & Friedman, 1995). Within each block, 80% of the sound probes were standards, and 20% were novels. The novel stimuli had the original duration of the Fabiani & Friendman dataset. That is, an average duration of 338 ms (*SD* = 60 ms; range 161 – 403 ms). All standard tones were set to the average of the novels: 338 ms.

Verb task with sound probes

Within each block where the verb task was combined with the sound probes, participants were probed 80 times with a sound probe (i.e., standard or novel stimulus). Of these, 40 probes were not preceded by a noun (control condition), and 40 were preceded by a noun, randomly alternating. The probability that a sound probe was a novel was kept at 20% (i.e., 8 of the 40 probes), and novels and standards also alternated randomly. The final combinations of stimulus pairings (standard / novel probe and word) per block are listed in Table 1.

Table 1: Frequency of stimuli pairings per block of trials.

	Standard	Novel
Control: no word, only probe	32	8
0 ms: word + probe	16	4
200 ms: word + probe	16	4

Note: Participants experience 9 blocks for listen, and 9 blocks for repeat / generate. Across all 18 blocks, each unique word is used 6 times.

For trials where a noun preceded a probe, the interstimulus interval between noun offset and probe onset was either 0ms (with 16 standards, 4 novels) or 200 ms (with 16 standards, 4 novels), for each of 2x9 blocks (see below). To avoid that the measurement of the next probe was influenced by any remnant process related to the verb stimulus and/or speaking, in cases where a probe followed a word, the next probe was presented 4000 ms after offset of the first probe. In all other cases (i.e., where a noun had not preceded a probe), the interval between a sound probe offset (i.e., standard or novel) and the next onset (i.e., either noun or sound probe) was 2000 ms. Figure 1 shows the relative timing of words, probes, and fP3 measurement. Figure 2 shows examples of the intervals between subsequent stimuli.

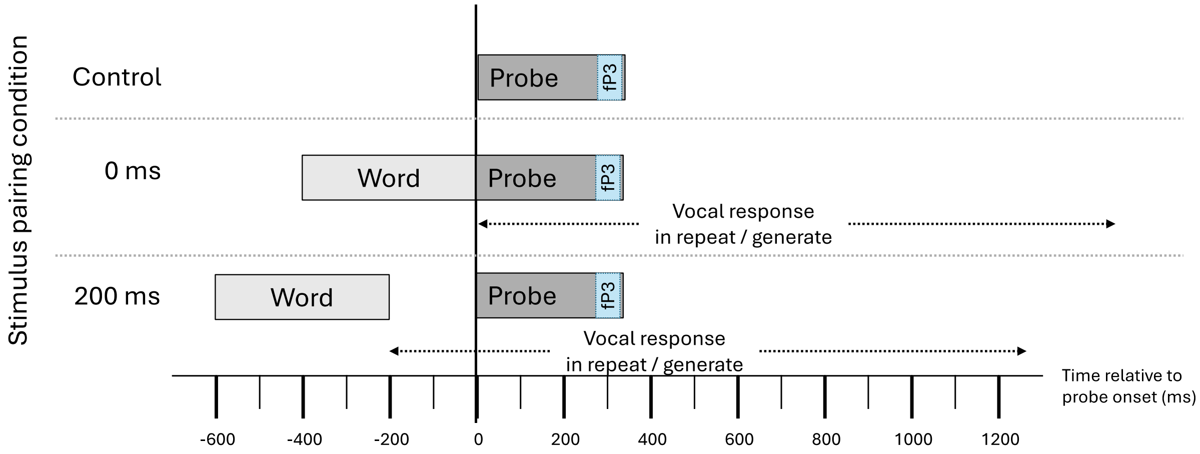

Figure 1: Relative timings of stimuli for different stimulus pairing conditions, with timing expressed relative to probe onset. If a word was presented, the probe (standard or novel) was either presented 0 or 200 ms after word offset. In the control conditions, no word preceded a probe. 275-325 ms after probe onset we measured the fP3 response. In the repeat and generate conditions, participants gave a vocal response to the word. In the listen condition they did not respond.

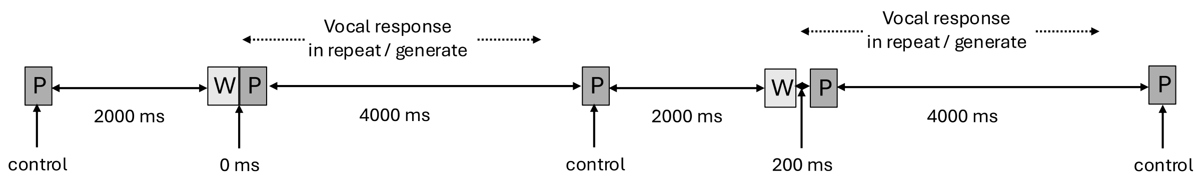

Figure 2: Example of relative timings of stimuli. There was alternation between presenting just a probe (P; "control" condition) or a word that was followed by a probe (W, P). If a previous trial only had a probe, there was 2000 ms before the next stimulus (Word). If a word was presented, the onset of the subsequent probe was either 0 or 200 ms after word offset (depending on stimulus pairing condition). If a word preceded a probe, then the onset of the next probe was 4000 ms. In repeat or generate trials, participants gave a vocal response after word offset.

2.3 Design

The study followed a 2 x 2 x 3 mixed-factorial design. First, we had two *groups* of participants. Half of the participants (listen-repeat group) completed the listen version and the repeat version of the task. The other half (listen-generate group) completed the listen version and the generate (verb) version of the task. Secondly, within subjects, we could analyze the *task*: whether trials came from a listen trial or from a trial where a vocal response was needed (repeat or generate, depending on group). Thirdly, within participants, we manipulated the *stimulus pairing* of the oddball probes and nouns (for listen, repeat, and generate). A sound probe (standard or novel) was presented either without a noun preceding it (control), 0 ms after noun offset, or 200 ms after noun offset. For this within-subjects manipulation of stimulus pairing, all three conditions (0ms, 200ms, control) were presented in each block at random positions. For the between-subjects manipulation of group, two equal groups were formed such that gender was balanced between groups.

For the manipulation of task, each participant performed 9 blocks of listen trials and 9 blocks of either repeat or generate trials (between subjects). Whether the task was listen or repeat/generate alternated from block to block, and across participants it was counterbalanced which block came first (listen or repeat / generate). Per block, 4 novels were measured per probing moment condition per participant, giving a total of 36 measurements per combination of stimulus pairing and task condition per participant.

We also controlled the order in which the nouns were presented. There were 120 unique nouns, but 9 x 40 = 360 noun exposures on listen trials and 360 exposures on the repeat/generate trials. Each participant was presented each unique noun 6 times: 3 times for listen and 3 times for repeat/generate.

We spaced out the nouns across the experiment such that the repeats were not close together. To achieve this practically, we made 3 sets (I, II, III) of 3 folders each (Ia, Ib, Ic; IIa, Iib, Iic; IIIa, IIIb, IIIc) with 40 words each (9 folders total, all 120 words used in set I, all in set II, and all in set III). Each folder (e.g., Ia vs IIIa) had a slightly different combination of words, with the constraint that each folder represented a comparable distribution of imaginability scores of the words. Specifically, the set of 120 words was split into four groups (top-25%, bottom-25%, upper-medium, and lower-medium imaginability score). Per folder, 10 words were selected from each group.

Each unique folder was used once for repeat/generate, and once for listen. However, no two subsequent folders were the same (achieved by e.g., first using the folders of I for listen and II for repeat/generate, then III for listen and I for repeat/generate, etc). In the end, although the words were presented in a structured manner, to the participant they appeared random and per set of 6 blocks (240 words) each word only occurred twice (but in total, across all 18 blocks they experienced each word six times).

2.4 Procedure

Figure 3 outlines the procedure. Upon arrival, participants received an instruction about the experimental procedure and signed a consent form. They were then asked to repeat words that were played through headphones. This was done to check audibility of the 120 words that were used in the experiment. An analysis of intelligibility is included in supplementary materials S1. No consistent patterns emerged, only in 3% of the cases was a word misheard or not heard at all.

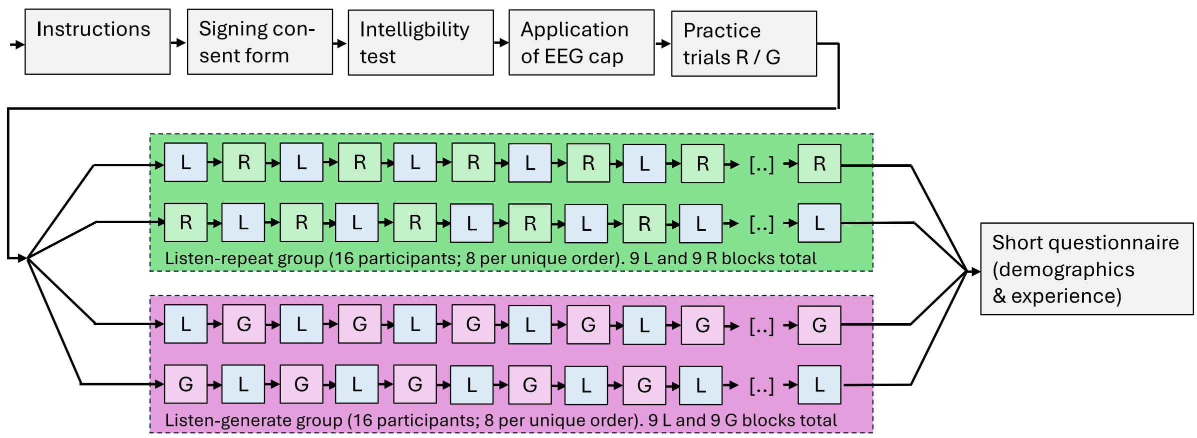

Figure 3: Experimental procedure. Experimental blocks are indicated with a letter as follows: L = Listen; R = Repeat; G = Generate. There were 2 groups: listen-repeat and listen-generate. Participants alternated between blocks with the Listen and the other condition (Repeat / Generate). Not pictured: there was a short break after the 4th and 14th block (1 minute) and a longer break (10 minutes) after the 9th block (i.e., midway).

Across all blocks, each novel sound was used 6 times: 3 times under listen conditions, and 3 times under repeat or generate. Counterbalancing of stimuli was done in such a way, that each novel stimulus was used twice per set of 6 blocks (i.e., once for listen and once for repeat/generate).

Subsequently, the EEG cap and electrodes were applied, and subjects were seated in a dimly lit room in front of a computer screen and received further task instruction. They then started with a short practice block, which had 20 sound probes of which 10 were preceded by a noun. The practice block was always either the generation or repeat condition (depending on whether the participant was assigned to the listen versus repeat or listen versus verb generate condition). The practice block was followed by the 18 main blocks (alternating between listen and generate, or between listen and repeat).

Before the start of each block, participants received on screen instructions about whether their task for this block was to listen, repeat, or generate. Breaks of about one and a half minute were provided after the fourth and fourteenth block, and a 10-minute break was provided halfway through the experiment. In addition, there was a possibility for a small, self-paced break after each block. The total duration of the task including breaks was approximately 1 hour and 43 minutes. At the end of the experiments, the EEG cap was removed, and participants completed a brief questionnaire for demographics information and subjective experience (analyzed in Supplementary Materials S2).

2.5 EEG data acquisition

For EEG data acquisition we used the Active-Two system with 64 Ag-AgCl electrodes (Biosemi, Amsterdam, The Netherlands). Scalp electrodes were placed according to the 10/10 system (Chatrian, Lettich, & Nelson, 1985). Vertical EOG electrodes were placed above and below the right eye and horizontal EOG electrodes were placed at the outer canthi of both eyes. Sampling rate was 2048 Hz and data were online low pass filtered at DC to 400 Hz. The Common Mode Sense/Driven Right Leg electrode was used as an online reference.

2.6 ERP data analysis

The choice for our main electrode of interest (FCz) and response interval (300 ms after onset of probe) follows preceding studies (e.g., van der Heiden, 2020; van der Heiden et al., 2018; van der Heiden, Kenemans, Donker, & Janssen, 2022; Wester, Böcker, Volkerts, Verster, & Kenemans, 2008). Although this perhaps qualifies as a sufficient rationale for the FCz choice, we note here that a confirmatory visualization is presented in Figure 6, and a confirmatory quantitative analysis in section 3.3.

ERP data were analyzed using Brainvision Analyzer 2.0 (Brain Products GmbH). Data were down sampled to 250 Hz (using Analyzer's Change-Sampling-Rate transformation with the spline-interpolation option, including anti-aliasing filtering with 112.5 Hz low-pass, 24 dB/ octave). and re-referenced to the average signal of the left and right mastoids. Data were filtered with a 0.16-30 Hz band pass (24 dB/oct), and an additional 50 Hz notch filter. ERP data were epoched into windows from -100ms to 1000 ms surrounding the onset of novel and standard sounds.

Segments with extreme artifacts (i.e., a difference of more than 1500 µV between two consecutive sample points) in the EOG channels and/or electrode FCz. This was done to prevent such extreme artifacts to contaminate the estimation of transfer coefficients from VEOG and HEOG channels to especially FCz, the envisaged electrode site for statistical analysis. We then performed ocular artifact correction using the Gratton & Coles algorithm (Gratton, Coles, & Donchin, 1983).

Data were baseline corrected using the 100-ms period before sound onset. An automatic artifact rejection procedure was applied for all individual channels using the following criteria: maximal allowed absolute difference between two values in an epoch of 100 µV, and lowest allowed activity within a 100-ms interval of 0.5 µV. Average ERPs were computed for each condition and subject. These data were used to calculate difference waves between the standard and novel probes. Note that there were more observations for the "standard" probe (compared to novel; see Table 1). We used all these samples to get a reference signal for the standard that has the least impact of noise.

Our criterion for keeping participant data after artifact rejection was that for each participant in each condition at least half of the segments for electrode FCz remained; this pertained to all participants. The lowest we observed was 69% of segments (25 of 36 segments in the novel+word after 0ms repeat condition). In all conditions, the first quantile of number of segments kept was 95% or higher. Similarly, scalp distribution plots were based on data of subjects for which at least half of the segments remained for each condition and channel. This pertained to 26 subjects.

Visual inspection of the novel-minus-standard waveforms across all experimental conditions indicated that the fP3 peaked at 300 ms after the sound onset. We therefore exported the average novel-minus-standard ERP activity between 275-325 ms for each condition and subject for further statistical testing.

Based on the grand averages of the difference wave, we performed additional control analyses of whether a novelty-related negativity effect (Wester, Verster, Volkerts, Böcker, Kenemans, 2010) was present and did additional analyses to rule out that this NRN-effect had an effect on the fP3 analysis.

2.7 Manipulation checks

As a manipulation check, we analyzed vocal response time and subjective experience rating. The literature suggests that generating a verb should take more time (i.e., later vocal response time compared to repeat condition) and be a harder task (i.e., higher subjective difficulty compared to repeat condition) (cf. e.g., Iqbal et al., 2010; Kunar et al., 2008; Strayer & Johnston, 2001; van der Heiden et al., 2019). The analyses are reported in Supplementary materials S2 and S3. Overall, these analyses suggest that the manipulation worked as intended: compared to the repeat condition, generating a verb required more time and was experienced as harder.

2.8 Statistical analysis

Unless otherwise noted, statistical analyses were conducted using SPSS 25, using a general linear model, repeated measures design with “group” as a between factor, and stimulus pairing (0, 200, control) and task (i.e., listen versus repeat/generate) as within-factors. Results of the univariate test (not involving the 3-level stimulus-pairing factor) and the multivariate test (involving the 3-level stimulus-pairing factor) are reported. Alpha was set at .05. For the stimulus pairing factor, we compared using within-subjects contrasts whether there was a linear effect (i.e., control and 200 ms condition differ from each other) and whether there was a quadratic effect (i.e., whether the 0 ms condition also differs from the average of the other conditions).

3. Results

3.1 ERP results: fP3

Figure 4 shows the grand averages of the difference wave (novel-standard) at electrode FCz, in the four conditions (see Figure caption)[2]. In addition, mean fP3 values are presented as barplots in Figure 5 and scalp distributions are shown in Figure 6. If a novel frontal P3 effect is present, we would expect a peak around 275-325 ms after probe presentation. This time interval is highlighted in blue in the Figure, and this is the area of which we calculated the mean value for statistics (means are presented in the bar plot in Figure 5). If there is an effect of cognitive load, we would expect this peak to be reduced in the conditions with a word present (0, 200 ms) and/ or where an action is required (i.e., repeat / generate) compared to the condition without a word preceding the sound probe ("control") and with an action requirement (i.e., listen).

[2] Separate plots of the waveforms of standard and novel can be found in supplementary material S4

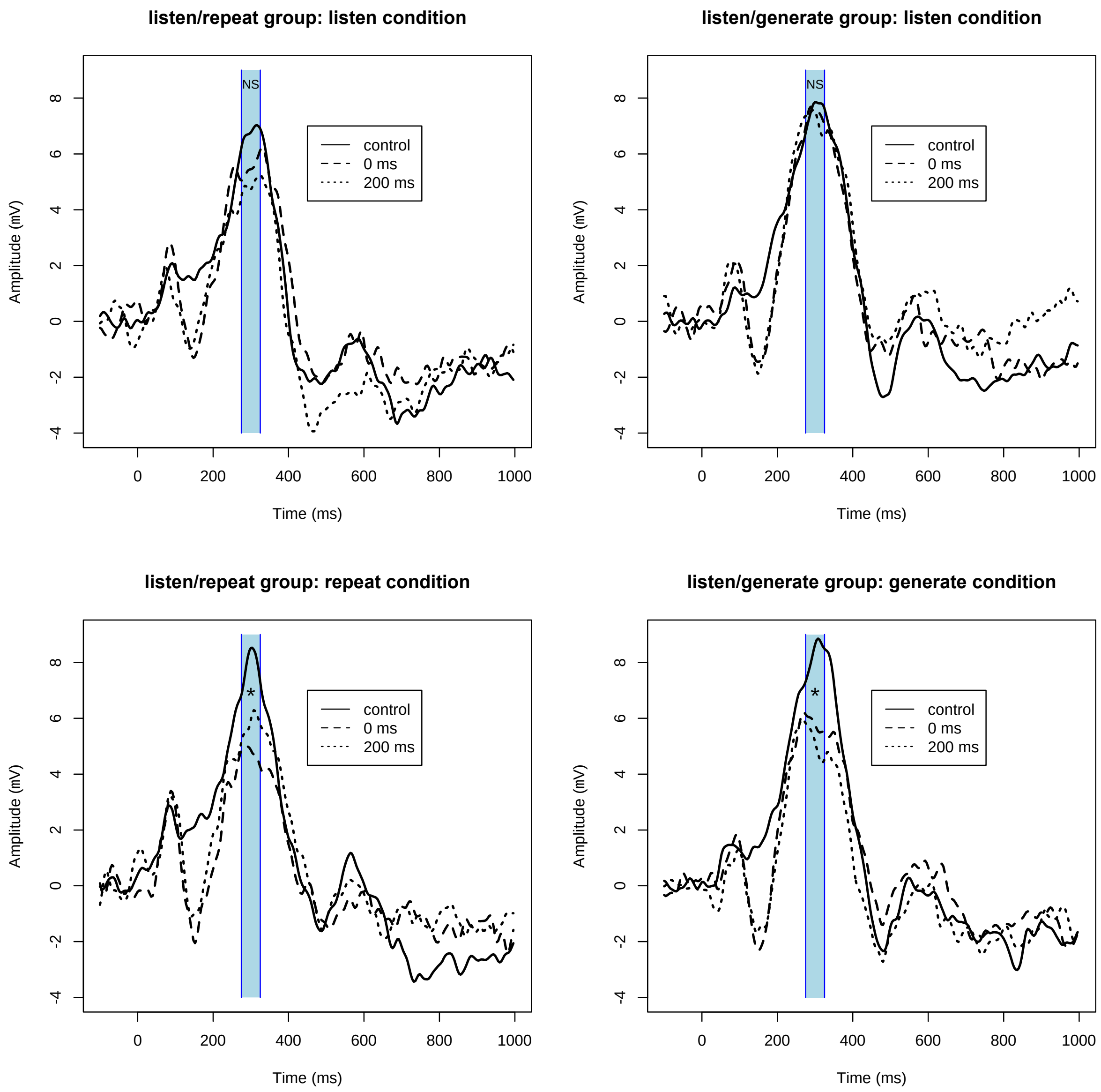

Figure 4: Difference waves (novel-standard) at electrode FCz for the listen-repeat group (left) and listen-generate group (right); for the listen condition (top row) or response conditions (bottom row: repeat / generate). The blue area highlights the area of which the mean is calculated and statistical analyses are performed. Different lines show different stimulus pairing of the sound probe and nouns (control, or 0ms, 200 ms). *Note:* NS stands for not-significant, and "*" for significant difference between the control condition compared to 0ms and 200ms stimulus pairing conditions.

The ANOVA revealed no effect of group, $F(1, 30) = 0.73$, $p = .40$. There was also no effect of task (listen versus repeat/generate), $F(1, 30) = 0.31$, $p = .58$. There was a main effect of stimulus pairing, $F(2, 29) = 4.72$, $p = .017$, $\eta_p^2 = .245$. This effect was also slightly influenced by an interaction between stimulus pairing and task, $F(2, 29) = 3.36$, $p = .049$, $\eta_p^2 = .188$. No other interaction effects were significant[3].

To test the nature of the small interaction effect between stimulus pairing and task, we conducted separate ANOVAs on the factor stimulus pairing for each of the two levels of task (listen versus repeat/generate). The factor group was also included in this ANOVA, but its effect was not interpreted (in both conditions there was no significant effect).

For the listen task condition, there was no effect of stimulus pairing, $F(2, 29) = 1.94$, $p = .162$. In contrast, in the task condition with a vocal response (repeat or generate), there was a main effect of stimulus pairing, $F(2, 29) = 6.16$, $p = .006$, $\eta_p^2 = .298$. The test of within-subject contrasts showed that there was both a linear effect, $F(1, 30) = 5.67$, $p = .024$, $\eta_p^2 = .159$, and a quadratic effect, $F(1, 30) = 5.36$, $p = .028$, $\eta_p^2 = .152$. The effect was such that in vocal response trials (repeat/generate), the fP3 was high in the control condition where there was no word preceding the oddball-probe ($M = 8.12$ µV; $SD = 3.73$ µV), but the amplitude of the fP3 was reduced when it was preceded in the word condition, independent of whether it was the 0 ms ($M = 5.27$ µV; $SD = 4.03$ µV) or 200 ms condition ($M = 5.45$ µV; $SD = 5.08$ µV). Figure 5 illustrates this effect. It suggests that fP3 was reduced when participants had to respond to a noun (repeat/generate with probe at 0 ms and 200 ms), but

[3] There were no other interaction effects: not between task and group ($F(1, 30) = 2.09$, $p = .158$), stimulus pairing and group ($F(2, 29) = 0.405$, $p = .671$), or between task, stimulus pairing, and group, $F(2, 29) = 0.693$, $p = .508$.

not when they either needed to listen to a sound (listen 0 ms, 200 ms) or when there was no word (control condition in listen and repeat/generate condition).

To specifically test for differences between repeat and generate in only the word conditions, the analysis was repeated for word conditions only. This again yielded non-significance for the Group contrast between generate versus listen and repeat versus listen ($F(1,30)=2.0$, $p=.16$). A similar result was obtained when additionally omitting the listen condition so that the generate and repeat conditions were directly compared (averaged across word conditions): $F(1,30)=0.01$, $p=.91$ (see Supplementary Materials S6)."

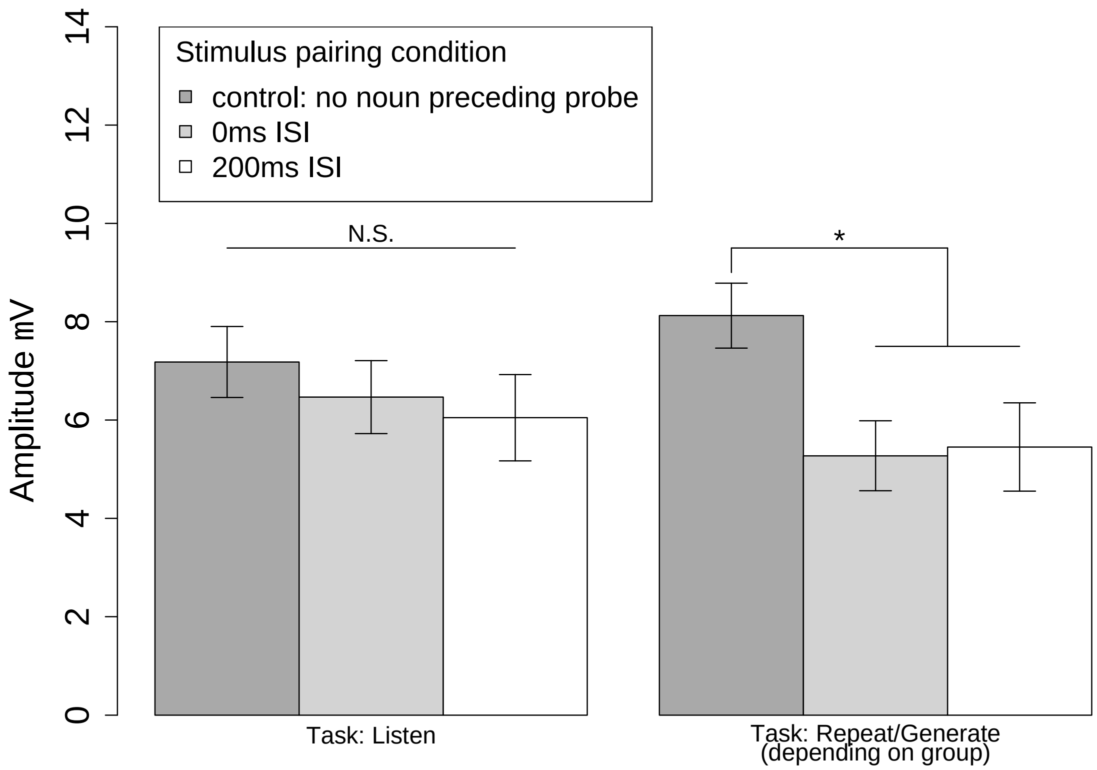

Figure 5: Bar plot of the interaction between task (listen versus repeat/generate) and timing of the sound probe (control, 0 ms, 200 ms). Error bars show standardized error of the mean. *Note:* NS stands for not-significant, and "*" for significant difference between the control condition compared to 0ms and 200ms stimulus pairing conditions.

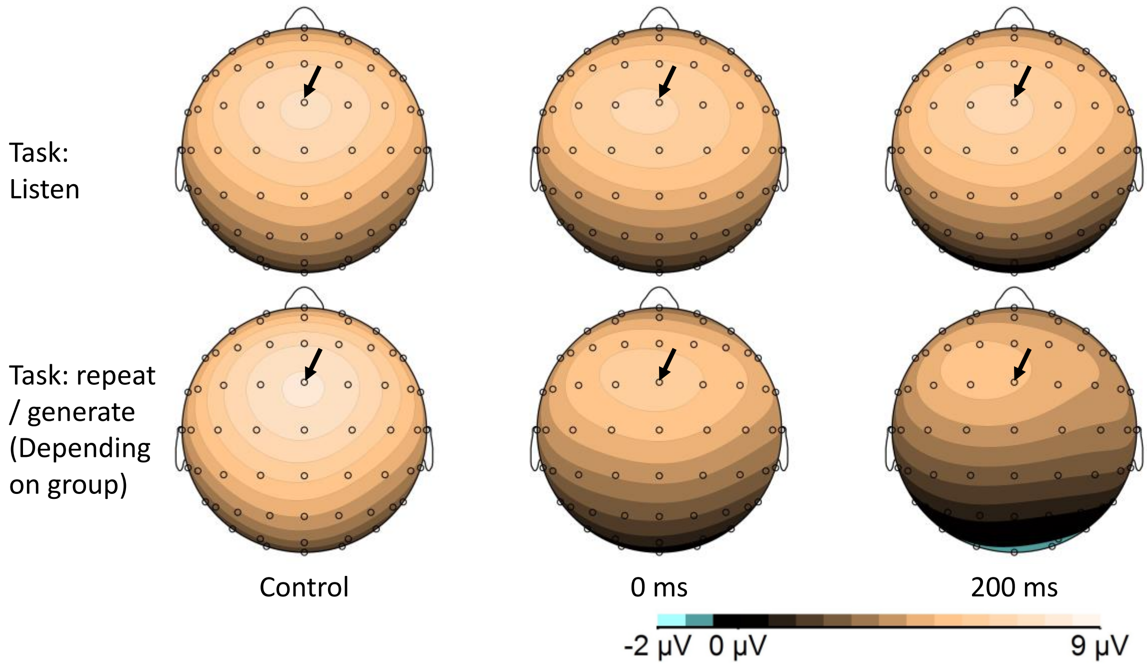

Figure 6: Scalp distribution of novel minus standard during time interval 275-325 ms (interval where fP3 occurs at FCz). Results were aggregated based on task (rows: listen or repeat/generate) and stimulus pairing (columns: control, 0 ms, 200 ms). *Note:* The arrow shows the location of the FCz electrode.

3.2 Additional control analysis 1: Novelty-Related Negativity (NRN) at 125-175 ms around electrode Pz, P1, and P2

One surprising aspect in Figure 4 is that there seems to be a negative dip before the fP3 peak, around 150 ms. This dip seems to occur only in word conditions (0ms and 200 ms; not in control). To further analyze this, we first explored scalp distributions to find out at which electrode and around which time interval this negative peak seemed strongest after averaging over all conditions where a word was presented. This exploration suggested electrode Pz, P1, and P2 had the strongest negative peak around 150 ms. Previous work by

Wester and colleagues (Wester, Verster, Volkerts, Böcker, Kenemans, 2010) also found a negative component at Pz at 150 ms, which they labelled novelty-related negativity.

For our analysis, we averaged the difference wave (novel-standard) value of electrodes Pz, P1, and P2. Figure 7 shows the resulting grand averages, with results averaged over groups (listen-repeat and listen-generate). The average value between 125 and 175 ms (highlighted in red in Figure 7; scalp distribution shown in Figure 8) was analyzed with a 2 (Group) x 2 (task: Listen versus repeat/generate) x 3 (stimulus pairing: control, 0, 200 ms) repeated measures ANOVA.

The analysis only revealed a main effect of stimulus pairing of the noun, $F(2, 29) = 13.84$, $p < .001$, $\eta_p^2 = .488$. The test of within-subject contrasts showed that there was both a linear effect, $F(1, 30) = 12.49$, $p = .001$, $\eta_p^2 = .294$, and a quadratic effect, $F(1, 30) = 9.80$, $p = .004$, $\eta_p^2 = .246$. In the control condition there was hardly any negativity (*M* = -0.09 μV, *SD* = 1.56 μV), whereas it was very pronounced in the conditions preceded by a word for the 0 ms (*M* = -1.61 μV, *SD* = 1.93 μV) and 200 ms conditions (*M* = -1.39μV, *SD* = 1.96μV). There were no other significant main effects or interaction effects, all *F*s < 1.

In summary, the additional analysis suggests that there was a novel related negativity when participants heard a noun, independent of whether they needed to listen to the noun, repeat the noun, or generate a verb. The effect was also independent of whether the time between word and probe was 0 ms or 200 ms.

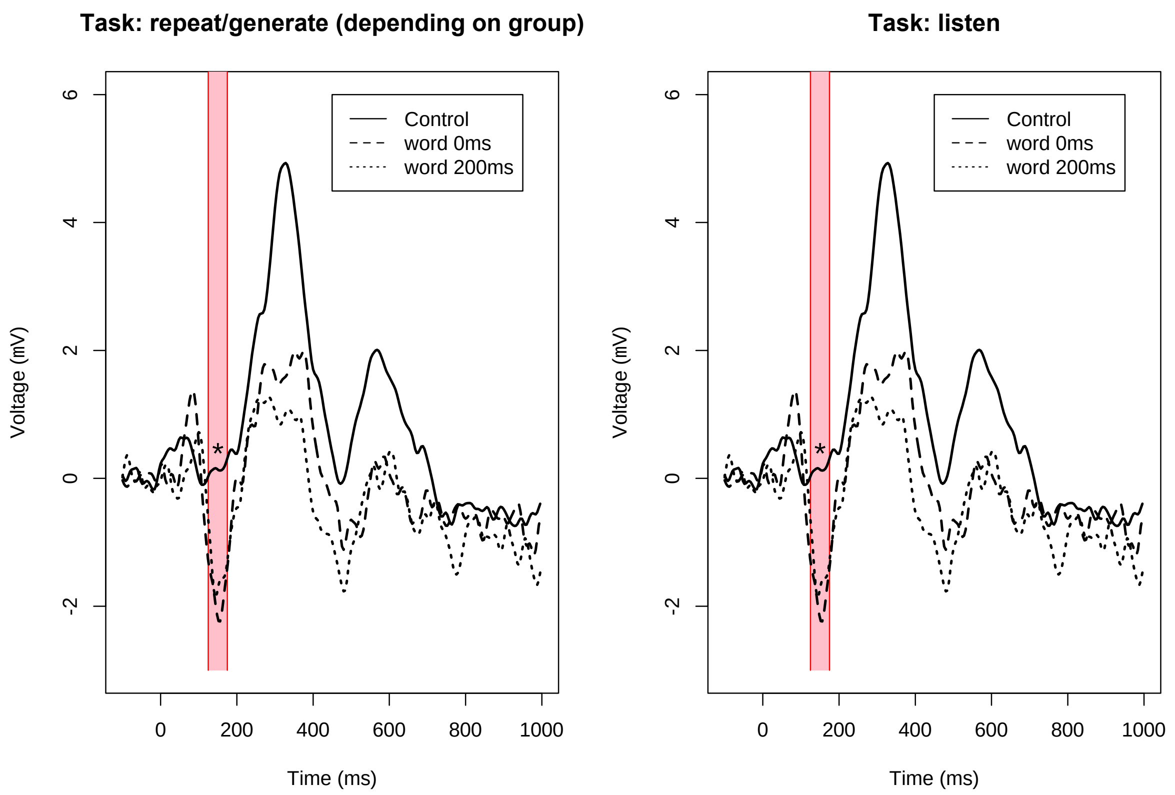

Figure 7: Grand average plot of novel-related negativity. Difference wave (novel-standard) is averaged over electrodes Pz, P1 and P2. Whenever a word was presented, there was novel-related negativity in the interval 125-175 ms. *Note:* NS stands for not-significant, and “*” for significant difference between the control condition compared to 0ms and 200ms stimulus pairing conditions.

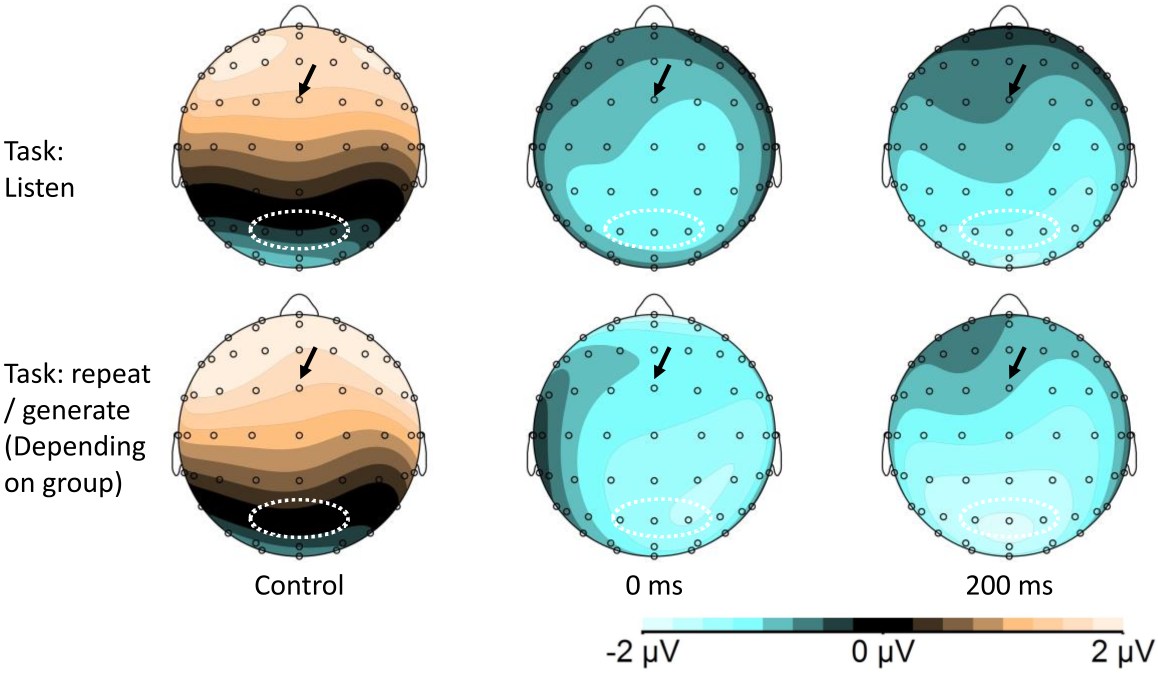

Figure 8: Scalp distribution of novel minus standard during time interval 124-172 ms (interval where NRN occurs at Pz, P1 and P2). Results were aggregated based on task (rows: listen or repeat/generate) and stimulus pairing (columns: control, 0 ms, 200 ms). *Note:* The arrow shows the location of the FCz electrode, and the dashed circle the location of Pz, P1, and P2.

3.3 Additional control analysis 2: Controlling fP3 results for the NRN

A potential concern was that the timing effects on fP3 might have been biased by the presence of the Novelty-Related Negativity (NRN). We therefore conducted additional control analyses to rule out this concern. Details are reported in Supplementary file S5.

In our analysis (see S5) we calculated mean amplitudes for the 275-325 ms interval for each point on the midline (Oz – Fz), when averaging over all conditions where stimulus pairing was preceded by a word (i.e., 0ms and 200 ms) versus when it was not preceded by a word (i.e., control). In both the word and the control/non-word condition, the amplitude peaks at FCz. However, the difference between the conditions with a word preceding a probe and the control trials without a word was largest more posteriorly, consistent with a contribution of the NRN.

To deal with this potential contamination, a variant of the vector filter method of Gratton, Coles and Donchin (1989), was applied (see S5 for details). Briefly, the currently used spatial vector filter capitalizes on a linear increase of the 'true' fP3 from Oz to FCz electrodes, effectively filtering out the contribution of a parietally maximal NRN. Figure 9 plots the mean amplitude of the data (solid lines) and the linearly filtered data (dashed lines). The result showed that in addition to separating the word (0 ms and 200 ms combined) versus non-word (control) probe conditions, the difference between these two lines was now also maximal at FCz.

When the filtered FCz data was further analysed using a 2 x 2 x 3 mixed-factorial ANOVA (group x task x stimulus pairing), the original findings were replicated. There was again a main effect of stimulus pairing, $F(2, 29) = 4.62$, $p = .018$. This effect was also again (less) slightly influenced by an interaction between stimulus pairing and task, $F(2, 29) = 4.35$, $p = .022$. Finally, again there were no other main effects and no other interaction effects.

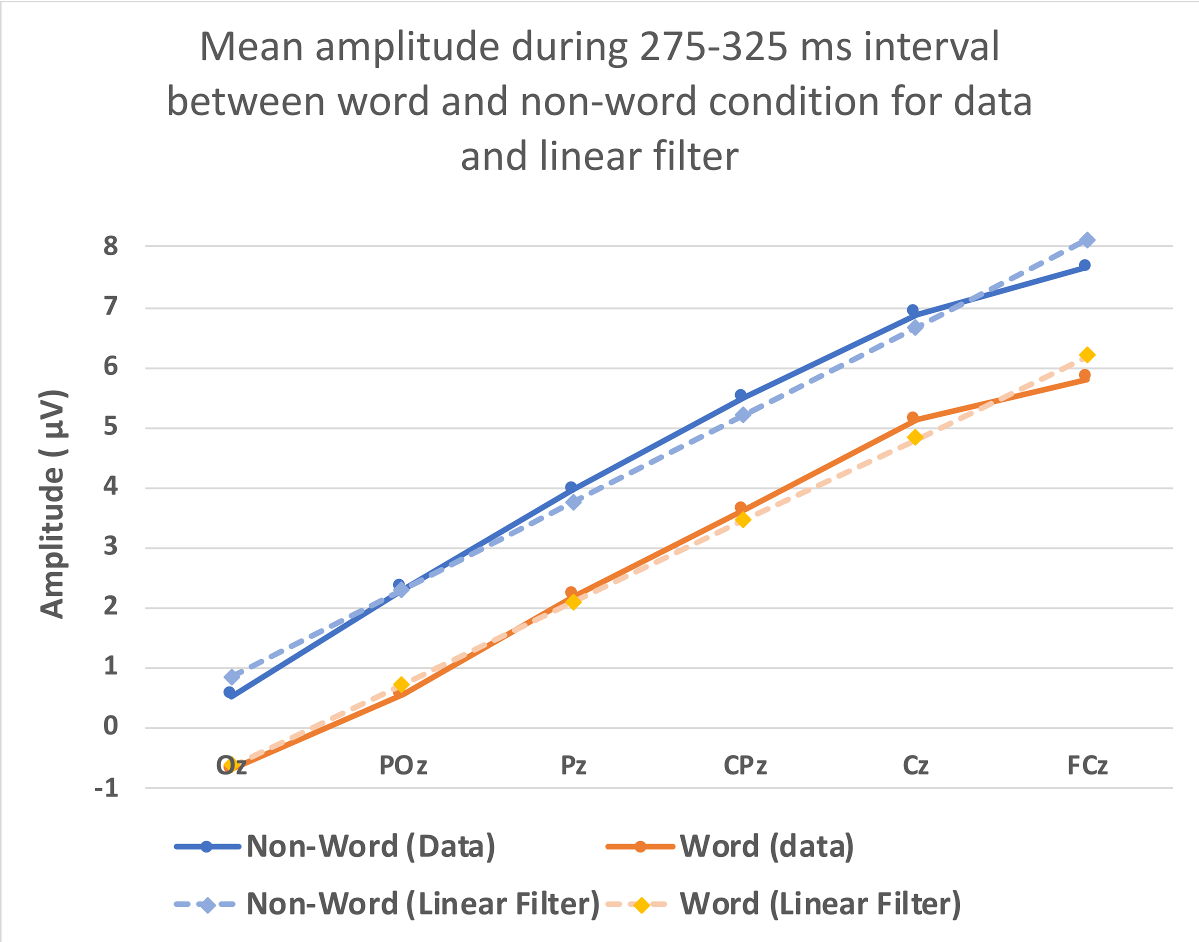

Figure 9: Mean amplitude of novel-standard data in sound probes preceded by a word and not by a word (non-word) across the midline.

4. Discussion

This study investigated how different levels of cognitive load affect susceptibility to novel sounds, as measured through the fP3 event-related potential. We compared fP3 response after participants heard a noun and either had to listen to it, repeat the noun, or generate a verb (i.e., the verb task; Abdullaev & Posner, 1998; Petersen et al., 1988; Snyder et al., 1995). These three conditions are hypothesized to create increasing cognitive load (cf. e.g., Iqbal et al., 2010; Kunar et al., 2008; Strayer & Johnston, 2001; van der Heiden et al., 2019). We found that fP3 is reduced in all conditions after a word is presented (listen, repeat, generate). It did not matter whether the interval between the noun and the sound probe was 0 or 200 ms (factor: stimulus-pairing). In other words, cognitive load seems to not affect fP3 response. As one small nuance, we also found an interaction effect between stimulus-pairing and whether participants had to listen to the stimuli or verbally respond (i.e., repeat/generate). This effect was such that there was a reduction in fP3 in the repeat/generate condition, but not in the listen condition. However, this interaction effect had a very small effect size, prompting replication before drawing more firm conclusions. One factor in the modesty of this interaction effect was noted by a reviewer: That during listening participants may have been anticipating repeating or generating in the subsequent block and adjusted their word processing accordingly, resulting in a relatively small difference between listening and repeating/ generating, that was however still significant.

The statement of fP3 being reduced in all word conditions should be qualified by the observation of an enhanced, more parietally distributed negativity in these same conditions (Novelty-Related Negativity or NRN). A possible concern in this respect was mitigated by applying a linear spatial filter to remove the possible contamination of the fP3 effects by the NRN effects (to be discussed in more detail at the end of this Discussion section). Furthermore, even without this kind of correction, not all variation in fP3 can be (partly) explained by variation in NRN, as the word effect on fP3 was modulated by the task condition whereas that on NRN was not.

We did not find a difference between repeat and generate, similar to Van der Heiden and colleagues (2022). As the Van der Heiden study also included an automated driving task, there it was not clear whether the lack of a difference between repeat and generate was due to some floor level that was reached in combination with driving. In our current study, there was no other task on top of the verb task, providing a purer estimate of the difference in susceptibility between repeat and generate conditions – in both it is reduced.

Our manipulation checks (see supplementary materials S3) suggest that our experiment did impact behavior as expected: verbal responses were slower in the generate condition compared to repeat, and the generate condition was subjectively rated as harder compared to the repeat condition (cf. e.g., Iqbal et al., 2010; Kunar et al., 2008; Strayer & Johnston, 2001; van der Heiden et al., 2019). Therefore, although we do see behavior differences between conditions, our conclusion is that fP3 amplitude, and associated susceptibility to novel sounds, is mostly affected by stimulus processing (i.e., hearing a word; supported by main effect of timing in analysis of fP3), and it has the *strongest* effect when participants need to verbally respond to the stimulus (i.e., repeat or generate; supported by subtle interaction between the effect of listen versus active response and that of timing).

These results can be considered in the context of alert processing in tasks such as responding to an alert in (automated) driving. Whereas previously we suggested that cognitive distraction of a (generate or repeat) task can reduce the susceptibility to novel sounds in semi-automated driving conditions (van der Heiden et al., 2022), the current work suggests that this can mostly be attributed to stimulus processing. This is consistent with earlier work where visual and visual-manual tasks (Scheer et al., 2016; 2018) including manual driving (van der Heiden et al., 2018; Wester et al., 2008; 2010) reduce the fP3 amplitude and associated susceptibility to novel sounds.

Our additional analysis (e.g., Figure 7) suggests that the fP3 response can be differentiated from activity in parietal areas of the brain that precede the fP3 response (around 150 ms). The earlier parietal response seems to reflect initial detection of a stimulus – as this negative peak was present equally in the listen, repeat, and generate conditions. The fP3 effect occurs more strongly when participants had to actively respond to the stimulus (repeat and generate). One of the co-authors has only observed this early parietal negative peak once before in a study by Wester et al. (2010), where participants had to perform a driving task while also occasionally responding to a visual task and being probed by oddball-probes. This effect was labelled as novelty-related negativity (NRN). In other studies in our lab both with driving tasks (van der Heiden et al., 2018) and versions of the verb task (van der Heiden et al., 2020; 2022) we did not detect the NRN, either because it was not present or not sufficiently prominent on the frontocentral sensors for which fP3 was analyzed. It may be then, that the current data are the first in which NRN is both prominent as such and sensitive to manipulations of task and/or stimuli.

One possibility is that the amplitude of the NRN depends on a combination of preceding stimuli related to a different task and of the inverse probability that these preceding stimuli are followed by task-unrelated but salient events (i.e., the novels). Whereas in the previous two Van der Heiden studies (van der Heiden et al., 2020; 2022) standards and novels were equally probable after a word, in the present study the novel followed a word considerably less frequently than standards did (see Table 1 for distribution per block). It could be that NRN reflects some interrupt function that is activated only in the context of on-going task-related stimuli and only by signals that are both unique and infrequent. For example, when you listen to the radio or to linguistic utterances otherwise, NRN may signal the occurrence of an unexpected but potentially relevant event, which may or may not be relevant in relation to yet another task (e.g., process control, or more specifically driving). Further studies are needed to test the interrupt hypothesis and to explore how NRN differs between single and dual-task situations (e.g., conversing and driving). Of note, a similar sequence of posterior negativity and frontal positivity has been reported for unexpectedly omitted stimuli in a regular series of task-relevant auditory stimuli with intervals of 1 or 2 seconds (Busse & Woldorff, 2003). It may be that in that paradigm a similar infrequent and therefore unexpected interruption of the task occurs as in the current setup.

Another possible interpretation is that our finding of a smaller fP3 under conditions with a preceding noun might reflect a more general sensory reduction or ERP overlap and distortion (Woldorff, 1993), for example, because participants started speaking early (which is very rare, see Supplementary materials S3). However, there are four reasons that reduce the likelihood of such effects. First, sensory reduction would affect ERPs equally in the Listen and in the Repeat/ Generate conditions, which is not the case in the present data. Second, we incorporated a baseline of standard-elicited ERPs which should at least partially correct for common sensory reduction, and especially for potential ERP overlap. Third, typical sensory-reduction phenomena such as PPI and P50 gating are usually observed on time scales exceeding 0 ms (interval between first and second stimulus, possibly because of the involvement of prefrontal neurons to set the gate (Oranje et al., 2006); in the present data however, there was no difference between 0- and 200-ms intervals. Fourth, in the present data there was in fact another ERP response that was only dependent on the presentation of the probe, not on the task condition (NRN); but much more than a result of sensory reduction, this seems to be an enhanced interrupt-like response to the quality of infrequent novels.

Future work can explore what happens if tasks are performed entirely cognitively, without any external stimuli, for example, by having participants count numbers subvocally. Does the observed reduction in susceptibility then still persist? If so, that would provide evidence that there is more than just stimulus processing that reduces susceptibility under load conditions.

Future work can also explore how reduced susceptibility impacts performance more broadly. Our manipulation checks (see supplementary materials S2 and S3) already demonstrated that the generate condition led to slower verbal responses compared to the repeat task, and that the generate task was subjectively experienced as being harder (cf. e.g., Iqbal et al., 2010; Kunar et al., 2008; Strayer & Johnston, 2001; van der Heiden et al., 2019). However, we did not test the correctness or quality of the verbal response (e.g., how semantically related a spoken verb was to the presented noun) or other metrics of performance.

Another open question is how a reduction in susceptibility can be counter-acted. Previous work suggests that, at least within the context of the oddball task, having an occasional active response requirement to a subset of the sound probes can boost susceptibility to novel sounds (e.g., van der Heiden et al., 2018; Wester et al., 2008). It can be tested whether such response requirements boost the susceptibility under control circumstances and under listening conditions.

In our analyses we had theory-informed hypotheses that could be tested with conventional hypothesis tests. However, an open question is whether other markers in the brain signal can predict the impact of cognitive load on susceptibility. Other techniques, such as machine learning, can be applied to these more exploratory analyses (e.g., Gogna, Tiwari, & Singla, 2024; Prasad, Tarai, & Bit, 2023).

A final note concerns the methodology that was applied here for dealing with the potential overlap between NRN and fP3. The current linear filtering is one instantiation of the more general vector-filter procedure, which can principally take any form. Other procedures such as Principal Component Analysis (PCA) and Independent Component Analysis (ICA) have been proposed for overlap issues. However, whereas PCA and ICA are completely bottom-up and multivariate, the present and earlier applied vector filters are a form of (univariate) hypothesis testing, which makes them more powerful (in statistical terms) to test specific hypotheses.

Especially the (polynome-derived) 0- to k-order trend vectors (k+1 being the number of, in this case, sensors, or features in terms of multivariate pattern analysis) may be considered a powerful vector-filter variety, as they are mutually orthogonal (independent) and orthonormal. They essentially function as a Fourier analysis, with linear corresponding to a half cycle, quadratic one full cycle, cubic 3/2 cycle, and so forth. The current application essentially boils down to a low-pass spatial filter, with only the linear component and the intercept (0-order or average) being retained in the data model. Note again that this was based on good grounds, as the data were tested for the robust presence of certain trends across subjects. The spatial vector filter can be easily tailored to other specific hypotheses, e.g. that the process of interest is distributed in a parabolic fashion across a sensor array and must be isolated from another process marked by a uniform posterior-anterior gradient; in that case it would be the quadratic component that should be isolated, rendering the model essentially equivalent to a band-pass spatial filter.

5. Conclusion

We conclude that cognitive load does not affect fP3-indexed susceptibility. Instead, fP3-indexed susceptibility is reduced when a word is presented (in listen, repeat, and generate conditions). This effect is the most pronounced in the conditions where a vocal response is needed (repeat and generate). No additional effect of cognitive load is present.

6. 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.

Busse L., & Woldorff, M.G. (2003). The ERP omitted stimulus response to "no-stim" events and its implications for fast-rate event-related fMRI designs. *Neuroimage*, 18(4):856-864.

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

Escera, C., Alho, K., Schröger, E., & Winkler, I. (2000). Involuntary Attention and Distractibility as Evaluated with Event-Related Brain Potentials. *Audiology and Neurotology*, *5*(3–4), 151–166.

Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural Mechanisms of Involuntary Attention to Acoustic Novelty and Change. *Journal of Cognitive Neuroscience*, *10*(5), 590–604.

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

Figalová, N., Pichen, J., Chandrayan, V., Pollatos, O., Chuang, L. L., & Baumann, M. (2023). Manipulating Drivers' Mental Workload: Neuroergonomic Evaluation of the Speed Regulation N-Back Task Using NASA-TLX and Auditory P3a. In *Adjunct Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 145–149. https://doi.org/10.1145/3581961.3609887

Figalová, N., Bieg, H. J., Reiser, J. E., Liu, Y. C., Baumann, M., Chuang, L., & Pollatos, O. (2024). From Driver to Supervisor: Comparing Cognitive Load and EEG-Based Attentional

Resource Allocation Across Automation Levels. *International Journal of Human-Computer Studies*, *182*, 103169. https://doi.org/10.1016/J.IJHCS.2023.103169

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. http://doi.org/10.1016/S0149-7634(01)00019-7

Gogna, Y., Tiwari, S., & Singla, R. (2024). Evaluating the performance of the cognitive workload model with subjective endorsement in addition to EEG. *Medical & Biological Engineering & Computing*, 1-18.

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*(4), 468–484. http://doi.org/10.1016/0013-4694(83)90135-9

Gratton, G., Coles, M. G., & Donchin, E. (1989). A procedure for using multi-electrode information in the analysis of components of the event-related potential: Vector filter. *Psychophysiology*, *26*(2), 222-232.

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.

Kenemans, J. L. (2015). Specific proactive and generic reactive inhibition. *Neuroscience and Biobehavioral Reviews*, *56*, 115–126. http://doi.org/10.1016/j.neubiorev.2015.06.011

Kunar, M. A., Carter, R., Cohen, M., & Horowitz, T. S. (2008). Telephone conversation impairs sustained visual attention via a central bottleneck. *Psychonomic Bulletin and Review*, *15*(6), 1135–1140. http://doi.org/10.3758/PBR.15.6.1135

Muller-Gass, A., & Schröger, E. (2007). Perceptual and cognitive task difficulty has differential effects on auditory distraction. *Brain Research*, *1136*(1), 169–177.

Oranje, B., Geyer, M. A., Bocker, K. B. E., Leon Kenemans, J., & Verbaten, M. N. (2006). Prepulse inhibition and P50 suppression: Commonalities and dissociations. *Psychiatry Research*, *143*(2–3), 147–158.

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. *The Journal of Neuroscience*, 29(26), 8604-8611. doi:10.1523/jneurosci.5967-08.2009

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, *331*, 585–589.

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

Prasad, R., Tarai, S., & Bit, A. (2023). Investigation of frequency components embedded in EEG recordings underlying neuronal mechanism of cognitive control and attentional functions. *Cognitive Neurodynamics*, *17*(5), 1321-1344.

Sabri, M., Liebenthal, E., Waldron, E. J., Medler, D. A., & Binder, J. R. (2006). Attentional Modulation in the Detection of Irrelevant Deviance: A Simultaneous ERP/fMRI Study. *Journal of Cognitive Neuroscience*, *18*(5), 689–700.

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*, 211. http://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. http://doi.org/10.1177/0018720818760919

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

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.

van der Heiden, R. M. A. (2020). *Auditory alerts and driving: Susceptibility to audio during (automated) driving*. [Doctoral dissertation, Utrecht University] https://dspace.library.uu.nl/handle/1874/396727

van der Heiden, R. M. A., Janssen, C. P., Donker, S. F., & Kenemans, J. L. (2020). The influence of cognitive load on susceptibility to audio. *Acta Psychologica*, *205*, 103058. http://doi.org/10.1016/j.actpsy.2020.103058

van der Heiden, R. M. A., 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. http://doi.org/10.1016/j.trf.2018.02.024

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

van der Heiden, R. M. A., Kenemans, J. L., Donker, S. F., & Janssen, C. P. (2022). The effect of cognitive load on auditory susceptibility during automated driving. *Human Factors*, 64 (7), 1195-1209

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

Volosin, M., & Horváth, J. (2020). Task difficulty modulates voluntary attention allocation, but not distraction in an auditory distraction paradigm. *Brain Research*, *1727*, 146565.

Wessel, J. R., & Aron, A. R. (2013). Unexpected Events Induce Motor Slowing via a Brain Mechanism for Action-Stopping with Global Suppressive Effects. *The Journal of Neuroscience*, 33(47), 18481-18491. doi:10.1523/jneurosci.3456-13.2013

Wester, A. E., Böcker, K., 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.

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.

Wiens, S., van Berlekom, E., Szychowska, M., & Eklund, R. (2019). Visual perceptual load does not affect the frequency mismatch negativity. *Frontiers in Psychology*, *10*(AUG), 461129.

Woldorff, M. G. (1993). Distortion of ERP averages due to overlap from temporally adjacent ERPs: Analysis and correction. *Psychophysiology*, *30*(1), 98–119.