

## P3 event-related brain potential reflects allocation and use of central processing capacity in language production



Natalia Shitova<sup>a,b,\*</sup>, Ardi Roelofs<sup>a</sup>, Caitlin Coughler<sup>a</sup>, Herbert Schriefers<sup>a</sup>

<sup>a</sup> Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

<sup>b</sup> International Max Planck Research School for Language Sciences, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

### ARTICLE INFO

#### Keywords:

Complexity  
ERP  
P3  
Phrase production  
Processing capacity  
Switching

### ABSTRACT

Allocation and use of central processing capacity have been associated with the P3 event-related brain potential amplitude in a large variety of non-linguistic tasks. However, little is known about the P3 in spoken language production. Moreover, the few studies that are available report opposing P3 effects when task complexity is manipulated. We investigated allocation and use of central processing capacity in a spoken phrase production task: Participants switched every second trial between describing pictures using noun phrases with one adjective (size only; simple condition, e.g., “the big desk”) or two adjectives (size and color; complex condition, e.g., “the big red desk”). Capacity allocation was manipulated by complexity, and capacity use by switching. Response time (RT) was longer for complex than for simple trials. Moreover, complexity and switching interacted: RTs were longer on switch than on repeat trials for simple phrases but shorter on switch than on repeat trials for complex phrases. P3 amplitude increased with complexity. Moreover, complexity and switching interacted: The complexity effect was larger on the switch trials than on the repeat trials. These results provide evidence that the allocation and use of central processing capacity in language production are differentially reflected in the P3 amplitude.

### 1. Introduction

Language production involves planning and monitoring, which does not happen fully automatically but requires some central processing capacity (e.g., Roelofs and Piai, 2011, for a recent review). It is generally assumed that central processing capacity is limited and needs to be selectively allocated to ongoing mental processes. The amount of capacity that is allocated depends on an evaluation of the circumstances and task demands (e.g., Kahneman, 1973). In planning a word, phrase, or sentence, speakers need to engage in conceptual, syntactic, morphological, phonological, and phonetic encoding processes (e.g., Levelt, 1989). How much central processing capacity is allocated to some of these processes (e.g., conceptual encoding) may depend on the complexity of the intended utterance. For example, planning the production of “a cup” would seem to require less processing capacity than “a big cup”, while planning to say “a big red cup” presumably requires even more capacity.

The allocation and use of central processing capacity have been studied extensively in non-linguistic domains by measuring event-related brain potentials (ERPs), thereby focusing on the P3 component (e.g., Kok, 2001; Polich, 2007; Polich and Kok, 1995). However, to our

knowledge, there exist only three studies that investigated central processing capacity in language production using ERPs (Habets et al., 2008; Marek et al., 2007; Sikora et al., 2016a). In all three studies, allocation and use of central processing capacity were taken to be modulated by phrase and sentence complexity, which was reflected in the P3 component. This corresponds to the effect of complexity on the P3 that has been observed for a variety of non-linguistic tasks (e.g., Brydges et al., 2014; Evans et al., 2011; Kramer et al., 1987; Kok, 2001; Strayer and Drews, 2007; Watter et al., 2001). However, the directionality of the P3 effect in language production differed between the studies of Habets et al. and Marek et al., on the one hand, and the study of Sikora et al., on the other.

In the ERP study of Habets et al. (2008), participants saw two sequentially presented pictures of objects (e.g., a car and a book), and they had to produce sentences describing two actions (e.g., driving and reading) involving these objects. A cue presented after the objects indicated whether the participants had to express the chronological order of object presentation (e.g., “After I drove the car, I read a book”) or the non-chronological order (e.g., “Before I read a book, I drove the car”). It was assumed that participants would allocate more central capacity to the planning processes in the non-chronological than in the

\* Correspondence to: Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, Radboud University, Nijmegen, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands.  
E-mail addresses: [n.shitova@donders.ru.nl](mailto:n.shitova@donders.ru.nl) (N. Shitova), [a.roelofs@donders.ru.nl](mailto:a.roelofs@donders.ru.nl) (A. Roelofs), [c.coughler@gmail.com](mailto:c.coughler@gmail.com) (C. Coughler), [h.schriefers@donders.ru.nl](mailto:h.schriefers@donders.ru.nl) (H. Schriefers).

chronological condition. In line with this assumption, in the ERPs time-locked to the presentation of the cue, the amplitude of the P3 component was higher in the non-chronological (i.e., complex) than in the chronological (i.e., simple) condition.

A similar effect on the amplitude of the P3 was observed by Marek et al. (2007). In their study, the participants were shown networks of colored shapes together with an arrow cue. They were asked to describe the direction of the arrow alone (e.g., “downwards”, the simple condition), the direction and shape (e.g., “downwards to the triangle”, the medium condition), or the direction, shape, and color (e.g., “downwards to the grey triangle”, the complex condition). Instructions on which type of phrase to produce were given before the network stimulus appeared on the screen. Marek et al. observed a higher P3 amplitude in the medium and complex conditions as compared to the simple condition. Thus, again, the ERP data suggest that speakers allocate more central capacity to the planning processes in more complex conditions.

However, an opposite effect of complexity on the P3 amplitude was observed in a study by Sikora et al. (2016a). Their participants had to produce noun phrases in response to pictures in simple (e.g., “the cup”) or complex (e.g., “the green cup”) conditions. Sikora et al. observed that the P3 amplitude was lower in the complex than in the simple condition, which is exactly the reverse of what Habets et al. (2008) and Marek et al. (2007) empirically obtained.

As suggested by Sikora et al. (2016a), one potential explanation of this discrepancy in results between studies was offered by Kok (2001). Based on an extensive literature review of P3 studies, Kok (2001) suggested that the P3 amplitude primarily reflects the amount of capacity allocated to “event-categorization”, which refers to perceptual-conceptual processes or stimulus identification. According to Kok, attention to stimulus processing (i.e., “task emphasis”) increases the P3 amplitude, while concurrent working-memory load (i.e., in dual-task performance) reduces the P3 amplitude. The amplitude is reduced because the working-memory load is thought to take away some of the capacity that is allocated to perceptual-conceptual processes. To map these factors to the present work, we refer to the capacity allocation aspect of Kok's view as “allocation of processing capacity” and the working-memory load aspect as the “use of processing capacity”. Importantly, the account by Kok suggests that changes in task difficulty per se do not predict the direction of the effect on the P3 amplitude. Task difficulty will increase P3 amplitude when participants allocate more capacity on complex than on simple trials and when capacity use does not counteract this difference in allocation. In contrast, task difficulty will decrease P3 amplitude when participants allocated more capacity on complex than on simple trials but use counteracts this difference in allocation. Correspondingly, Sikora et al. explained the difference in directionality of the P3 effects between studies in terms of a difference in how much capacity was allocated and how much capacity was used when performing the task. According to their proposal, the participants of Habets et al. (2008) and Marek et al. (2007) allocated substantially more central capacity to planning processing on complex trials than on simple trials, and this difference was preserved despite greater use of capacity on the complex trials. As a consequence, the P3 amplitude was higher on the complex trials than on the simple trials in these studies. In contrast, the participants of Sikora et al. allocated the same amount of capacity to processing on complex and simple trials but more capacity was used on complex trials, reducing the P3 amplitude. Alternatively, more capacity was allocated on complex than on simple trials, but capacity use counteracted this difference, also reducing the P3. To conclude, the difference in the directionality of the P3 effect between the studies of Habets et al. and Marek et al., on the one hand, and the study of Sikora et al., on the other, may have been due to a different balance between capacity allocation and use.

Such a different balance between capacity allocation and use may have been due to design differences between the studies. Overall, the task in the study of Sikora et al. (2016a) was simpler than the tasks in

the studies of Habets et al. (2008) and Marek et al. (2007). Perhaps, the participants of Sikora et al. therefore did not bother to allocate different amounts of capacity depending on the trial type. First, the utterances produced by the participants of Sikora et al. were relatively simple compared to the utterances produced in the studies of Habets et al. and Marek et al. Second, Sikora et al. employed only a very limited stimulus set (only four pictures) presented many times throughout the experiment. In contrast, in the studies of Marek et al. and Habets et al., there were more stimuli, which were more complex and repeated less often. If more capacity is allocated on complex trials than on simple trials, this may reduce the behavioral consequences of differences in complexity. In line with this, Sikora et al. observed an effect of complexity in response time (RT), whereas Habets et al. observed no significant difference in RT and error rate between the non-chronological (complex) and chronological (simple) conditions. Marek et al. did not report mean RTs, although the complexity effect was present in the error rates.

If the difference between studies in the directionality of the P3 effect is due to an overall difference in task complexity, then we should be able to replicate the P3 findings of Habets et al. (2008) and Marek et al. (2007) by increasing task complexity using the study design of Sikora et al. (2016a). Sikora et al. asked participants to respond to black-and-white or colored pictures by producing noun phrases with no adjectives (e.g., “the cup”, the simple condition) or with one adjective (e.g., “the green cup”, the complex condition), respectively. To increase overall task complexity in the present ERP study, we used small and large pictures and asked participants to produce noun phrases with one adjective to express size (e.g., “the small chair”, the simple condition) or with two adjectives expressing size and color (e.g., “the small red chair”, the complex condition). Furthermore, we increased the stimulus set size to a total of 16 pictures. If participants now allocate different amounts of capacity depending on the trial type, then the P3 amplitude should be larger in the complex than in the simple condition, as Habets et al. and Marek et al. observed. As in the study of Sikora et al., the type of phrase switched every second trial. Sikora et al. did not observe an effect of switching in the amplitude of the P3, whereas a number of studies on non-linguistic performance have observed such switching effect (see Sikora et al. for a review). For example, Lorist et al. (2000) observed that switching reduced the amplitude of the P3, suggesting greater use of central capacity on switch than on repeat trials. Thus, the switching manipulation in the present study may possibly index an effect of capacity use on the P3 (i.e., lowering the P3 amplitude on switch trials). Furthermore, switching has been observed to influence an earlier component, namely the N2 (e.g., Jackson et al., 2001; Sikora et al.), with the ERP deflecting more negatively on switch trials than on repeat trials. Therefore, an N2 switching effect would be expected in our alternating-runs experiment.

## 2. Methods

### 2.1. Participants

Twenty-three native Dutch speakers from the participant pool of Radboud University Nijmegen participated in this experiment for monetary compensation (16 female,  $M = 22.91$  years,  $SD = 3.19$  years, age range: 19–30). One participant's data were excluded from analysis because of excessive error rate: 28% of trials contained errors in naming. All participants reported having normal or corrected-to-normal vision, normal hearing, being non-dyslexic and having no speech impediments (i.e., stutter). The participants signed an informed consent form prior the experiment. The study followed the World Medical Association Declaration of Helsinki and it was performed within the line of research that was approved by the local ethics committee.

### 2.2. Materials

Sixteen line drawings of objects were selected from the picture

database of the Max Planck Institute for Psycholinguistics, Nijmegen. These pictures showed high name agreement in previous studies of our lab. The depicted objects belonged to four semantic categories: human body parts, animals, kitchenware and furniture (see the [Appendix A](#) for the stimulus list). Dutch nouns have grammatical gender, namely common or neuter, and definite articles are gender-marked. The article “de” is used for nouns with common gender and “het” for nouns with neuter gender. Objects were chosen so that the gender-marked determiners associated with the object names were balanced: eight object names were associated with the determiner “de” and the other eight with the determiner “het”. Pictures were manipulated by color (white, red, blue) as well as size (small, big). Small pictures were 4-by-4 cm and large pictures were 8-by-8 cm.

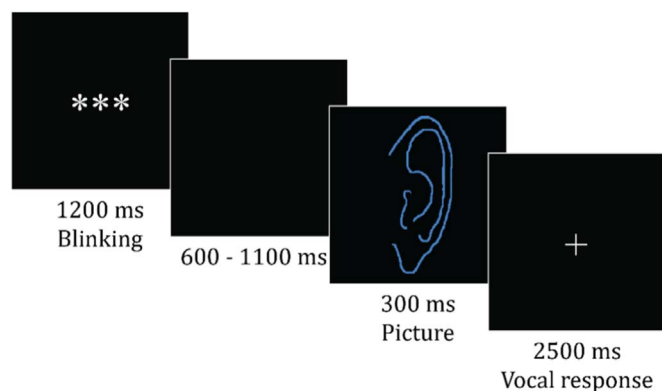
In the simple condition, a picture appeared on the screen in default (white) color in either big or small size. In the complex condition, a picture appeared in either red or blue color and in either big or small size. Each picture appeared 32 times in total throughout the experiment: 16 times either big (8 times) or small (8), and 16 times either big red (4), big blue (4), small red (4), or small blue (4).

To control for trial-by-trial effects, a pseudo-randomization technique was applied using the Mix software ([Van Casteren and Davis, 2006](#)). Nouns were separated by a minimum distance of 10 trials, no initial phoneme repetitions of nouns were permitted, determiners could be repeated four trials in a row maximally, and there was a minimum distance of three trials separating nouns of the same semantic category. All participants received a unique stimulus list.

### 2.3. Design and procedure

Stimulus presentation and recording of behavioral data were implemented with Presentation Software (Neurobehavioral Systems, Albany, CA). During the experiment, participants were seated in a regularly lit sound-attenuating room, approximately 75 cm away from the screen. All vocal responses were recorded through Presentation, later to be used for error coding and response-time registration.

At the beginning of the experiment, participants were familiarized with the pictures and corresponding nouns and determiners. An experimental session consisted of eight blocks of 64 trials each. A trial began with presentation of stars centrally for 1200 ms, during which a participant was instructed to blink. This was followed by presentation of a blank (black) screen for  $850 \text{ ms} \pm$  a random jitter of up to 250 ms. Then a picture was presented on the screen for 300 ms (see [Fig. 1](#) for the trial events). Participants were instructed to name this picture in the format “determiner + adjective(s) + noun” as quickly and accurately as possible, but not until they had planned out everything they would like to say. The picture presentation was followed by a fixation cross,



**Fig. 1. Trial events.** After a blinking interval and a black screen, pictures of different size and color are presented for 300 ms, and participants have to respond within 2500 ms after picture offset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

presented centrally for 2500 ms, during which participants made their vocal response. Before testing, participants were given a training session consisting of ten trials with a longer time to respond (3500 ms), followed by five trials with the regular response interval (2500 ms). Following this mandatory training, participants were given the choice of practicing further (2 blocks of five trials each), after which testing began.

The trials were presented in alternating runs ([Rogers and Monsell, 1995](#)), and the participants were informed about the sequence before the start of the experiment. Each block began with two trials of the simple condition (determiner + size adjective + noun) followed by two trials of the complex condition (determiner + size adjective + color adjective + noun). This pattern continued throughout a block of trials. On *switch* trials, the complexity level was different compared to the previous trial (i.e., a simple trial following a complex trial or a complex trial following a simple trial), whereas on *repeat* trials, the complexity level was the same between trials (i.e., a simple trial following a simple trial or a complex trial following a complex trial).

### 2.4. EEG data acquisition

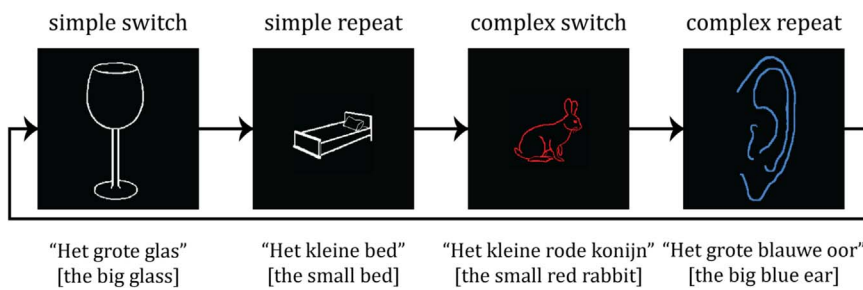
EEG data were recorded from 32 active electrodes (ActiCAP 32Ch Standard-2, Brain Products) at a sampling rate of 1000 Hz. Data were referenced online to FCz, and then re-referenced offline to the average of the left and right mastoids. Six additional active electrodes were used in order to record eye movements and blinks (vertical and horizontal EOGs), as well as movements of the mouth (EMG). Vertical EOG was recorded from below and above the orbital rim of the left eye. Horizontal EOG was recorded from the lateral orbital rim of the left and right eye. EMG measurements were taken one above the right side of the mouth (right orbicularis oris superior), and one below the left side of the mouth (left orbicularis oris inferior). The difference of the signals recorded from these pairs of electrodes (i.e., EOG and EMG) was used for artifacts rejection. All EEG signals were online filtered with the low cutoff of .016 Hz and the high cutoff of 125 Hz.

### 2.5. Analysis

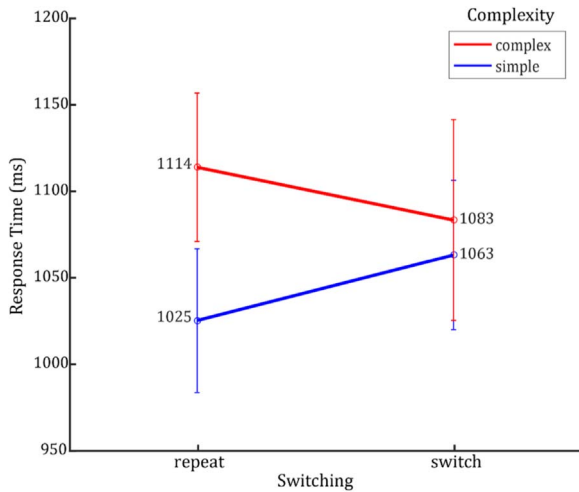
Trials that contained errors in naming were excluded from analysis. Errors concerned the production of an incorrect determiner, adjective or noun, or an incorrect ordering of the adjectives (size had to be expressed before color). Trials with missing determiners were not considered to be errors and were kept for analysis. After an experimental session, the recordings of the utterances were used to determine the RTs. RTs were measured semi-automatically using PRAAT ([Boersma, 2002](#)).

EEG data were analyzed using FieldTrip ([Oostenveld et al., 2011](#)) and custom analysis scripts, run through Matlab v. 8.1.0.604 (R2013a, The MathWorks, Inc.). Trials were defined to start 500 ms before the onset of a picture stimulus and finish at speech onset. To account for muscular artifacts that contaminate the EEG data close to articulation onset, trials were inspected visually and cut off at the onset of such preparatory muscular activity, usually approximately 150 ms before the speech onset. Since the average RT for the simple phrases was expected to be at least 800 ms (based on [Bürki and Laganaro, 2014](#); [Sikora et al., 2016a](#)), such approach would yield sufficient data to analyze the P3 component. Blinks and other eye movements were rejected manually through trial-to-trial visual inspection of vertical and horizontal EOG.

Artifact-free data were then low-pass filtered at 30 Hz and baseline-corrected per trial using a pre-stimulus baseline window of 300 ms, following previous studies that employed per-trial baseline correction in the alternating-runs paradigm (e.g., [Capizzi et al., 2016](#); [Hung et al., 2016](#); [Kamijo and Takeda, 2010](#); [Sikora et al., 2016a](#); [Swainson et al., 2003](#); [Fig. 2](#)). In order to obtain ERPs, the data were then averaged across trials per condition per participant. These averaged waveforms were analyzed using within-subjects cluster-based permutation test ([Maris and Oostenveld, 2007](#)) to assess statistical significance between conditions.



**Fig. 2. Illustration of the trial sequence.** Following two simple trials (at which only size had to be specified in the utterance), two complex trials were presented (at which size and color had to be specified), which was followed by two simple trials, and so forth. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



**Fig. 3. Behavioral results.** Mean response time as a function of complexity (simple, complex) and switching (repeat, switch). The error bars indicate the standard error of the mean per condition.

### 3. Results

#### 3.1. Behavioral results

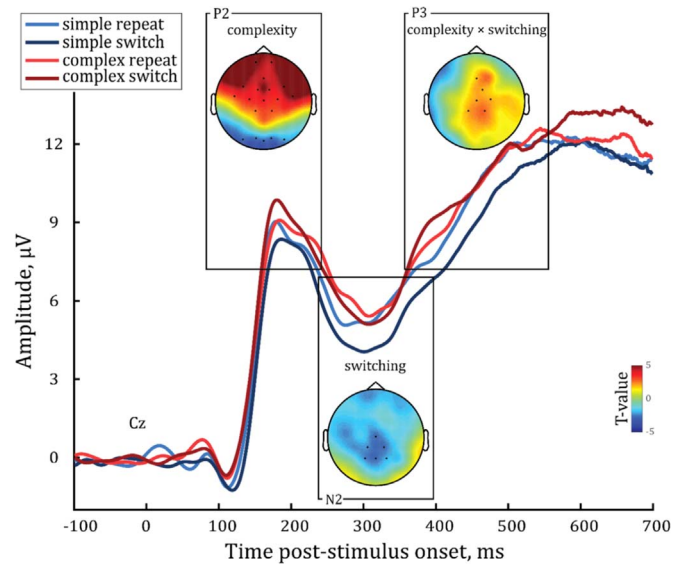
Group-averaged RTs on trials manipulated by complexity and switching are displayed in Fig. 3. Participants were slower on complex trials than on simple trials ( $F(1,21) = 43.2, p < .01$ ), while there was no main effect of switching ( $F(1,21) = 1.87, p = .19$ ). Complexity interacted with switching ( $F(1,21) = 28.83, p < .01$ ). Naming on complex switch trials was faster than on complex repeat trials ( $t(21) = -4.42, p < .01$ ), while naming on simple switch trials was slower than on simple repeat trials ( $t(21) = 5.47, p < .01$ ).

Error rates for the simple repeat, simple switch, complex repeat, and complex switch trials were 2.4%, 2.1%, 3.3%, and 2.5%, respectively. The analysis of error rates revealed an only marginally significant effect of complexity ( $F(1,21) = 3.37, p = .07$ ), while there was no main effect of switching ( $F(1,21) = 2.27, p = .15$ ) or interaction between complexity and switching ( $F(1,21) = .63, p = .44$ ).

#### 3.2. ERP results

Group-averaged ERPs for the vertex electrode Cz are displayed in Fig. 4 (for group-averaged ERPs over nine electrode sites, see Appendix B). Differential time-locked EEG activity was present in three time-windows, roughly corresponding to three standard ERP components: P2 (from 150 to 250 ms post-stimulus onset), N2 (250–350 ms), and P3 (350–500 ms).

During the P2 time-window, ERP waveforms on complex trials deflected more positively than those on simple trials over a broad range of fronto-central electrode sites ( $p = .002$ ) and more negatively over occipital electrode sites ( $p = .018$ ). ERPs on switch and repeat trials did not differ. Complexity and switching interacted ( $p = .02$ ), with only



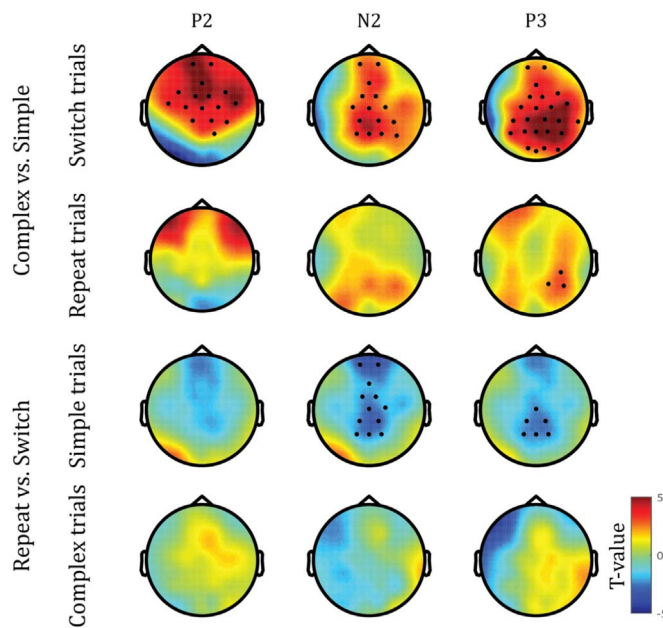
**Fig. 4. ERPs of simple repeat, simple switch, complex repeat, and complex switch trials.** Topographies represent T-values of the cluster-based permutation tests performed on every comparison within the following time-windows: 150–250 ms post-stimulus onset for the P2; 250–350 ms post-stimulus onset for the N2; 350–500 ms post-stimulus onset for the P3. The complexity effect is the difference between complex and simple trials. The switching effect is the difference between switch and repeat trials. The interaction effect is a difference between complexity effects in switch and in repeat trials. Electrode sites that entered spatio-temporal clusters based on which the null hypothesis was rejected are marked.

switch trials showing a complexity effect ( $p < .001$ ).

During the N2 time-window, the main effect of complexity persisted ( $p = .005$ ). Moreover, ERPs on switch trials deviated more negatively than those on repeat trials ( $p = .036$ ). There was no interaction of complexity and switching.

During the P3 time-window, the complexity effect was present over centro-parietal electrode sites ( $p < .001$ ), while no switching effect was obtained. However, an interaction of complexity and switching was observed ( $p = .037$ ). Although complex trials differed from simple trials on both switch ( $p < .001$ ) and repeat ( $p = .042$ ) trials, the complexity effect on the switch trials was larger than on the repeat trials. The switching effect was only present on the simple trials ( $p = .049$ ). The topographies of the simple ERP effects are shown in Fig. 5.

As we described in the Methods section, we performed per-trial baseline-correction of the data before studying differences in ERPs. This baseline correction may be problematic when participants are certain about the upcoming stimulus type. Per-trial subtraction of averaged pre-stimulus activity indeed led to an elimination of effects of preparation in the current experiment. However, if such correction is not performed before computing ERPs, it becomes impossible to assess whether any observed effects are real or due to random fluctuations in sustained EEG activity levels across trials. Therefore, at expense of not being able to investigate preparatory effects, we chose to baseline-correct the data. Furthermore, baseline-correction is a standard practice



**Fig. 5. Topographies of simple ERP effects.** Topographies represent T-values of the cluster-based permutation tests performed on every comparison. Motivated by the literature, positive-sided tests were performed on complexity effects in the P3 time-window, and negative-sided tests were performed on switching effects in the N2 and P3 time-windows. Otherwise, two-sided cluster-based permutation tests were performed on all simple effects. Electrode sites that entered spatio-temporal clusters based on which the null hypothesis was rejected are marked.

in research employing the alternating runs paradigm, including the study by Sikora et al. (2016a), which we used as a critical comparison for our experiment. Nevertheless, when we re-analyzed the current data without baseline-correction, the critical comparisons yielded the same results as the previously reported tests performed on the baseline-corrected data. The complex trials deflected more positively than the simple trials within the P3 time-window ( $p = .002$ ) and *complexity* and *switching* interacted within the P3 time-window ( $p = .004$ ), with only switch-trials showing significant complexity effect ( $p = .003$ ), while no difference was observed between complex-repeat and simple-repeat trials.

#### 4. Discussion

In previous ERP research on the allocation and use of processing capacity in language production, the directionality of the P3 effect differed between studies. Whereas Habets et al. (2008) and Marek et al. (2007) obtained a higher P3 amplitude for complex trials than for simple trials, Sikora et al. (2016a) observed the reverse. We hypothesized that the difference in directionality of the P3 effect may have been due to a difference in overall task complexity between studies. We tested this by increasing task complexity using the study design of Sikora et al. As in the study of Sikora et al., participants had to switch between utterance types, which may reduce the P3 amplitude, as has been shown in non-linguistic studies. The mean RTs were much longer in the present study (around 1070 ms) than in the study of Sikora et al. (around 700 ms), which suggests that our attempt to increase the overall complexity of the task was successful. In the ERPs, we observed a larger P3 amplitude on complex than on simple trials, as obtained by Habets et al. (2008) and Marek et al. (2007) but different from Sikora et al. This suggests that participants in our study allocated more capacity to the planning processes on complex than on simple trials. Also different from Sikora et al., we observed an interaction of complexity

and switching, with the complexity effect being larger on switch trials than on repeat trials. An RT switch cost was present only on simple trials. This would suggest that more capacity is used in switching to simple than to complex trials, in line with the interaction between complexity and switching in the P3 amplitude.

Following Sikora et al. (2016b), we take central processing capacity to determine “how well a speaker keeps in mind the requirements of the task (e.g., to be fast and accurate) while engaging in conceptual and linguistic processes” (p. 1720). Sikora et al. took processing capacity to concern the updating component of executive control. According to an influential theory of Miyake et al. (2000), three main components of executive control are updating, inhibiting, and shifting. Sikora et al. assumed that conceptual and linguistic processes critically require the updating of working memory. Long and short noun phrases differ in the extent of conceptual and linguistic processing and the corresponding demand on the updating of working memory (see Sikora et al. for discussion).

Several conceptual and linguistic processes underlie the production of noun phrases (e.g., Levelt, 1989). For example, in response to a picture of a big blue ear (Fig. 1), participants have to conceptually identify the object as well as its size and color, and the corresponding concept representations need to be temporarily maintained in working memory. Moreover, participants have to retrieve from long-term memory the lemmas of the corresponding Dutch noun (*oor*), size adjective (*groot*), color adjective (*blauw*), and gender-marked definite article (*het*). Using these lemmas, a syntactic structure for the noun phrase has to be generated, which includes serial ordering of the determiner, size adjective, color adjective, and noun. The syntactic structure has to be maintained temporarily. Moreover, a morpho-phonological representation has to be generated and maintained, which includes retrieving the appropriate morphemes and inflecting the adjectives (e.g., *grote*; morphological encoding), retrieving the phonemes and syllabifying them (phonological encoding), and accessing articulatory programs for the syllables (phonetic encoding), followed by the initiation of articulation (for “het grote blauwe oor”). These conceptual and linguistic processes and their time course have been extensively investigated in previous RT experiments (e.g., Meyer, 1996; Schriefers, 1992) and more recently in electrophysiological experiments (e.g., Bürki and Laganaro, 2014; Bürki et al., 2016; Eulitz et al., 2000; Michel Lange et al., 2015; Pykkänen et al., 2014). However, none of these earlier studies examined influences of capacity demands on the P3 (i.e., analyses of the P3 were not reported), which was central to the present experiment. In our experiment, conceptual and linguistic processing was presumably more extensive for the long than for the short noun phrases. As a consequence, keeping in mind the requirements of the task while engaging in conceptual and linguistic processes was more demanding for the long than for the short phrases, leading to differential allocation of central processing capacity, as reflected in the amplitude of the P3.

We obtained effects of our complexity and switching manipulations in the P2 (from 150 to 250 ms post-stimulus onset), N2 (250–350 ms), and P3 (350–500 ms). It is difficult to relate these ERP effects (or corresponding time windows) to specific conceptual and linguistic processes. In a meta-analysis of neuroimaging studies on word production, Indefrey and Levelt (2004; Indefrey, 2011) estimated that perceptual and conceptual processes in picture naming (i.e., producing a single noun) are completed around 200 ms after picture onset, while linguistic planning processes (i.e., lemma retrieval and morpho-phonological encoding up to phonetic encoding) are completed around 145 ms before articulation onset. These estimates hold for a mean picture naming RT of 600 ms. However, in the present experiment, the conceptual and linguistic processes concerned the planning of a phrase containing three or four words (i.e., a determiner, one or two adjectives, and a noun)

rather than a single noun. The mean RT in the present experiment was around 1070 ms. Clearly, the estimates of Indefrey and Levelt for single word production cannot be directly applied to noun phrase production. At the very least, the estimates have to be rescaled (Roelofs and Shitova, 2017). Under proportional rescaling (i.e.,  $1070/600 \times 200$  ms), the estimate for the completion of perceptual and conceptual processes in the present experiment would be about 360 ms. This would imply that the observed effects in the P2 and N2 occurred during perceptual and conceptual encoding, whereas the effects in the P3 arose during linguistic planning processes.

In previous ERP studies of picture naming, Strijkers et al. (2010) and Strijkers et al. (2011) argued that the P2 reflects the “intention to speak”. The time windows of the P2 in these studies were 160–240 ms and 140–210 ms, and the mean picture naming RTs were about 720 ms and 800 ms, respectively. Under proportional scaling of the estimates of Indefrey and Levelt (2004; Indefrey, 2011), completion of perceptual and conceptual processing was, respectively, around 240 ms and 266 ms in these studies. This would suggest that the P2 effects in the studies of Strijkers et al. occurred during perceptual and conceptual processing.

In our study, complexity effects were present in the P2, N2, and P3 components of the ERP. This is not surprising, since complex and simple trials differed at all stages of stimulus processing from visual perception (colored vs. non-colored stimuli) to response planning (expressing size and color vs. size only). A switching effect was only present in the N2 component. Moreover, complexity and switching interacted in the P2 and P3 components.

The P2 component has been associated with late visual processing modulated by salience and attention (see Luck and Hillyard, 1994, for a color pop-out P2 effect). Thus, a P2 complexity effect in our paradigm is expected: Colored stimuli were more salient events in a stream of colored and non-colored stimuli. Moreover, the color of the colored stimuli was relevant for the response, and therefore the colored stimuli triggered a larger P2 component. Interestingly, this complexity effect was present on switch trials but not on repeat trials. This can be explained by the fact that the visual pattern of the stimuli was changed on switch trials but remained the same on repeat trials. There were no early visual P1 effects associated with our color manipulation. Martinovic et al. (2008) showed that such effects occur in comparisons of objects pictured in prototypical or non-prototypical colors, as well as if objects differ in amount of surface detail and linear complexity. The line drawings that we employed in the current study for complex and simple conditions were not manipulated in these dimensions and, therefore, did not differ in their ease of recognition, which explains absence of the P1 effect. Although no canonical N1 effect was observed in the current study, there was a negative-going complexity effect over the occipital-most electrode sites that temporally overlapped with the frontal-central-parietal P2 complexity effect. In particular, the P2 on Cz corresponds to an N170 on posterior sites. We have marked out the relevant time window on posterior channels in the figure in Appendix B. Therefore, the observed P2-time-window effect can be more broadly considered as an N1-P2 complex effect. This would not change the original interpretation of the P2 effect in terms of perceptual processing regulated by attention (see Proverbio et al., 2004, for an N1 effect in color manipulation associated with attention).

The N2 component has been associated with inhibition and strategic

monitoring (Folstein and Van Petten, 2008). In task switching, the amplitude of the N2 tends to be larger on switch than repeat trials (e.g., Jackson et al., 2001). In our study, the ERPs were also more negative-going on switch trials than on repeat trials, replicating the pattern reported by Sikora et al. (2016a). The authors also observed an interaction of complexity and switching in the N2 amplitude, with a switch effect only being present on simple trials. Sikora et al. argued that colored pictures require long-phrase responses (e.g., “the red cup”) but afford a short-phrase response as well (e.g., “the cup”), and therefore the task set for short phrases needs to be inhibited. As a consequence, later switching to a short-phrase response requires overcoming the previous inhibition, which increases the N2. There was no interaction of complexity and switching in the N2 time-window in the current data set, but an additional post-hoc analysis showed that a switch effect was present on simple trials ( $p = .017$ ), but not on complex trials. According to Sikora et al., overcoming the inhibition on short switch trials explains the switch cost in RTs on the short trials. However, in our RTs, both a switch cost on the simple trials and a switch benefit on the complex trials were present. One possible explanation of the switch benefit is that this effect might be due to additional lexical competition with a different color adjective produced on the previous trial, which may have occurred on a proportion of the complex repeat trials (e.g., “the big blue ear” following “the small red desk”). However, a post-hoc analysis of complex repeat trials showed no RT difference between trials that repeated the color adjective and trials that did not. Therefore, the switch benefit effect in RTs remains to be explored further in future research.

It may also be argued that the asymmetrical switch cost is due to a potential need to suppress the adjective *white* when responding to the simple stimulus and particularly after using a color adjective on the previous complex trial. However, we believe that this was not the case. First, if such suppression had to be performed on every simple-switch trial, then there should have been considerable numbers of errors that include the word *white*, which we did not observe. Second, the suppression should have led to increased RTs for simple trials compared to complex trials, on which the participant did not have to suppress the color name. However, we observed that mean RTs were about 55 ms smaller on simple than on complex trials.

Complexity and switching interacted in the P3 time-window: The P3 effect of complexity was larger on switch trials than on repeat trials. The RTs suggest that processing on the short switch and the complex repeat trials was more demanding than processing on their switching counterparts (i.e., the short repeat and the complex switch trials). This relative increase in RTs parallels the relative decrease in P3 amplitude on these trials, suggesting that additional use of processing capacity decreased the P3 amplitude. Therefore, although in general complex trials were associated with a higher P3 amplitude than simple trials, reflecting a difference in allocation of processing capacity, the greater use of capacity on short switch and complex repeat trials reduced the P3. This differential allocation and use of processing capacity explains the interaction of complexity and switching that we observed in the P3 time-window.

To conclude, we observed that allocation and use of processing capacity differentially influence the P3 amplitude in phrase production. Thus, our study demonstrated both directionalities of the P3 effect proposed by Kok (2001).

Appendix A

| Category    | Determiner | Noun            |
|-------------|------------|-----------------|
| Animals     | het        | hert (deer)     |
|             | het        | konijn (rabbit) |
|             | de         | geit (goat)     |
| Kitchenware | de         | zwaan (swan)    |
|             | de         | fles (bottle)   |
|             | de         | kan (pitcher)   |
|             | het        | glas (glass)    |
| Furniture   | het        | bord (plate)    |
|             | de         | stoel (chair)   |
|             | de         | kast (cupboard) |
|             | het        | bed (bed)       |
| Body parts  | het        | bureau (desk)   |
|             | de         | arm (arm)       |
|             | de         | neus (nose)     |
|             | het        | been (leg)      |
|             | het        | oor (ear)       |

Appendix B

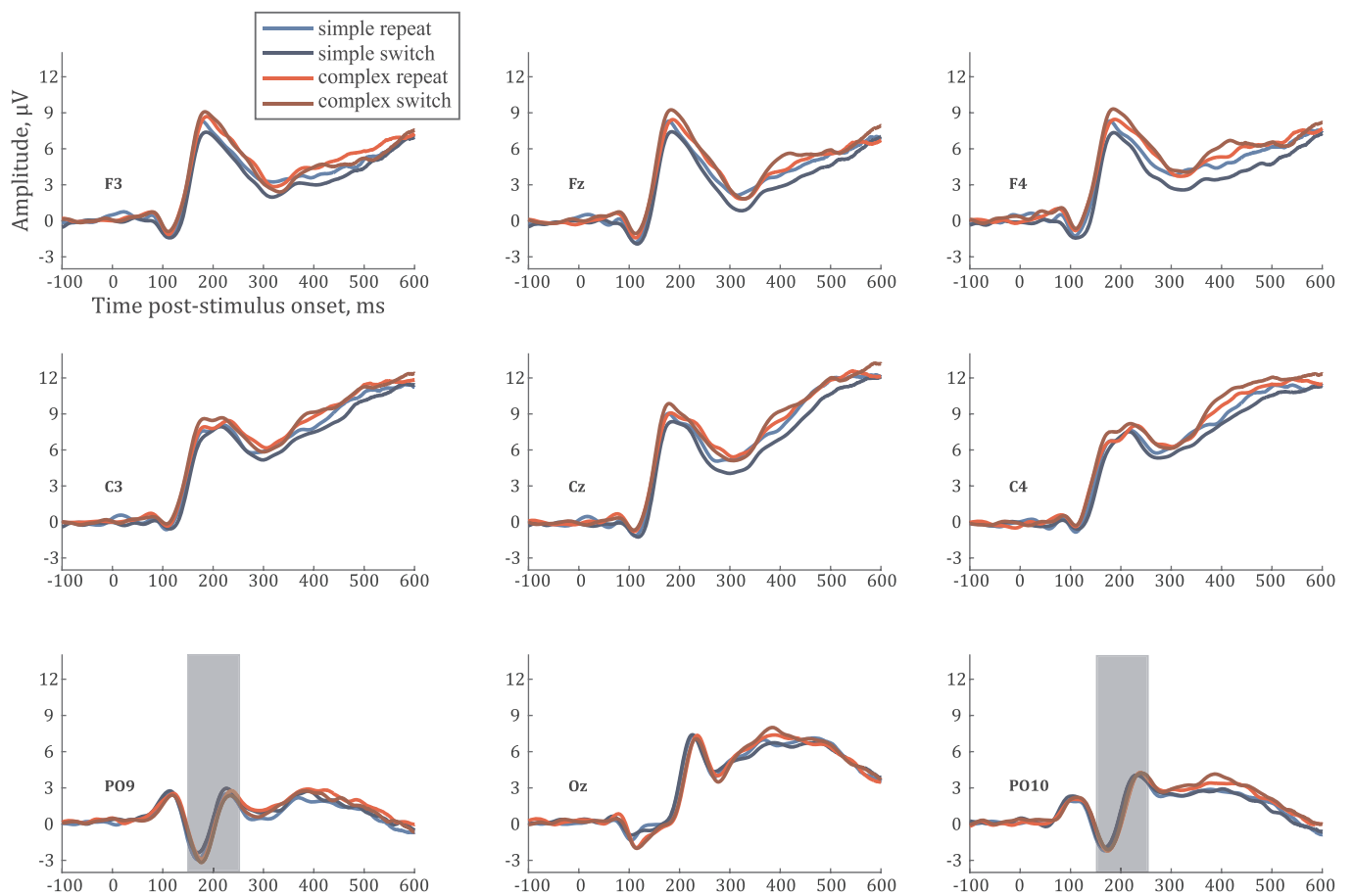


Fig. B. Group-averaged ERPs for nine electrode sites (i.e., F3, Fz, F4, C3, Cz, C4, PO9, Oz, and PO10) for the simple repeat, simple switch, complex repeat, and complex switch conditions.

## References

- Boersma, P., 2002. Praat, a system for doing phonetics by computer. *Glottol. Int.* 5, 341–345.
- Brydges, C.R., Fox, A.M., Reid, C.L., Anderson, M., 2014. Predictive validity of the N2 and P3 ERP components to executive functioning in children: a latent-variable analysis. *Front. Hum. Neurosci.* 8, 1–10.
- Bürki, A., Laganaro, M., 2014. Tracking the time course of multi-word noun phrase production with ERPs or on when (and why) cat is faster than the big cat. *Front. Psychol.* 5.
- Bürki, A., Sadat, J., Dubarry, A.-S., Alario, F.-X., 2016. Sequential processing during noun phrase production. *Cognition* 146, 90–99.
- Capizzi, M., Ambrosini, E., Arbula, S., Mazzonetto, I., Vallesi, A., 2016. Electrophysiological evidence for domain-general processes in task-switching. *Front. Hum. Neurosci.* 10.
- Eulitz, C., Hauk, O., Cohen, R., 2000. Electroencephalographic activity over temporal brain areas during phonological encoding in picture naming. *Clin. Neurophysiol.* 111, 2088–2097.
- Evans, J.L., Selinger, C., Pollak, S.D., 2011. P3 as a measure of processing capacity in auditory and visual domains in specific language impairment. *Brain Res.* 1389, 93–102.
- Folstein, J.R., Van Petten, C., 2008. Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology* 45, 152–170.
- Habets, B., Jansma, B.M., Münte, T.F., 2008. Neurophysiological correlates of linearization in language production. *BMC Neurosci.* 9, 1–8.
- Hung, C.L., Huang, C.J., Tsai, Y.J., Chang, Y.K., Hung, T.M., 2016. Neuroelectric and behavioral effects of acute exercise on task switching in children with attention-deficit/hyperactivity disorder. *Front. Psychol.* 7 (Article 1589).
- Indefrey, P., 2011. The spatial and temporal signatures of word production components: a critical update. *Front. Psychol.* 2 (Article 255).
- Indefrey, P., Levelt, W.J., 2004. The spatial and temporal signatures of word production components. *Cognition* 92 (1), 101–144.
- Jackson, G.M., Swainson, R., Cunnington, R., Jackson, S.R., 2001. ERP correlates of executive control during repeated language switching. *Biling.: Lang. Cogn.* 4, 169–178.
- Kahneman, D., 1973. *Attention and Effort*. Prentice-Hall, Englewood Cliffs, N.J.
- Kamijo, K., Takeda, Y., 2010. Regular physical activity improves executive function during task switching in young adults. *Int. J. Psychophysiol.* 75, 304–311.
- Kok, A., 2001. On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology* 38, 557–577.
- Kramer, A.F., Sirevaag, E.J., Braune, R., 1987. A psychophysiological assessment of operator workload during simulated flight missions. *Hum. Factors* 29, 145–160.
- Levelt, W.J.M., 1989. *Speaking: from Intention to Articulation*. MIT Press, Cambridge, MA.
- Lorist, M.M., Klein, M., Nieuwenhuis, S., De Jong, R., Mulder, G., Meijman, T.F., 2000. Mental fatigue and task control: planning and preparation. *Psychophysiology* 37, 614–625.
- Luck, S.J., Hillyard, S.A., 1994. Electrophysiological correlates of feature analysis during visual search. *Psychophysiology* 31, 291–308.
- Marek, A., Habets, B., Jansma, B.M., Nager, W., Münte, T.F., 2007. Neural correlates of conceptualisation difficulty during the preparation of complex utterances. *Aphasiology* 21, 1147–1156.
- Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190.
- Martinovic, J., Gruber, T., Müller, M.M., 2008. Coding of Visual Object Features and Feature Conjunctions in the Human Brain. *PLoS ONE* 3 (11), e3781.
- Meyer, A.S., 1996. Lexical access in phrase and sentence production: results from picture-word interference experiments. *J. Mem. Lang.* 35, 477–496.
- Michel Lange, V.M., Perret, C., Laganaro, M., 2015. Comparison of single-word and adjective-noun phrase production using event-related brain potentials. *Cortex* 67, 15–29.
- Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., Wager, T.D., 2000. The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100.
- Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 1–9.
- Polich, J., 2007. Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148.
- Polich, J., Kok, A., 1995. Cognitive and biological determinants of P300: an integrative review. *Biol. Psychol.* 41, 103–146.
- Proverbio, A.M., Burco, F., del Zotto, M., Zani, A., 2004. Blue piglets? Electrophysiological evidence for the primacy of shape over color in object recognition. *Cogn. Brain Res.* 18, 288–300.
- Pylkkänen, L., Bemis, D.K., Elorrieta, E.B., 2014. Building phrases in language production: an MEG study of simple composition. *Cognition* 133, 371–384.
- Roelofs, A., Piai, V., 2011. Attention demands of spoken word planning: a review. *Front. Psychol.* 2 (Article 307).
- Roelofs, A., Shitova, N., 2017. Importance of response time in assessing the cerebral dynamics of spoken word production: Comment on Munding et al. (2016). *Lang. Cogn. Neurosci.* 8, 1064–1067.
- Rogers, R.D., Monsell, S., 1995. Costs of a predictable switch between simple cognitive tasks. *J. Exp. Psychol.: Gen.* 124, 207–231.
- Schriefers, H., 1992. Lexical access in the production of noun phrases. *Cognition* 45, 33–54.
- Sikora, K., Roelofs, A., Hermans, D., 2016a. Electrophysiology of executive control in spoken noun-phrase production: dynamics of updating, inhibiting, and shifting. *Neuropsychologia* 84, 44–53.
- Sikora, K., Roelofs, A., Hermans, D., Knoors, H., 2016b. Executive control in spoken noun-phrase production: contributions of updating, inhibiting, and shifting. *Q. J. Exp. Psychol.* 69, 1719–1740.
- Strayer, D.L., Drews, F.A., 2007. Cell-phone-induced driver distraction. *Curr. Dir. Psychol. Sci.* 16, 128–131.
- Strijkers, K., Costa, A., Thierry, G., 2010. Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cereb. Cortex* 20, 912–928.
- Strijkers, K., Holcomb, P.J., Costa, A., 2011. Conscious intention to speak proactively facilitates lexical access during overt object naming. *J. Mem. Lang.* 65, 345–362.
- Swainson, R., Cunnington, R., Jackson, G.M., Rorden, C., Peters, A.M., Morris, P.G., Jackson, S.R., 2003. Cognitive control mechanisms revealed by ERP and fMRI: evidence from repeated task-switching. *J. Cogn. Neurosci.* 15, 785–799.
- Van Casteren, M., Davis, M.H., 2006. Mix, a program for pseudorandomization. *Behav. Res. Methods* 38, 584–589.
- Watter, S., Geffen, G.M., Geffen, L.B., 2001. The n-back as a dual-task: P3 morphology under divided attention. *Psychophysiology* 38, 998–1003.