The nature of the verbal self-monitor

Ganushchak, A.Y.

Chapter 5

Motivation and semantic context affect brain error-monitoring activity: An event-related brain potentials study[9]

Abstract

During speech production, we continuously monitor what we say. In situations in which speech errors potentially have more severe consequences, e.g. during a public presentation, our verbal self-monitoring system may pay special attention to prevent errors than in situations in which speech errors are more acceptable, such as a casual conversation. In an event-related potential study, we investigated whether or not motivation affected participants' performance using a picture naming task in a semantic blocking paradigm. Semantic context of to-be-named pictures was manipulated; blocks were semantically related (e.g., *cat*, *dog*, *horse*, etc.) or semantically unrelated (e.g., *cat*, *table*, *flute*, etc.). Motivation was manipulated independently by monetary reward. The motivation manipulation did not affect error rate during picture naming. However, the high-motivation condition yielded increased amplitude and latency values of the error-related negativity (ERN) compared to the low-motivation condition, presumably indicating higher monitoring activity. Participants showed semantic interference effects in reaction times, error rates and the ERN amplitude, presumably indicating that semantic relatedness induces more conflict between possible verbal responses.

---

9 This chapter is based on Ganushchak, L. Y. & Schiller, N. O. (2008). Motivation and semantic context affect brain error-monitoring activity: An event-related brain potentials study. *NeuroImage*, 39, 395-405.

Introduction

Speaking is a very fast and seemingly effortless process. In overt speaking, we produce up to 150 words per minute. However, the speech error rate in normal individuals is not more than one error in every 1,000 words (Levelt, 1989). Such low error rates may be the result of a verbal self-monitor that detects and corrects errors. The most prominent theory of verbal monitoring is the *perceptual-loop theory* proposed by Levelt (1983, 1989). According to this theory, there is a single, central verbal monitor that checks the message for its appropriateness, inspects the speech plan, and detects errors prior to its articulation via the speech comprehension system (Postma & Noordanus, 1996; Schiller, 2005, 2006; Schiller, Jansma, Peters, & Levelt, 2006; Wheeldon & Levelt, 1995; Wheeldon & Morgan, 2002), as well as after speech has become overt (Postma, 2000).

As stated above, the error rate under normal circumstances is very low indicating that verbal monitoring generally has low susceptibility to interference. However, there may be specific circumstances that produce interference with the working of the monitor. For instance, it is possible that in situations in which speech errors potentially have more significance because they are less acceptable, e.g. during giving an interview vs. having a casual conversation, the verbal self-monitoring system works harder in order to prevent errors. One question to ask is about the role of the verbal context in which a conversation takes place. If we hear or see information that is related to what we are planning to say, does that information interfere with verbal monitoring, thereby leading to more erroneous speech output? We will try to answer this question in the present study.

One way to study monitoring is by looking at error monitoring. An electrophysiological measure related to error processing is the so-called *Error-Related Negativity* (ERN; Falkenstein et al., 1991; Gehring et al., 1993), a component of the event-related potential (ERP) that has a fronto-central scalp distribution and peaks about 80 ms after an overt incorrect response (Bernstein et al., 1995; Holroyd & Yeung, 2003; Scheffers et al., 1996). Originally, the ERN was thought to reflect conscious error detection (Bernstein et al., 1995). However, according to the conflict hypothesis, the ERN arises not due to error detection *per se* but rather as a result of response conflict that arises when multiple responses compete for selection (Botvinick et al., 2001; Carter et al., 1998). Presence of conflicting responses reflects situations in which errors are likely to occur. Thus, according to the conflict hypothesis error detection is not an independent process but based on the presence of response conflict.

Alternatively, the reinforcement-learning theory proposed that the ERN may reflect a negative reward-prediction error signal that is elicited when the monitor detects that the consequences of an action are worse than expected. This reward-prediction error signal is coded by the mesencephalic dopamine system and projected to the anterior cingulated cortex (ACC), where the ERN is elicited (Holroyd & Coles, 2002).

Interestingly, a number of studies demonstrated the influence of emotional/motivational factors on the ERN (e.g., Boksem et al., 2006; Luu et al., 2000; Pailing & Segalowitz, 2004; Ullsperger & Von Cramon, 2004). The general finding is that the ERN increases when monetary incentives are offered for accuracy (Gehring et al., 1993; Hajack et al., 2005; Pailing & Segalowitz, 2004). For instance, Pailing and Segalowitz (2004) manipulated value of response error by selectively financially rewarding one type of response over another in a four-choice letter task. Pailing and Segalowitz found that more costly types of errors were associated with higher amplitude of the ERN. However, this dependency was only present for participants who scored high on neuroticism. Hajack and colleagues (2005) also investigated whether the ERN is sensitive to value of errors. They manipulated motivational significance or errors by administrating monetary punishment for them. Consistent with previous studies, these authors showed that the ERN was significantly larger on high-value errors than low-value errors. Consistently with the EEG studies, Ullsperger and Von Cramon (2004) performed an fMRI study in which they also modulated the relevance of errors by a financial reward manipulation. Ullsperger and Von Cramon found that error-related activation in posterior fronto-medial cortex, previously shown to be involved in performance monitoring, was modulated by error relevance.

Most studies on the ERN investigate the working of action monitoring. In the present study, however, we use the ERN to explore the workings of the *verbal* monitoring system. There are only few studies that looked at the ERN after verbal errors (see Ganushchak & Schiller, 2006, in press; Masaki et al., 2001; Möller, Jansma, Rodríguez-Fornells, & Münte, 2007; Sebastián-Gallés, Rodríguez-Fornells, Diego-Balaguer, & Díaz, 2006), which we will briefly review below.

Masaki and colleagues (2001) examined whether or not the ERN occurs in relation to speech errors in the Stroop color-word task. Participants in their study were instructed to overtly name the color of each stimulus as quickly and accurately as possible. Masaki and colleagues found an ERN-like response after speech errors, e.g. when participants named the wrong color.

Sebastián-Gallés and colleagues (2006) assessed Spanish-dominant and Catalan-dominant bilinguals using an auditory lexical decision task in Catalan. The authors showed that Spanish-dominant bilinguals had great difficulty in rejecting experimental non-words, and did not show an ERN in their erroneous non-word decisions either. According to Sebastián-Gallés et al. this suggests that Spanish-dominant bilinguals activated the same lexical entry from experimental words and non-words (in the experimental stimuli, the vowel change involved a Catalan-specific /e – ɛ/ contrast) and therefore showed no differences between correct and erroneous responses. In contrast, Catalan-dominant bilinguals demonstrated a clear ERN.

Recently, Möller et al. (2007) employed a laboratory task known to elicit speech errors to investigate verbal monitoring. In this task, participants are presented with inductor word pairs such as '*ball doze*', '*bash door*', and '*bean deck*', which are followed by a target word pair such as '*darn bore*' (see Motley et al., 1982). The reversal of initial phonemes in the target pair

compared to the inductor pairs is supposed to lead to speech errors such as '*barn door*'. Möller and colleagues asked their participants to covertly read the inductor word pairs and vocalize the target word pair immediately preceding a response cue. They found a negative deflection on error trials, as compared to correct trials, preceding the response cue. Möller et al. proposed that this activity reflects the simultaneous activation of competing speech plans. However, these authors do not make an explicit link between the negativity they found in their study and the ERN.

Ganushchak and Schiller (2006) used a phoneme-monitoring task to investigate the effects of verbal monitoring under time pressure. Participants were presented with pictures and had to indicate whether the target phoneme was present in the name of the picture. For example, if the presented picture was *table* and target phoneme was /t/, then participants had to press a button; however, if the target phoneme was /m/, they had to withhold their response. Ganushchak and Schiller obtained an ERN following verbal errors which showed a typical decrease in its amplitude under severe time pressure.

In more recent study by the same authors (Ganushchak & Schiller, in press), a similar phoneme-monitoring task was employed to investigate the effect of auditory distractors on verbal monitoring. Participants were requested to press a button when a target phoneme was present in the pictures' name. However, simultaneously with the picture participants heard a semantically related distractor, a semantically unrelated distractor, or no distractor at all. Ganushchak and Schiller observed a larger ERN when auditory distractors were semantically related to the picture than when distractors were unrelated or no distractors were present at all. Presence of distractors, by activating more related concepts, presumably increased conflict at the time of response and therefore led to higher amplitudes of the ERN. This result may indicate that the ERN after verbal errors, as well as after general performance errors, is sensitive to conflict present at the time of response (see Botvinick et al., 2001). The goal of the present study was to further investigate the relationship between the ERN and verbal monitoring.

In the study described above, Ganushchak and Schiller (in press) used a phoneme-monitoring task in which button-press responses were required, and not pure verbal responses. In contrast, in the current study, we employed a blocked picture naming task in which recorded responses were overt verbal responses. The blocked naming paradigm manipulates the context in which to-be-named pictures appear. In semantically related blocks, pictures from the same semantic category appear on successive trials, for example *table, chair, couch*, and *closet*. In contrast, in semantically unrelated, mixed blocks, pictures from different semantic categories appear one at a time, for instance *table, snake, apple*, and *car*. Speakers take longer to name pictures from the same semantic category than from different categories. This increase in naming latencies is attributed to the increased competition for lexical selection from semantically related competitors (Belke, Meyer, & Damian, 2005; Damian, Vigliocco, & Levelt, 2001; Levelt et al., 1999; Schnur, Schwartz, Brecher, & Hodgson, 2006).

In our own study, we employed this semantic blocking picture naming paradigm to

investigate the effects of the semantic context on verbal self-monitoring and the ERN. How does semantic blocking relate to the verbal self-monitor? According to the Levelt's perceptual loop theory (1983, 1989), the verbal self-monitoring system not only monitors for errors, but also for semantic appropriateness. In semantically related blocks, the monitor presumably checks whether or not the correct alternative has indeed been chosen as the target response from the set of competing items (Maess et al., 2002). This check is less urgent in the mixed blocks, where co-activation of competing items supposedly occurs less than in the semantically related blocks. In the present study, we expected to find more errors and slower reaction times while naming pictures in semantically related blocks, compared to naming pictures in mixed blocks.

Moreover, we were interested to examine what happens when people commit errors. Will error signals be different when an error occurred during semantically related blocks compared to error signals during mixed blocks? In semantically related blocks, as opposed to mixed blocks, there are multiple semantically related entries active which compete for the lexical selection, thus leading to a higher conflict between various semantically related competitors. The ERN is sensitive to the amount of conflict present at the time of the response (Botvinick et al., 2001). Therefore, one may hypothesize that the amplitude of the ERN will be larger following errors in the semantically related blocks than following errors in the mixed blocks.

In the present study, we investigated not only the effects of the semantic context, but also effects of motivation on verbal monitoring and the ERN. Pictures were presented in two colors, i.e. orange and purple, and participants were told that the more errors they make while naming orange pictures, the smaller their financial reward for participation would be, i.e. the high-motivation condition. If naming errors were made on purple pictures, participants received neither financial punishment, nor financial reward: the low-motivation condition. In the high-motivation condition, making errors had more consequences for participants than making errors during the low-motivation condition. Therefore, we expected to find higher amplitudes of the ERN during the high-motivation than the low-motivation condition. In previous research, it has been shown that the amplitude of the ERN is increased when response accuracy is emphasized over speed (e.g., Falkenstein et al., 2000; Gehring et al., 1993). Furthermore, participants were instructed to name pictures as quickly and as accurately as possible in both high-motivation and low-motivation conditions in the current study. Therefore, we did not expect to find differences in reaction times in naming pictures in the high-motivation and low-motivation conditions.

Note that in semantically related and mixed blocks participants were instructed to name pictures as fast and as accurately possible. However, between these two types of blocks, we expected to find reaction time differences because the semantic interference effect is a robust effect which occurs automatically without participants' awareness. In contrast, the behavioral differences in the high-motivation versus the low-motivation conditions most likely do not occur due to automatic processes but due to strategies participants applied, for instance, in order to gain accuracy in the high-motivation condition by slowing the responses down, compared to

the low-motivation condition. Therefore, by stressing the importance of reacting as fast and as accurately as possible we hoped to counteract the potential differences in response strategies in the high-motivation versus the low-motivation conditions, and therefore expected no behavioral differences between these conditions.

To summarize, we predicted that participants will be slower and make more errors naming pictures in semantically related blocks than in mixed blocks. Moreover, we expected to obtain an ERN after erroneous trials across all conditions. However, the amplitude of the ERN should increase while naming pictures in semantically related blocks compared to mixed blocks. Finally, the amplitude of the ERN should also be larger during the high-motivation condition than low-motivation condition.

Methods

Participants

Twenty-two students of Maastricht University (19 female) took part in the experiment. All participants were right-handed, native Dutch speakers, and had normal or corrected-to-normal vision. Participants gave written informed consent prior to participating in the study. They received a small financial reward for their participation in the experiment. Due to technical problems, the data of three participants were lost.

Materials

Seventy-five simple line drawings were selected from fifteen semantic categories of five exemplars each (see Appendix). Items for categories were selected minimizing within-category visual similarity. In a control study, we asked another 20 participants to ignore the semantic similarity of the pictures and judge all 150 pairs of pictures in terms of visual similarity. Participants were required to rate visual similarity on a five-point scale (1 = not similar at all, 5 = very similar). Within-category similarity was only slightly higher (mean: 2.4) than between-category similarity (mean: 1.6). These judgments are similar to the ones reported in Damian et al. (2001), who obtained a within-category similarity of 2.4 and a between-category similarity of 1.9.

Picture names were on average 1.3 syllables long (range: 1 – 3) and had a moderate frequency of occurrence between 10 and 100 per million according to the CELEX database (CEnter for LEXical information, Nijmegen; Baayen, Piepenbrock, & Gulikers, 1995). Furthermore, each picture was presented once in orange and once in purple (i.e. 255, 127, 0 and 158, 73, 161, respectively, on the RGB scale), and degraded with 9 pt dashes and 16 pt spaces, and the weight of the lines was 3 pt. Pictures were dashed to make the task a bit more difficult and provoke participants to commit more speech errors. In the simple picture naming task, participants could recognize and name all dashed pictures correctly (for examples of stimuli, see Figure 1 on the back side of the cover).

Design

The experiment consisted of learning, practice, and main task. During the learning and practice phases, participants saw all pictures in black-on-white in the middle of the screen. In the learning phase, each picture was presented simultaneously with its label written underneath it. Pictures stayed on the screen for 2,000 ms. In the practice task, participants saw the same pictures without the picture names written underneath. In practice and main tasks, a trial consisted of a fixation point with variable duration (between 500 and 800 ms), a blank screen for 500 ms, and the target stimulus, i.e. a picture. Pictures disappeared from the screen as soon as the voice key was activated or after 500 ms maximally. The inter-trial interval was variable, depending on the response latency, such that each trial had a total duration of 2,000 ms.

For the main task, five-item sets were formed. In semantically related blocks, five exemplars from the same semantic category were presented together in a set (e.g., *giraffe* 'giraffe', *kameel* 'camel', *hert* 'deer', *olifant* 'elephant', *zebra* 'zebra'). In mixed blocks, the five-item sets comprised exemplars from different semantic categories (e.g., *giraffe* 'giraffe', *bank* 'couch', *arm* 'arm', *piano* 'piano', *citroen* 'lemon'). Each block consisted of five pictures that were repeated four times in different order, resulting in blocks of 20 trials each. In total, there were 15 semantically related and 15 mixed blocks. Each block was presented twice: once with all pictures in a block colored orange and once with all pictures colored purple. In such a way, motivation was manipulated on a block basis. Each participant saw 15 semantically related and 15 mixed blocks of 20 trials each and repeated in both colors, i.e. 30 blocks x 20 trials x 2 colors = 1,200 trials altogether. The order of pictures was pseudo-randomly varied for each participant in such a way that identical pictures did not appear on consecutive trials. The order of blocks followed a Latin square design.

Procedure

Participants were tested individually while seated in a sound-proof room. They were presented with the learning phase, the practice task, and, finally, the main task. During the learning task, participants were familiarized with the pictures and their corresponding names. In the practice task, participants were asked to overtly name the pictures with the labels they learned during the learning phase. If errors were made, participants were told about their mistakes and correct responses were provided by the experimenter. During the main task, participants were asked to overtly name the same pictures as in the practice task. Participants were told that their total financial reward for participation would depend on their performance during the task. If they made errors during naming pictures presented in orange, they would receive €12.50, €10, €7.50, or €5 depending on amount of errors made. For errors made on pictures presented in purple, there was neither financial gain nor punishment. Participants were asked to overtly name all orange and purple pictures as fast and as accurately possible. At the end of the experiment, all participants received the full financial reward independent of their performance.

Apparatus and Recordings

The electroencephalogram (EEG) was recorded from 29 scalp sites (extended version of the 10/20 system) using tin electrodes mounted to an electro cap. The EEG signal was sampled at 250 Hz with band-pass filter from 0.05 to 30 Hz. An electrode at the left mastoid was used for on-line referencing of the scalp electrodes. Off-line analysis included re-referencing of the scalp electrodes to the average activity of two electrodes placed on the left and right mastoid. Eye movements were recorded to allow off-line rejection of contaminated trials. Lateral eye movements were measured using a bipolar montage of electrodes placed on the right and left external canthus. Eye blinks and vertical eye movements were measured using bipolar montage of electrodes placed above and below the left eye. Impedance level for all electrodes was kept below 5kΩ.

Data analysis

Epochs of 1,300 ms (from –400 ms to +900 ms) were computed. A 100 ms pre-response baseline was used. The EEG signal was corrected for vertical EOG artifacts, using the ocular reduction method described in Anderer, Satety, Kinsperger, and Semlitsch (1987). The ERN was measured in response-locked ERP averages. For the ERN, averaging was carried out across error trials at the voice-key onset of the erroneous response. For the correct trials, averaging was done for the voice-key onset of the correct responses. To compute the difference between correct and error trials, a mean area amplitude analysis was used in a time window between 0 and 100 ms after response onset. For this analysis, we used a mean area amplitude analysis since it was impossible to identify peaks on correct trials. The amplitude of the ERN was derived from each individual's average waveforms after filtering with a band pass, zero phase shift filter (frequency range: 1 – 12 Hz). The ERN was quantified by peak-to-peak measurements that were calculated to determine baseline-independent amplitudes of negative deflections by subtracting the amplitude of the preceding positive peak from the negative peak of this component (Falkenstein et al., 2000). Thus, the amplitude of the ERN was defined as the difference between the most negative peak in a window from 0 to 150 ms after the response and the most positive peak from –50 to 0 ms preceding the ERN (Falkenstein et al., 2000). The latency of the ERN was defined as a point in time when the negative peak was at its maximum. The amplitude and the latency of the ERN were recorded for each condition at electrode sites Fz, FCz, Cz, and Pz.

Results

Behavioral data

Latencies shorter than 300 ms and longer than 1,500 ms were excluded from the analysis. Effects of Motivation and Semantic Relatedness on naming latencies were assessed by repeated measures ANOVAs. These analyses revealed a significant effect of Semantic Relatedness (F(1,

18) = 10.43, MS_e = 457.42, p < .001), but no effect of Motivation (F(1, 18) = 1.08, MS_e = 553.29, n.s.) and no interaction between the two factors (F(1, 18) = 2.71, MS_e = 373.42, n.s.). Naming latencies were 15 ms longer during semantically related blocks than during mixed blocks (see Table 1 for mean naming latencies).

Similar analyses are reported for number of errors. Participants made on average 3.0% phonological and semantic errors. Eighty percent of these errors were phonological errors and only 20% were semantic errors. An example of a semantic error is when a participant incorrectly named the picture of a *giraffe* as *zebra.* An example of a phonological error is the non-word *drood* instead of *brood* 'bread'. Trials on which participants failed to give a response or the voice key was triggered by an inappropriate response (e.g., sneezing or coughing) were not included in the analysis (0.8 % of all trials). A repeated measures ANOVA with number of errors as dependent variable revealed a significant effect of Semantic Relatedness (F(1, 18) = 7.77, MS_e = 17.55, p < .01). Participants made more errors during semantically related blocks than during mixed blocks (see Table 1 for error rates). There was no effect of Motivation nor was there an interaction between Motivation and Semantic Relatedness (both Fs < 1).

Table 1. Overview of the behavioral data. Mean (± standard deviation) reaction times (in ms) and error rates (% relative to the number of trials per condition) as a function of motivation and semantic context manipulation.

	Semantically related blocks	*Mixed blocks*	*High-motivation*	*Low-motivation*
Reaction times	683 (65)	668 (63)	673 (64)	678 (65)
Error rates	3.3 (7)	2.8 (6)	2.9 (6)	3.0 (6)

Electrophysiological data

Inspection of the grand averages for error trials revealed a clear negative deflection on error trials but not on correct trials (see Figures 2 and 3). Figure 4 shows the scalp distribution of the ERN by means of topographic maps. A repeated measures ANOVA with mean ERN amplitude as dependent variable revealed a significant effect of Condition (correct vs. incorrect; F(1, 18) = 6.79, MS_e = 40.42, p < .05). As expected, the ERN appeared only on erroneous trials and not on correct trials. First, we looked at errors made during semantically related and semantically unrelated blocks separately in high-motivation and low-motivation conditions. This analysis revealed no significant interaction between Motivation and Semantic Relatedness (F(1, 18) < 1). Therefore, for all analyses described below, errors made during semantically related and mixed blocks in the high-motivation condition were collapsed. Errors made during semantically related and mixed blocks in the low-motivation condition were also pulled together. Similar, we collapsed errors

made in semantically related and mixed blocks across high- and low-motivation conditions.

In the remaining analyses, amplitudes and latencies of the ERN were submitted to a repeated-measures General Linear Model (GLM) analysis. All ANOVAs included two factors, i.e. Electrode Site (Fz, FCz, Cz, and Pz) and either Motivation (high vs. low) or Semantic Relatedness (semantically related vs. mixed).

To investigate the effect of motivation on the ERN, repeated measures ANOVAs were run with Motivation as independent variable and amplitude of the ERN as dependent variable. This analysis demonstrated that the amplitude of the ERN was modulated by the motivation manipulation ($F(1, 18) = 5.44$, $MS_e = 26.26$, $p < .05$). The amplitude of the ERN was significantly larger in the high-motivation condition (-5.44 µV, SD = 3.50) compared to the low-motivation condition (-3.51 µV, SD = 1.94). There was no significant effect of Electrode Site ($F(3, 54) = 1.97$, $MS_e = 1.98$, n.s.), nor was there an interaction between Electrode Site and Motivation ($F < 1$).

Interestingly, similar analyses with latency of the ERN as dependent variable revealed a significant effect of Motivation ($F(1, 18) = 18.38$, $MS_e = 2986.56$, $p < .001$). The ERN peaked significantly later in the high-motivation condition (92 ms, SD = 36) compared to the low-motivation condition (54 ms, SD = 28). There was no effect of Electrode Site ($F(3, 54) = 1.49$, $MS_e = 218.73$, n.s.), nor was there an interaction between Electrode Site and Motivation ($F(3, 54) = 2.32$, $MS_e = 332.22$, n.s.).

The corresponding analysis with Semantic Relatedness as independent variable and amplitude of the ERN as dependent variable demonstrated a significant effect of Semantic Relatedness ($F(1, 18) = 8.63$, $MS_e = 68.42$, $p < .01$). The amplitude of the ERN was significantly larger in the semantically related blocks (-0.73 µV, SD = 2.75) as opposed to the mixed blocks (1.18 µV, SD = 2.86). The analysis of Electrode Site ($F < 1$) and the interaction between Electrode Site and Semantic Relatedness ($F < 1$) revealed no significant effects. Similar analyses with latency of the ERN as dependent variable showed no significant results (all $Fs < 1$).

Figure 3 provides an overview of the response-locked averaged ERP waveforms for errors and correct trials across semantically related and mixed blocks. There was a significant effect of Electrode Site ($F(3, 54) = 4.01$, $MS_e = 218.06$, $p < .05$). Further investigation of this effect revealed that the ERN peaked significantly later at electrode site Pz than at electrode site Fz ($F(1, 18) = 5.43$, $MS_e = 381.80$, $p < .05$). However, there was no interaction between Electrode Site and Semantic Relatedness ($F(3, 54) = 2.51$, $MS_e = 305.53$, n.s.)

High-motivation   Low-motivation

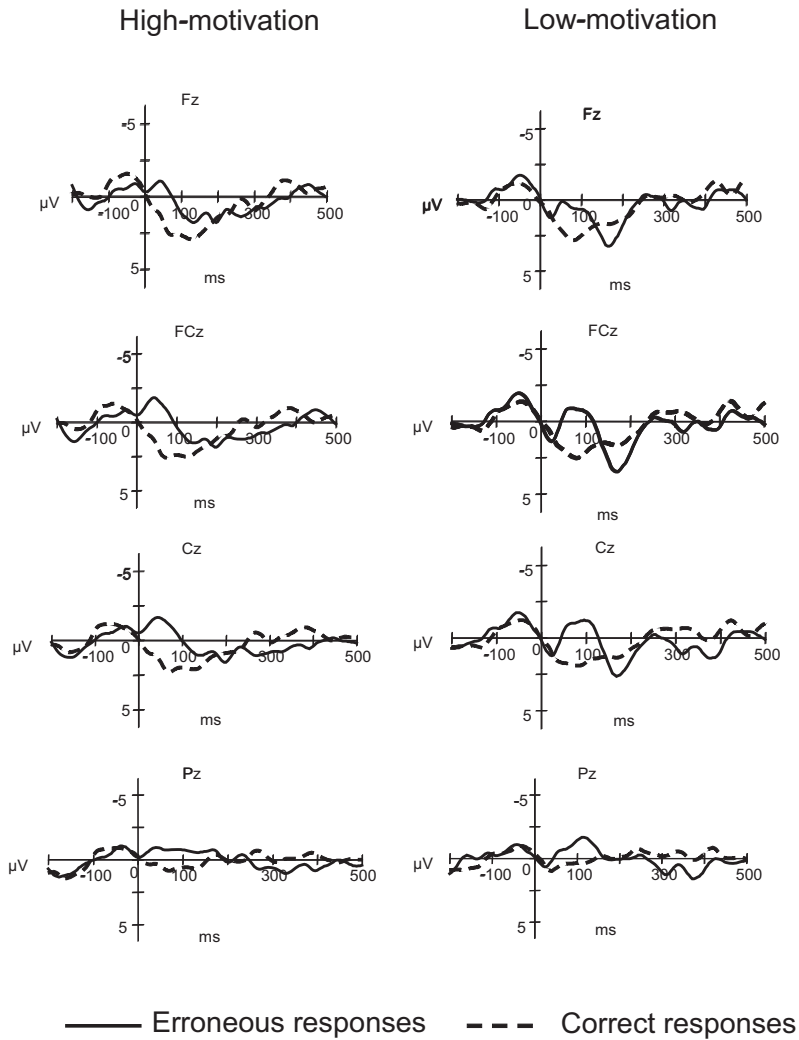――― Erroneous responses   ― ― ― Correct responses

Figure 2. Averaged response-locked ERP waveforms for all error trials (solid lines) versus correct trials (dashed lines) across high-motivation and low-motivation conditions. Correct and incorrect trials were matched on RTs and number of trials. For graphical representation, waveforms where filtered with a high-pass filter; all analyses were done prior to the filtering.

Mixed blocks   Semantically related blocks
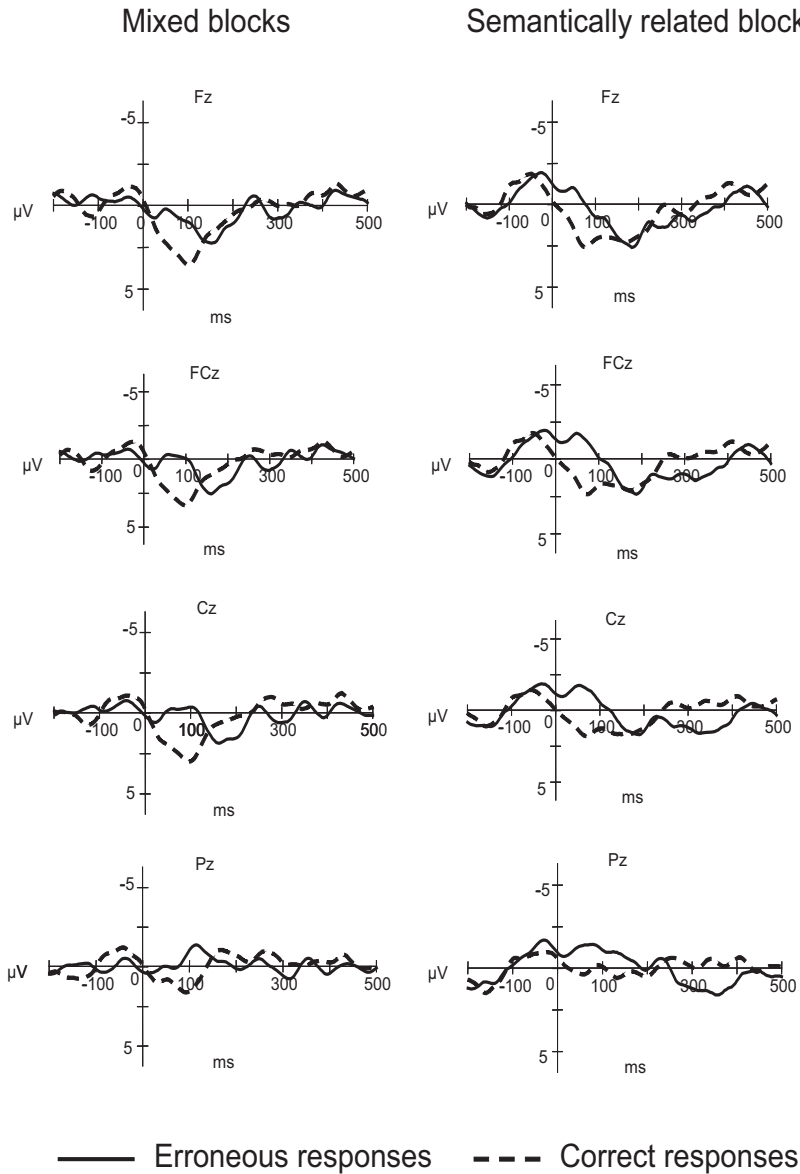

—— Erroneous responses   – – – Correct responses

Figure 3. Averaged response-locked ERP waveforms for all error trials (solid lines) versus correct trials (dashed lines) across mixed and semantically related blocks. Correct and incorrect trials were matched on RTs and number of trials. For graphical representation, waveforms where filtered with a high-pass filter; all analyses were done prior to the filtering.
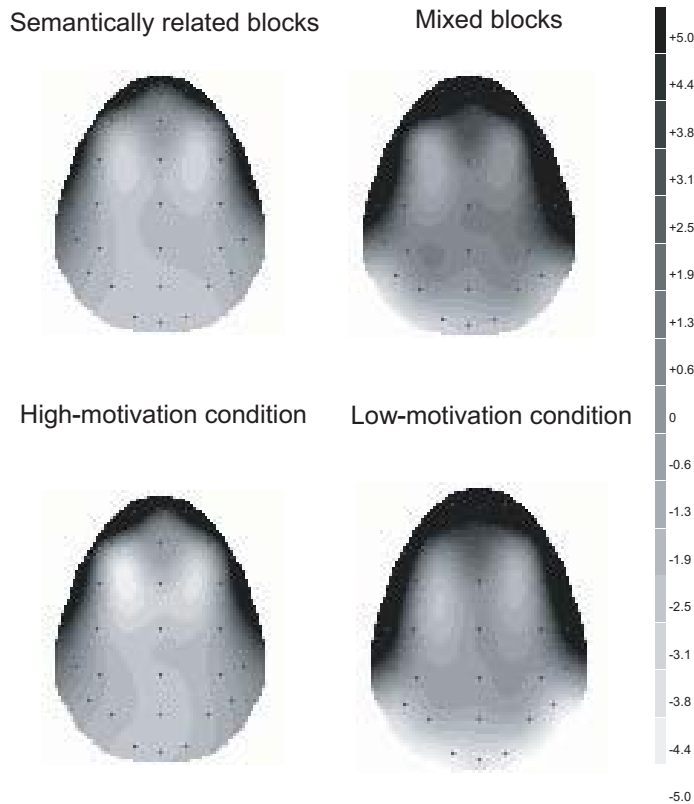
Figure 4. Topographic maps of the ERN amplitude between 0 and 100 ms after response onset. Negative regions depicted in light gray.

Discussion

The goal of the present study was to investigate how verbal self-monitoring and the ERN are affected by motivation and semantic context. As expected, we obtained a typical semantic interference effect (for instance, Lupker, 1979; Schriefers, Meyer, & Levelt, 1990; etc.). Participants were slower and made more errors during picture naming in semantically related blocks as compared to mixed blocks. Our motivation manipulation had no effect on naming latencies or error rates indicating that participants did not employ different response strategies in the high- and low-motivation conditions. In the electrophysiological data, we saw a clear negative deflection on error trials compared to correct trials. Investigation of topographic maps revealed that this negative deflection has a fronto-central distribution, typical to the classic ERN (e.g., Gehring et al., 1993; Falkentein et al., 1991). However, in our statistical analysis, there

was no significant effect of Electrode Site suggesting that the negative deflection found in our study was more evenly distributed across the scalp. Note, however, that in our analysis only central electrodes were included. Thus, while statistically the ERN was no different at frontal and posterior electrode sites, it is statistically unclear whether or not the ERN was largest at central versus lateral electrode sites. One possible reason for non-significant effect of Electrode Site is that we measured non-masked overt speech as opposed to button presses. It is possible that the effect we have demonstrated in the present study has a more wide-spread distribution than the effect previously shown with motor tasks. However, all other characteristics of the negative deflection in our study are in correspondence with the classic ERN. The negative deflection in our study is present only on erroneous responses and absent from correct trials, it peaks within 100 ms of the overt erroneous response, and is descriptively largest at fronto-central electrode sites. Therefore, we would like to propose here that the negative deflection found in our study is an ERN-like response. In the remaining part of the discussion, we will refer to the ERN-like response in our study as the ERN.

The electrophysiological data obtained in the present study agree with our predictions. The amplitude of the ERN was significantly larger in semantically related blocks than in mixed blocks. Further, the amplitude of the ERN was also significantly larger in high- than low-motivation conditions. Interestingly, the latency of the ERN was also affected by the motivation manipulation. The ERN peaked significantly later in the high-motivation condition than in the low-motivation condition. Neither in the behavioral nor in the electrophysiological data was there an interaction between the context in which pictures were named and the motivation manipulation. These findings will be discussed in more detail below.

In accordance with previous studies (e.g., Damian et al., 2001, Schnur at al., 2006; Vitkovitch & Humphreys, 1991), we showed that pictures were named slower and responses were more erroneous in the context of the same-category items than in the context of items from different semantic categories. This semantic interference can be accounted for by competition between co-activated lexical entries in the same semantic context. This competition, in turn, affects selection latencies (see Levelt et al., 1999 for a review; but see Finkbeiner, Gollan, & Caramazza, 2006 as well as Mahon, Costa, Peterson, Vargas, & Caramazza, 2007 for an alternative view). Co-activation of multiple lexical entries may lead to multiple potential responses active at the same time and therefore increased conflict present at the time of response. The verbal self-monitor has to verify on-line whether or not the correct entry was chosen from the set of competing candidates, presumably resulting in slower naming responses. In the mixed context, however, such verification may be faster, since unrelated words do not lead to the activation of related concepts, and therefore less competition may be present at the time of the response.

Our EEG data are in accordance with this assumption. We showed that the amplitude of the ERN increases in the semantically related context as opposed to the mixed context. In previous research, Ganushchak and Schiller (in press) demonstrated an increase in the amplitude of the

ERN in the presence of semantically related distractors to target pictures as compared to unrelated distractors. They concluded that this increase in ERN amplitude during the semantically related condition was most likely due to the simultaneous activation of competing lexical items. The present study replicates and extends this finding. Ganushchak and Schiller (in press) employed a phoneme-monitoring task, in which participants were instructed to press a button when a target phoneme was present in the name of the picture. Even though a phoneme monitoring task is verbal in nature and involves monitoring of internal speech production, the authors could not completely exclude the possibility that at least some of the errors observed in their study were motor slips (of the hand) and not verbal errors *per se*. In contrast, in the present study, we employed a more natural picture naming task in which all responses given were verbal responses, and we demonstrated an enhancement of the amplitude of the ERN in the semantically related context compared to the mixed context.

A reinforcement-learning theory cannot fully account for the increase in the amplitude of the ERN in the semantically related blocks compared to the mixed blocks. According to this theory, errors induce a phasic decrease in mesencephalic dopaminergic activity when ongoing events are determined to be worse than expected (Holroyd & Coles, 2002). However, there is no reason to suggest that the monitoring system could not make an optimal evaluation of current events and events that were predicted in the mixed blocks compared to the semantically related blocks.

Alternatively, it is possible the difference in the ERN amplitude between semantically related and mixed blocks was due to the differences in naming latencies. As stated above, participants were slower in naming pictures in the semantically related blocks than mixed blocks. However, to our knowledge, differences in the response latencies only have an effect on the ERN when accuracy is emphasized over speed (e.g., Gehring et al., 1993). This, however, is not applicable to our case, since we did not observe a speed-accuracy trade-off: Participants were not only slower but also made more errors in the semantically related blocks than in the mixed blocks. There is a whole range of studies that demonstrate an increase in amplitude of the ERN on incongruent as compared to congruent trials, despite differences in the behavioral responses on these trials. Participants are generally slower and make more errors on incongruent compared to congruent trials (e.g., Hajcak et al., 2005; Fiehler et al., 2005; Yeung et al., 2004). This increase in amplitude of the ERN is attributed to the increased amount of response conflict during incongruent trials compared to congruent trials and not due to behavioral differences. Therefore, we think that in our case the higher amplitude of the ERN during semantically related blocks relative to mixed blocks is also due to the increased response conflict. This, in turn, provides stronger evidence that the ERN is sensitive to the conflict that arises during lexical competition.

Interestingly, the motivation manipulation also yielded an effect of ERN amplitude. The high-motivation condition was associated with higher amplitudes and longer latencies of the ERN compared to the low-motivation condition. However, in our behavioral data, we found no

effect of the motivation manipulation. Participants were equally fast and accurate in the high- and low-motivation condition. Due to the fact that the high- and low-motivation condition did not differ with respect to any behavioral measure, it is unlikely that the increased ERN in the high-motivation condition could be driven by performance-related differences. This is supported by Gehring and colleagues (1993) who showed in their original paper that the ERN was increased when accuracy was emphasized over speed. However, no such trade-off was observed in our behavioral data.

It is likely that in the high-motivation condition there was more conflict present than in the low-motivation condition, which consequently led to an enhanced ERN. According to the conflict monitoring theory, this could be explained by assuming that participants striving for accurate responses tended to focus more effectively on the task at hand in the high-motivation than in the low-motivation condition and therefore an increased tendency to correct errors in the former condition. This increased error-correcting activity, in turn, may have led to increased conflict with the error just produced, resulting in an increased ERN (Yeung, Botvinick, & Cohen, 2004).

In terms of the reinforcement learning theory (Holroyd & Coles, 2002), participants could have had a higher expectation of a good outcome in the high-motivation condition than in the low-motivation condition, since making errors in the high-motivation condition was associated with financial punishment. Therefore, errors in the high-motivation condition lead to larger violations of the prediction than errors in the low-motivation condition, and thus a more pronounced ERN was generated.

Alternatively, in the high-motivation condition, errors had a higher significance for participants than in the low-motivation condition. It is plausible that the ERN is sensitive to the motivational significance of errors. For instance, Hajcak and colleagues (2005) used an arrowhead version of a flanker task and varied monetary value on trial-by-trial basis. Interestingly, they found that errors committed in a condition with higher monetary value gave rise to a higher ERN. Note that Hajcak and colleagues demonstrated the effect of monetary value on the ERN despite it having no systematic effect on participants' behavioral performance. The authors concluded that the ERN reflects motivation significance of errors. Our own results are in agreement with their findings. It is possible that in circumstances when errors are less acceptable and potentially have more severe consequences, e.g. when giving a public speech, the verbal monitoring system has to be more alert in order to verify whether or not the selected response was correct.

Besides the increased amplitude of the ERN in the high-motivation condition, we also showed prolonged ERN latency, compared to the low-motivation condition. We expected to find no difference in ERN latency, since it seemed to be invariant with respect to the erroneous response and largely independent of experimental manipulation (Falkenstein et al., 2000; Scheffers & Coles, 2000). Recently, however, more studies showed a modulation of ERN latency depending on the experimental manipulations (e.g., Fiehler, Ullsperger, & Von Cramon, 2005; Johannes et al., 2001; Krigolson & Holroyd, 2007). It is possible that the latency of the ERN reflects prolonged

error-monitoring mechanisms (Johannes et al., 2001). In the high-motivation condition, errors had more severe consequences for participants than in the low-motivation condition. Therefore, it is plausible to assume that in the high-motivation condition error monitoring was slowed down in order to verify whether or not the selected response was indeed the correct response. In the low-motivation condition, such verification could be carried out faster and was less important, since errors did not have any consequences.

In semantically related and mixed blocks, there was no difference in ERN latency. According to the logic applied above, this lack of difference in ERN latency is not surprising. Errors made in semantically related blocks did not have more consequences for participants than errors made in mixed blocks. Alternatively, shift in ERN latency has been attributed to the corrective processes (e.g., Fiehler et al., 2005; Falkenstein et al., 1996). It has been proposed that slow error corrections seem to be based on a delayed correct response tendency resulting in a later peak of the ERN (Fiehler et al., 2005). However, this finding is not robust and there are studies that failed to show a relationship between ERN latency and error correction (Falkenstein et al., 1994; Rodríguez-Fornells et al., 2002). It is difficult to say whether or not our data can be interpreted with the help of this account. Possibly, participants internally corrected more errors during the high-motivation condition than the low-motivation condition and therefore showed a delayed ERN latency. However, in the behavioral data there was no significant difference in error rate between the high- and low-motivation conditions. If participants corrected more errors in the high-motivation condition than in the low-motivation condition, one would expect to see fewer errors in the former compared to the latter condition.

There is a seeming contradiction in our results. On the one hand, we showed that the amplitude of the ERN increased in the semantically related context as opposed to the mixed context, indicating that the ERN is sensitive to the presence of conflict. On the other hand, we demonstrated that the ERN was higher in the high-motivation condition than low-motivation condition. In the high-motivation condition, errors had a higher significance than in the low-motivation condition, indicating that the ERN is sensitive to the motivational manipulation. It is unlikely that errors in the semantically related context had higher significance than in the mixed context since the financial reward was independent of semantic context (see above). Furthermore, after the experimental session participants reported that they attempted to name pictures as accurately as possible independently of the context in which pictures were presented. To our knowledge, conflict and motivational accounts of the ERN are two mutually exclusive hypotheses in the existing literature, and it has not been shown so far that the ERN can be affected by both factors. We would like to propose here that possibly the conflict and motivational theories are closer related than previously thought. The detection of conflict or errors is likely to have direct affective consequences (Yeung, 2004). Ullsperger and Von Cramon (2004) also suggest that there might be a close interplay with emotional and motivational functions and performance monitoring. Therefore, a clear-cut distinction between theories that associate the ERN with a

process of error/conflict detection and theories that associate it with a process giving rise to affective/motivational changes related to error or conflict detection may not be possible, since both may refer to one and the same process (Yeung, 2004).

One potential methodological problem of the current study is the vocalization-related cortical potential (VRCP). The VRCP consists of a movement-related potential preceding vocalization and an auditory evoked potential which follows vocalization (Gunji, Hoshiyama, & Kakigi, 2000). The auditory evoked potential peaks around 100 ms after vocalization and therefore has a similar time course as the ERN. However, unlike the ERN, an auditory evoked potential is independent of response correctness. Masaki et al. (2001) used loud pink noise to suppress the vocalization-elicited components. In the current study, we did not use any masking procedure. Participants perceived their own voice as feedback to monitor their own speech. Removing such feedback, by masking participants' voices with pink noise, might disrupt and interfere with the normal working of the monitor. For instance, Christoffels, Formisano, and Schiller (2007) asked participants to name pictures when participants could hear their own voices and when they could not due to the presentation of masking pink noise during their responses. Christoffels and colleagues demonstrated that the masking of feedback was associated with a reduction of activity in areas found in overt speech production in comparison to the normal feedback condition. Therefore, in order to keep our task as natural as possible, we choose not to administer white or pink noise to our participants. We argue that, since auditory evoked potentials are independent of response correctness and are present during correct as well as erroneous trials, the changes between correct and erroneous responses found in the present study are due to the changes in the ERN and not auditory evoked potentials.

Finally, we would like to comment on the issue of individual differences since motivation is somewhat dependent on personality or mood characteristics of individual participants. For instance, Boksem et al. (2006) reported that individuals who score high on a measure of punishment sensitivity have larger ERN amplitudes than individuals who score low on such a measure. Pailing & Segalowitz (2004) demonstrated that individuals who score high on conscientiousness have smaller changes in ERN amplitudes with manipulations of motivation compared to individuals who score low on conscientiousness. However, our study had a within-subject design and participants were not pre-selected on their personality or mood characteristics but were randomly selected. Therefore, it is rather unlikely that our results could be accounted for by individual differences between participants.

To conclude, we argue that due to the simultaneous activation of competing items, the verbal self-monitor presumably needs to be more alert in the semantically related context than in the mixed context in order to validate whether a given response was correct or erroneous. Additionally, in circumstances when errors have more severe consequences for the speaker, e.g. when giving a public speech or during an interview, the monitor needs to work harder to prevent errors and to correct errors already made. Further, we showed that the ERN is sensitive to the

presence of lexical conflict. Similarly, in previous studies (i.e., Ganushchak & Schiller, in press) we reported an enhanced ERN after errors of verbal monitoring in the presence of semantically related distractors as opposed to semantically unrelated ones. Möller and colleagues (2007) – employing the inverse source localization method – identified a medial frontal generator in the supplementary motor area (SMA) as the main source of the negativity preceding erroneous vocalizations. A typical ERN has been located within the ACC/SMA regions (e.g., Dehaene, Posner, & Tucker, 1994). These findings provide converging evidence that the ERN could be used as an electrophysiological marker of error processing in language research. However, a note of caution may be in place here. To our knowledge, we are the first to use unmasked overt speech to investigate the ERN. As stated above, we did not find an unambiguous localization of the ERN in fronto-central electrodes. It is possible that the distribution of the ERN during overt speech is somewhat different than the classic ERN. The topography of the ERN during overt speech production clearly deserves further experimental investigation.

References

Anderer, P., Safety, B., Kinsperger, K. & Semlitsch, H. (1987). Topographic brain mapping of EEG in neuropsychopharmacology - Part 1. *Methodological aspects. Methods and Findings in Experimental Clinical Pharmacology,* 9, 371-384.

Baayen, R. H., Piepenbrock, R. and Gulikers, L. (1995). *The CELEX lexical database* (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. *The Quarterly Journal of Experimental Psychology*, 58A, 667-692.

Bernstein, P. S., Scheffers, M. K., & Coles, M. G. H. (1995). 'Where did I go wrong?' A psychophysiological analysis of error detection. *Journal of Experimental Psychology: Human perception and performance,* 21, 1312-1322.

Boksem, M. A. S., Tops, M., Wester, A. E., Meijman, T. F., Lorist, M. M. (2006). Error-related ERP components and individual differences in punishment and reward sensitivity. *Brain Research*, 1101, 92-101.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review,* 108, 624-652.

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D. C., & Cohen, J. D. (1998). Anterior cingulated cortex, error detection, and the online monitoring of performance. *Science,* 280, 747–749.

Christoffels, I. K., Formisano, E., & Schiller, N. O. (2007). Neural correlates of verbal feedback processing: an fMRI study employing overt speech. *Human Brain Mapping*, 28, 868-879.

Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a neural system for error detection and compensation. *Psychological Science,* 5, 3-23.

Damian, M. F., Vogliocco, G., & Levelt, W. J. M. (2001). Effects of semantic context in the naming of pictures and words. *Cognition,* 81, B77-B86.

Falkenstein, M., Hohnsbein, J., & Hoorman, J. (1994). Event-related potential correlates of errors in reaction tasks. In G. Karmos, M. Molnar, V. Csepe, I. Czigler, & J. E. Desmedt (Eds.), *Perspectives of event-related potentials research* (pp. 287-296). Amsterdam: Elsevier Science.

Falkenstein, M., Hohnsbein, J., & Hoorman, J. (1996). Differential processing of motor errors. In C. Ogura, Y. Koga, & M. Shimokochi (Eds.), *Recent advances in event-related brain potential research* (pp. 192-195). Tilburg: Tilburg University Press.

Falkenstein, M., Hohnsbein, J., Hoorman, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology,* 78, 447-455.

Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: a tutorial. *Biological Psychology,* 51, 87-107.

Fiehler, K., Ullsperger, M., & Von Cramon, D. Y. (2005). Electrophysiological correlates of error correction. *Psychophysiology*, 42, 72-82.

Finkbeiner, M., Gollan, T. H., & Caramazza, A. (2006). Lexical access in bilingual speakers: What's the (hard) problem? *Bilingualism: Language and Cognition*, 9, 153-166.

Ganushchak, L. Y., & Schiller, N. O. (2006). Effects of time pressure on verbal self-monitoring. *Brain Research*, 1125, 104-115.

Ganushchak, L. Y., & Schiller, N. O. (2008). Effects of auditory distractors on verbal self-monitoring. *Journal of Cognitive Neuroscience*, 20, 1-14.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science,* 4, 385-390.

Gunji, A., Hoshiyama, M., & Kakigi, R. (2000). Identification of auditory evoked potential of one's own voice. *Clinical Neurophysiology*, 111, 214-219.

Hajcak, G., Moser, J. S., Yeung, N., & Simons, R. F. (2005). On the ERN and the significance of errors. *Psychophysiology*, 42, 151-160.

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: reinforcement learning, dopamine and the error-related-negativity. *Psychological Review,* 109*,* 679-709.

Holroyd, C. B. & Yeung, N. (2003). Alcohol and error processing. *Trends in Neurosciences,* 26, 402 – 404.

Johannes, S., Wieringa, B. M., Nager, W., Rada, D., Dengler, R., Emrich, H. M., Münte, T. F., & Dietrich, D. E. (2001). Discrepant target detection and action monitoring in obsessive-compulsive disorder. *Psychiatry Research: Neuroimaging Science*, 108, 101-110.

Krigolson, O. E., & Holroyd, C. B. (2007). Predictive information and error processing: the role of medial-frontal cortex during motor control. *Psychophysiology*, 44, 586-595.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition,* 14, 41-104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press, Cambridge, MA.

Levelt, W. J. M., Roelofs, A. & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Science,* 22, 1-75.

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory & Cognition, 7*, 485-495.

Luu, P., Collins, P., & Tucker, D. (2000). Mood, personality, and self-monitoring: negative affect and emotionality in relation to frontal lobe mechanisms of error-monitoring. *Journal of Experimental Psychology: General*, 129, 43-60.

Maess, B., Friederici, A. D., Damian, M., Meyer, A. S., & Levelt, W. J. M. (2002). Semantic category interference in overt picture naming: sharpening current density localization by PCA. *Journal of Cognitive Neuroscience,* 14, 455-462.

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical Selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 33, 503-535.

Masaki, H., Tanaka, H., Takasawa, N., & Yamazaki, K. (2001). Error-related brain potentials elicited by vocal errors. *NeuroReport,* 12, 1851-1855.

Möller, J., Jansma, B. M., Rodriguez-Fornells, A., & Münte, T. F. (2007). What the brain does before the tongue slips. *Cerebral Cortex*, 17, 1173-1178.

Motley, M. T., Camden, C. T., & Baars, B. J. (1982). Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior,* 21, 578-594.

Pailing, P. E., & Segalowitz, S. J. (2004). The error-related negativity as a state and trait measure: motivation, personality, and the ERPs in response to errors. *Psychophysiology*, 41, 84-95.

Postma, A. (2000). Detection of errors during speech production: a review of speech monitoring models. *Cognition, 77*, 97-131.

Postma, A., & Noordanus, C. (1996). Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Language and Speech, 39*, 375-392.

Rodríguez-Fornells, A., Kurzbuch,, A. R., & Münte, T. F. (2002). Time course of error detection and correction in humans: Neurophysiological evidence. *Journal of Neuroscience*, 22, 9990-9996.

Scheffers, M. K., Coles, M. G. H., Bernstein, P. S., Gehring, W. J., & Donchin, E. (1996). Event-related brain potential and error-related processing: an analysis of incorrect responses to go and no-go stimuli. *Psychophysiology,* 33, 42-53.

Scheffers, M. K., & Coles, M. G. H. (2000). Performance monitoring in a confusing world: error-related brain activity, judgments of response accuracy, and types of errors. *Journal of Experimental Psychology, 26*, 141-151.

Schiller, N. O. (2005). Verbal self-monitoring. In A. Cutler (Ed), *Twenty-First Century Psycholinguistics: Four Cornerstones* (pp. 245-261). Mahwah, NJ: Lawrence Erlbaum Associates.

Schiller, N. O. (2006). Lexical stress encoding in single word production estimated by event-related brain potentials. *Brain Research, 1112*, 201-212.

Schiller, N. O., Jansma, B. M., Peters, J., & Levelt, W. J. M. (2006). Monitoring metrical stress in polysyllabic words. *Language and Cognitive Processes, 21*, 112-140.

Schnur, T. T., Schwartz, M. F., Brecher, A., & Hodgson, C. (2006). Semantic interference during blocked-cyclic naming: Evidence from aphasia. *Journal of Memory and Language*, 54, 199-227.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: picture-word interference studies. *Journal of Memory and Language, 26*, 86-102.

Sebastián-Gallés, N., Rodríguez-Fornells, A., De Diego-Balaquer, R., & Díaz, B. (2006). First- and second-language phonological representation in the mental lexicon. *Journal of Cognitive Neuroscience, 18*, 1277-1291.

Vitkovitch, M., & Humphreys, G. W. (1991). Perseverant responding in speeded naming of pictures: it's in the links. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 17, 664-680.

Ullsperger, M., & Von Cramon, D. Y. (2004). Neuroimaging of performance monitoring: error detection and beyond. *Cortex*, 40, 593-604.

Yeung, N. (2004). Relating cognitive and affective theories of the error-related negativity. In M. Ullsperger & M. Falkenstein (Eds.), *Errors, conflicts, and the brain. Current opinions on performance monitoring* (pp. 63-70). Leipzig, Germany: Max Planck Institute of Cognitive Neuroscience.

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, 111, 931-959.

Appendix

A list of the 15 semantic categories and 75 targets used in the current experiment. The approximate English translation is given in brackets.

FURNITURE: bank (couch), tafel (table), stoel (chair), bureau (desk), kruk (stool)

ANIMALS_1: giraffe (giraffe), kameel (camel), hert (deer), olifant (elephant), zebra (zebra)

ANIMALS_2: muis (mouse), slang (snake), vis (fish), eend (duck), libel (dragonfly)

BODYPARTS: been (leg), hand (hand), vinger (finger), voet (foot), arm (arm)

MUSICAL INSTRUMENTS: gitaar (guitar), fluit (flute), trompet (trumpet), harp (harp), piano (paino)

FRUITS: appel (apple), banana (banana), kers (cherry), peer (peer), citroen (lemon)

BIRDS: uil (owl), duif (pigeon), zwaan (swan), ooievaar (stork), pinguin (penguin)

HEADWARE: pet (pet), muts (bonnet), kroon (crown), helm (helm), hoed (hat)

TOOLS: hamer (hammer), zaag (saw), tang (tongs), vijl (file), boor (drill)

VEGETABLES: prei (leek), sla (lettuce), wortel (carrot), tomaat (tomato), ui (onion)

CLOTHING: broek (trousers), rok (skirt), trui (sweater), hemd (shirt), jurk (dress)

FOOD: brood (bread), donut (donut), ei (egg), worst (sausage), kaas (cheese)

VEHICLES: trein (train), auto (car), fiets (bicycle), schip (ship), bus (bus)

KITCHEN UTENSILS: pan (pan), vergiet (colander), rasp (grater), ketel (kettle), wok (wok)

UTENSILS: glas (glass), kop (cup), fles (bottle), schaal (dish), bord (plate)