

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

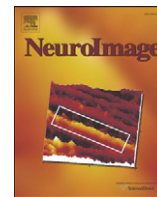
In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

NeuroImage

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

Detection of speech errors in the speech of others: An ERP study

Lesya Y. Ganushchak^{a,*}, Niels O. Schiller^{a,b}

^a Department of Cognitive Psychology, Leiden University-Institute for Psychological Research (LU-IPR), Wassenaarseweg 52, 2333 AK Leiden, The Netherlands

^b Maastricht University, The Netherlands

ARTICLE INFO

Article history:

Received 14 July 2009

Revised 20 November 2009

Accepted 23 November 2009

Available online 1 December 2009

Keywords:

Psycholinguistics

Speech errors

Verbal monitoring

ERP

LAN

PMN

N400

P600

ABSTRACT

The current event-related brain potential study examined the processing of observed speech errors. Participants were asked to detect errors in the speech of others while listening to the description of a visual network. Networks consisted of colored drawings of objects connected by straight or curved lines. We investigated the processing of two types of errors in the network descriptions, i.e., incorrect color and errors in determiners usage (gender agreement violations). In the 100- to 300-ms and 300- to 550-ms time windows, we found larger PMN and N400 amplitudes for both color and determiner error trials compared to correct trials. Furthermore, color but not determiner errors led to larger P600 amplitudes compared to correct color trials. Color errors also showed enhanced P600 amplitudes compared to determiner errors. Taken together, processing erroneous network descriptions elicits different brain potentials than listening to the corresponding correct utterances. Hence, speech is monitored for errors not only during speech production but also during listening to the naturally occurring speech of others.

© 2009 Elsevier Inc. All rights reserved.

Introduction

Speaking is a highly complex cognitive skill which involves the conceptualization of meaning, the retrieval of lexical items from memory, the encoding of phonological word forms, and the articulation of phonetic gestures (for an overview, see [Levelt et al., 1999](#)). Speakers also monitor their own speech output for correctness (for a review, see [Postma, 2000](#)). According to the *Perceptual Loop Theory* of self-monitoring ([Levelt, 1983](#)), the detection of errors in one's own speech is performed in similar ways as error detection in the speech of others, namely by means of the speech comprehension system, where the perceived information is parsed and then transferred to a so-called verbal monitor ([Hartsuiker and Kolk, 2001](#); [Oomen and Postma, 2002](#)). The verbal monitor compares the parsed speech and the intentions of the speaker to the linguistic standards.

Monitoring has been studied primarily in production tasks (for an overview, see [Postma, 2000](#); for an introduction, see [Schiller, 2005](#)). However, relatively little is known about the processing of spontaneously occurring *perceived* errors. In a conversation, it is important to understand exactly what an interlocutor is saying. For instance, when asking for directions, it is essential to know whether to turn *left* or *right*, whether to turn at the *church* or the *town hall*, or whether to look for a brown *gable* or a brown *stable*. Like self-monitoring of

speech production, when listening to and monitoring the speech of others, information from the auditory system is delivered to the speech comprehension system where it is parsed. The parsed information is then transferred to the monitoring system where it is checked for correctness. Therefore, one may assume that self-monitoring and monitoring the speech of others may involve similar processes ([Hartsuiker and Kolk, 2001](#); [Oomen and Postma, 2002](#)).

Considering the several stages in speech production and comprehension processes and the different types of monitoring, one may expect that monitoring for different types of speech errors is dissimilar ([Oomen and Postma, 2002](#)). For instance, [Kutas and Hillyard \(1980\)](#) were the first to demonstrate that reading sentences containing words that were semantically anomalous in a sentence context (e.g., "He spread the warm bread with *socks*") elicit a negative deflection in the ERP, the so-called N400, compared to words that were semantically appropriate in a sentence context (e.g., "He spread the warm bread with *butter*"). A typical N400 effect has a central-parietal maximum and peaks around 400 ms after stimulus onset, can begin as early as 200 ms and can last for as long as 600 ms post-stimulus (e.g., [Kutas and Hillyard, 1980](#)). In the literature, there are at least two ERP components that are elicited in response to syntactic errors: the P600 and the so-called left anterior negativity (LAN). The P600 is a relatively late, centro-parietally distributed positive deflection in the ERP, starting at about 500 ms and typically lasting up to at least 800 ms (e.g., [Osterhout and Holcomb, 1992](#)). An increase in P600 amplitude has been found during reading words that were syntactically anomalous in a sentence context compared to reading words that were syntactically appropriate (e.g., [Hagoort et al., 1993](#)).

* Corresponding author.

E-mail address: lganushchak@fsw.leidenuniv.nl (L.Y. Ganushchak).

An example of a syntactically anomalous sentence is “The spoilt child are throwing the toy on the ground” including a mismatch between the number of the grammatical subject (singular) and the finite verb (plural). In contrast to the P600, the LAN often has left frontal distribution of enhanced negativity associated with the detection of morphosyntactic mismatch (e.g., Coulson et al., 1998; Penke et al., 1997) or is considered a more general index of working memory (e.g., Kluender and Kutas, 1993).

The aim of the current study is to investigate processing of different types of speech errors (i.e., semantic and grammatical errors) in the speech of others. There are numerous studies that successfully demonstrated that the speech error processing is different for semantic and grammatical errors (e.g., Kutas and Hillyard, 1980; Osterhout and Holcomb, 1992). However, most of the previous studies investigated language processing by means of the uni-modal syntactic or semantic violation paradigm, where sentences were presented with unnaturally occurring errors (e.g., “He spread the warm bread with socks”; Kutas and Hillyard, 1980). In the present study, however, we visually presented a series of networks, consisting of colored drawings connected by one, two, or three lines (see also Levelt, 1983; Oomen and Postma, 2001, 2002). These networks were presented simultaneously with previously recorded auditory descriptions of these networks. We used audio recordings of spontaneously spoken descriptions of these networks to make sure that the experimental stimuli resembled naturally occurring speech – and speech errors – as much as possible. Participants were asked to detect these errors by pressing a button as soon as they perceived an error. Different types of speech errors (i.e., incorrect color or wrong determiners; for examples, see Materials section) were present in those descriptions of the networks. In a previous study, Schiller et al. (2009) asked participants to listen to sentences that described simultaneously presented picture. Sentences either correctly described the picture or contained a semantically or phonologically incongruent final word. When spoken sentences were task relevant, the authors found phonological mismatch negativity (PMN) effects for both semantic and phonologically incongruent conditions compared to the correct condition.

This negativity was followed by an N400 that was larger for semantic compared to phonological errors (Schiller et al., 2009). In the present study, we expect to obtain different ERPs for two types of errors. The color errors are semantic in nature and can only be detected when visual and auditory information is being integrated, whereas determiner errors are grammatical agreement errors (between the head noun and its determiner) and reside in the auditory modality alone. Further, we expected that ERPs for error trials would differ from correct trials.

Hypothetically, both color and determiner errors could elicit a PMN, since for both types of errors there is a violation between the expected and the actually heard word. The PMN is a negative-going ERP component that usually has its most negative peak between 150 and 350 ms after stimulus onset and has been associated with the initial phonological processing of a stimulus and is sensitive to a phonological mismatch in the onset of an expected and an actually heard word (Connolly et al., 1995; Connolly and Phillips, 1994). The PMN usually precedes the N400 – as was the case in the Schiller et al. (2009) study – and is independent of the semantic appropriateness of the spoken words (for an overview, see D’Arcy et al., 2004).

Furthermore, we expected to observe N400 and P600 components. For both N400 and P600, we expect to see an enhancement of the components for error trials compared to correct trials. Furthermore, considering that color errors are semantic errors, we expected to find an N400 effect for color errors, and a smaller or no N400 effect for determiner errors, which are grammatical in nature. In contrast, the P600 is typically larger for grammatical compared to semantic errors; therefore, we expected to see a larger P600 for determiner errors compared to color errors. Alternatively, it is possible that the

determiner errors and not the color errors will yield a LAN effect since determiner errors can be considered as grammatical gender violations. Gender violations have been shown to elicit LAN effects between 300 and 500 ms after stimulus onset followed by a P600 (e.g., Deutsch and Bentin, 2001; Gunter et al., 2000; Koester et al., 2004; Koester et al., 2007).

Experimental procedures

Participants

Thirty students of Maastricht University (25 female; average age 21 years; SD = 2.6) 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 with the EEG acquisition hardware, the data of four female participants were lost.

Materials

We used 36 visual networks that consisted of eight colored line drawings depicting simple objects (for an example, see Fig. 1). The objects were colored in purple, yellow, red, grey, blue, and green. The eight objects in the network were linked through one, two, or three lines that could be straight or curved. The straight lines were horizontal, vertical, or diagonal. The curved lines were curved to the left or to the right. An animated red dot moved along the lines through the network following an unpredictable route for 45 s.

Each network was presented simultaneously with a previously recorded auditory route description synchronized to the track of the red moving dot. We obtained the auditory recordings of the route descriptions by asking nine native speakers of Dutch to describe the route of the red dot through the network in Dutch. They were told that the auditory recordings of their descriptions would be played back to listeners, who only received a blank network and were required to fill in the connections and objects of the network following the participants’ description of the route of the red dot. Therefore,

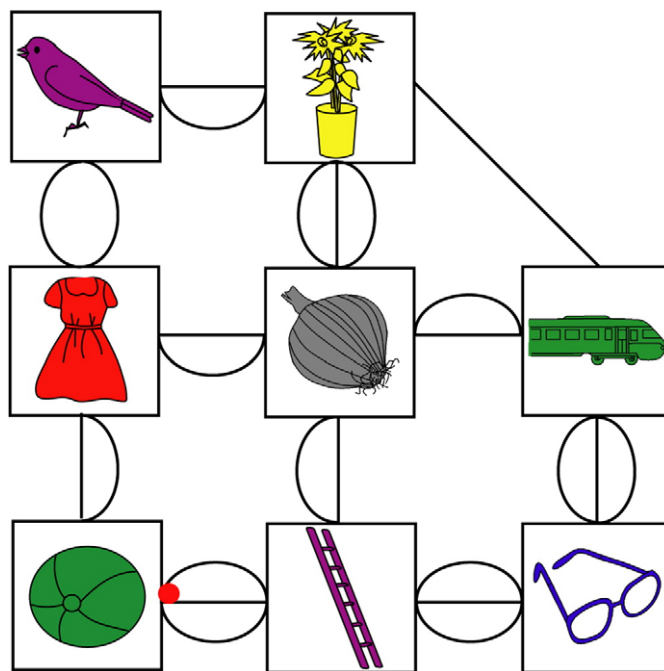


Fig. 1. Example of a visual network used in the experiment.

participants were instructed to describe the route of the dot as accurately as possible, including the object names, the color of the objects, the shape of the line (straight or curved), and the direction of the dot (right, left, horizontal, vertical, or diagonal; e.g., “from a green ball via a straight horizontal line the red dot moves to a purple ladder” [English translation of a possible Dutch route description]). We presented these previously recorded descriptions of the networks to the participants in the present study. Each network description contained naturally occurring speech errors, which were made by the Dutch speakers while describing the network. We classified the occurring speech errors into color errors (e.g., naming an object with an incorrect color) and determiner errors (e.g., saying *de boek* [the_{com} book] instead of *het boek* [the_{neu} book]; *de* and *het* are the gender-marked definite determiners in Dutch; *de* is used for common-gender nouns such as *hamer* “hammer”, *het* for neuter-gender nouns such as *boek* “book”). All types of errors contained partial errors (e.g., *g* [green]...*blue*) and whole errors (e.g., *de* instead of *het*). The ratio of partial and complete errors was comparable between different types of errors; for color errors, there were 4 partial errors and 41 complete errors. For determiner errors, there were 7 partial errors and 38 complete errors. In total, there were 45 errors of each type. Note, that because network descriptions contained naturally occurring speech errors, it was impossible to control for an equal distribution of the number of errors and the different types of errors across networks. Furthermore, network descriptions contained also smaller sets of other errors (e.g., naming an object with an incorrect name). However, due to the relatively low rate of such errors we could not include them separately into the analysis. Correct trials consisted of correct productions of the corresponding words included in error trials. Both error and correct trials were selected from the same speaker.

Design and procedure

Participants were tested individually while seated in a dimly lit soundproof room in front of a computer screen. They were told they were going to see an animation of a red dot moving through a visual network consisting of colored objects that were connected with each other by means of one, two, or three straight or curved lines. Simultaneously to viewing the networks, they were going to listen to other speakers' descriptions of the path of the dot. The networks were presented one by one on the computer screen. The dot ran through the networks for 45 s at a speed compatible with a normal speech rate, without hurrying and without too much pausing. The inter-trial interval was on 5000 ms. Participants were asked to press a button whenever they perceived an error in the description of the network. Participants were not informed about different types of errors. Before participants started the experiment, they received some practice trials including different error types. In total, the experiment lasted about 2 h.

Apparatus and recordings

The EEG was recorded from 29 electrode sites (extended version of the 10/20 system) using tin electrodes mounted onto an electrode cap. The EEG signal was sampled at 250 Hz and band-pass filtered 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 mastoids. Lateral eye movements were recorded using a bipolar montage of two electrodes placed on the right and left external canthus. Vertical eye movements were monitored using a bipolar montage of two electrodes placed above and below the left eye. Impedance level for all electrodes was kept below 5 k Ω .

Data analysis

Epochs from -200 to $+1400$ ms were obtained relative to the stimulus event, including a 100-ms prestimulus baseline. The EEG signal was corrected for ocular artifacts, using the ocular reduction method implemented in the NeuroScan 4.3.1 software. To correct for non-ocular artifacts, epochs with amplitudes above or below 50 μ V were manually rejected. The mean amplitude values were calculated per participant and condition for three time windows: 100–300, 300–550, and 550–900 ms post-stimulus, which were determined after careful visual inspection of the grand average ERP waveforms.

All analyses were performed on erroneous trials and correct trials. Please note that in the present study participants listened to correct and incorrect descriptions of the network. We use the term *error trial* to refer to incorrect descriptions of the network and not to incorrect button-press responses by the participants themselves (i.e., false alarms and misses). Mean button-press reaction times (RTs) and false alarm rates (i.e., indicating an error on an actually correct trial) as well as misses (i.e., failing to indicate an error) from each participant were submitted to repeated-measures ANOVAs only for error trials. Since participants were required to press a button when they perceived an error, there were no button presses for correct trials. The analysis was run separately for false alarms and misses and involved planned comparison with error type (color vs. determiner) as independent variable.

The mean amplitude of the ERP waveforms was submitted to a repeated-measures ANOVA for both error and correct trials with Position (prefrontal, i.e., Fp1, F3, F4, Fp2 vs. frontal, i.e., F7, FC3, FC4, F8 vs. parietal, i.e., TP7, P3, P4, TP8 vs. occipital, i.e., T5, O1, O2, T6; see also Christoffels et al., 2007; Federmeier and Kutas, 1999), Error Type (color vs. determiner), and Description Correctness (incorrect vs. correct). A similar ANOVA was run with Hemisphere (left: F7, F3, FC3, C3, CP3, P7, P3 vs. right: F4, F8, FC4, C4, CP4, P4, P8), Error Type (color vs. determiner), and Description Correctness (incorrect vs. correct). The spatial factors Position and Hemisphere were analyzed in separate analyses in order to include as many electrodes as possible and to keep a symmetrical arrangement. These analyses were performed for the three time windows specified above. Error trials included only recognized errors and not unrecognized errors.

Results

Behavioral data

Table 1 provides an overview of the behavioral results. For reaction times, the analyses revealed no significant effect of Error Type ($F < 1$).

Participants recognized on average 69% of all color and gender errors. A repeated-measures ANOVA with number of false alarms as dependent variable revealed no significant effects ($F < 1$). Similar analyses with number of misses revealed a significant effect of Error Type ($F(1, 25) = 30.15$, $MS_e = 19.02$, $p < .001$). Participants missed more determiner errors than color errors.

Electrophysiological data

Figs. 2 and 3 provide overviews of the stimulus-locked averaged ERP waveforms for error and correct trials across both Error Types.

Table 1
Overview of the behavioral data.

	Color errors	Determiner errors
Reaction times	1985 (506)	2401 (214)
Error rates		
False alarms	12 (4)	11 (5)
Misses	24 (6)	38 (7)

Mean (standard deviation) reaction times (in ms) and percentage of false alarms and misses (standard deviation) as a function of type of error.

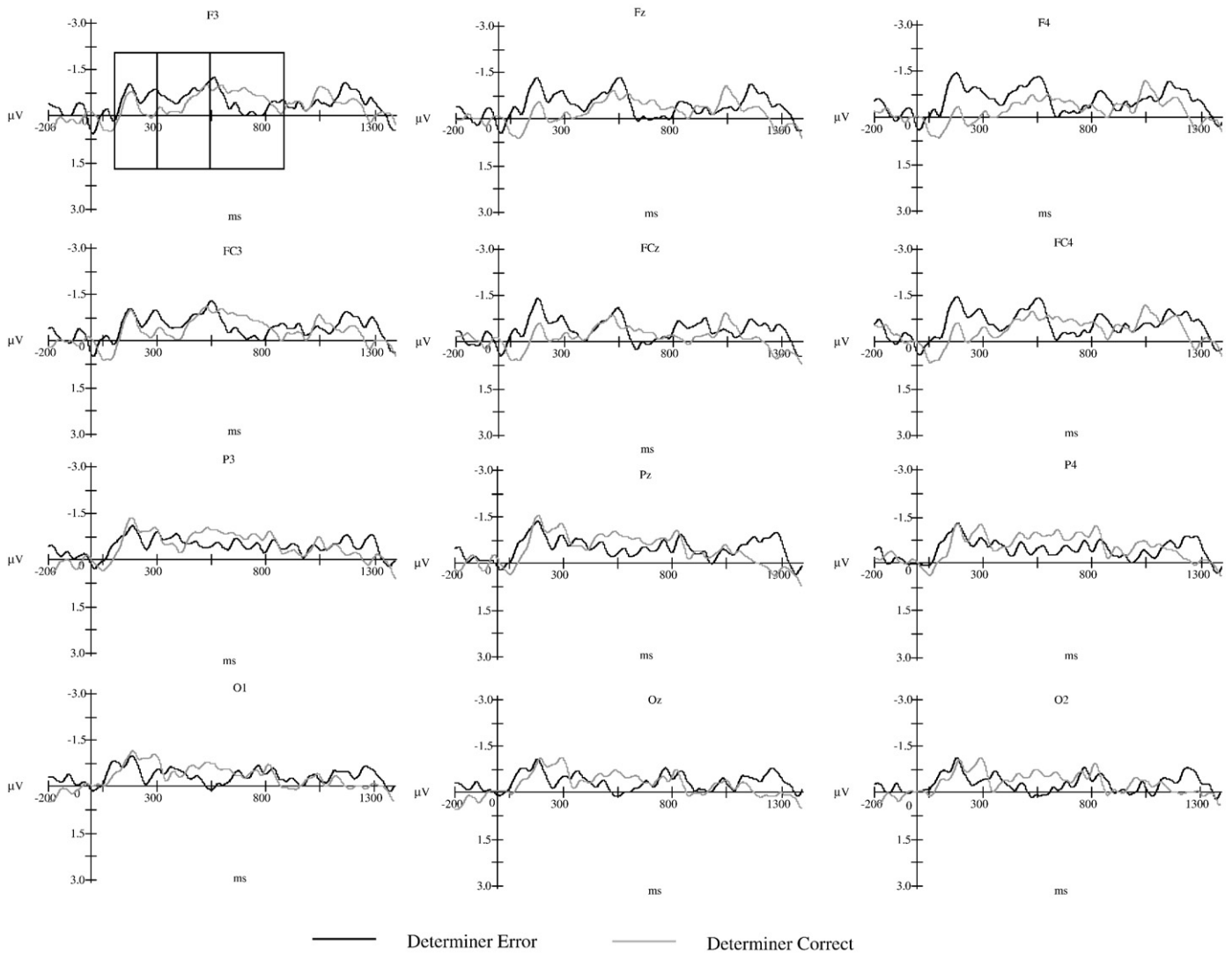


Fig. 2. Grand average event-related potentials for determiner error (black lines) and determiner correct (gray lines) trials. The time windows 100–300, 300–550, and 550–900 ms are framed.

Time window 100–300 ms

In this time window, there was neither main effect of Error Type nor an effect of Description Correctness ($F(1, 25) = 3.17, MS_e = 12.38, n.s.$; $F(1, 25) = 3.77, MS_e = 18.98, n.s.$, respectively). However, there was a significant interaction between Description Correctness and Position ($F(3, 75) = 7.79, MS_e = 2.71, p < .001$). At prefrontal and frontal sites, the amplitude of the ERP waveform was more negative on error trials ($-0.59 \mu V; SD = 1.45$; $-0.52 \mu V; SD = 1.28$, respectively) than on correct trials ($0.25 \mu V; SD = 1.79$; $-0.02 \mu V; SD = 1.52$; $F(1, 25) = 13.39, MS_e = 5.59, p < .001$; $F(1, 25) = 7.38, MS_e = 3.48, p < .01$, respectively). The difference between correct and error trials was not significant at parietal and occipital sites (both $F_s < 1$). Neither of the remaining interactions was significant (all $F_s < 1$). The interactions with Hemisphere were not significant, either (all $F_s < 1$).

Time window 300–550 ms

The analysis showed no significant effects of Description Correctness and Error Type ($F < 1$; $F(1, 25) = 2.99, MS_e = 22.87, n.s.$, respectively). The interaction between Description Correctness and Position was significant ($F(3, 75) = 4.78, MS_e = 2.92, p < .05$). At prefrontal sites, the amplitude of the ERP waveform was more negative on error trials ($-0.76 \mu V; SD = 1.66$) than on correct trials

($0.11 \mu V; SD = 1.89$; $F(1, 25) = 6.86, MS_e = 11.33, p < .05$). At frontal sites, the main effect of Description Correctness was approaching significance ($F(1, 25) = 3.97, MS_e = 6.82, p = .06$). The effect of Description Correctness was absent at parietal and occipital sites (both $F_s < 1$). Neither of the remaining interactions was significant (all $F_s < 1$). The interactions with Hemisphere were not significant, either (all $F_s < 1$).

Time window 550–900 ms

The analysis revealed significant effects of Error Type ($F(1, 25) = 6.80, MS_e = 14.76, p < .05$). Color errors ($0.03 \mu V, SD = 1.50$) were more positive than determiner errors ($-0.46 \mu V, SD = 1.34$). There was no effect of Description Correctness and no interaction between Error Type and Description Correctness ($F < 1$; $F(1, 25) = 1.79, MS_e = 14.86, n.s.$, respectively). However, there was a significant interaction between Description Correctness and Position ($F(3, 75) = 10.03, MS_e = 2.12, p < .001$). At prefrontal and frontal sites, there was no effect of Description Correctness ($F(1, 25) = 1.65, MS_e = 10.24, n.s.$; $F < 1$, respectively). At parietal sites, there was a significant main effect of Description Correctness ($F(1, 25) = 5.69, MS_e = 6.11, p < .05$), which was qualified by a significant interaction between Description Correction and Error Type ($F(1, 25) = 5.38, MS_e = 4.28, p < .05$). For descriptions of color, there was a significant difference between error

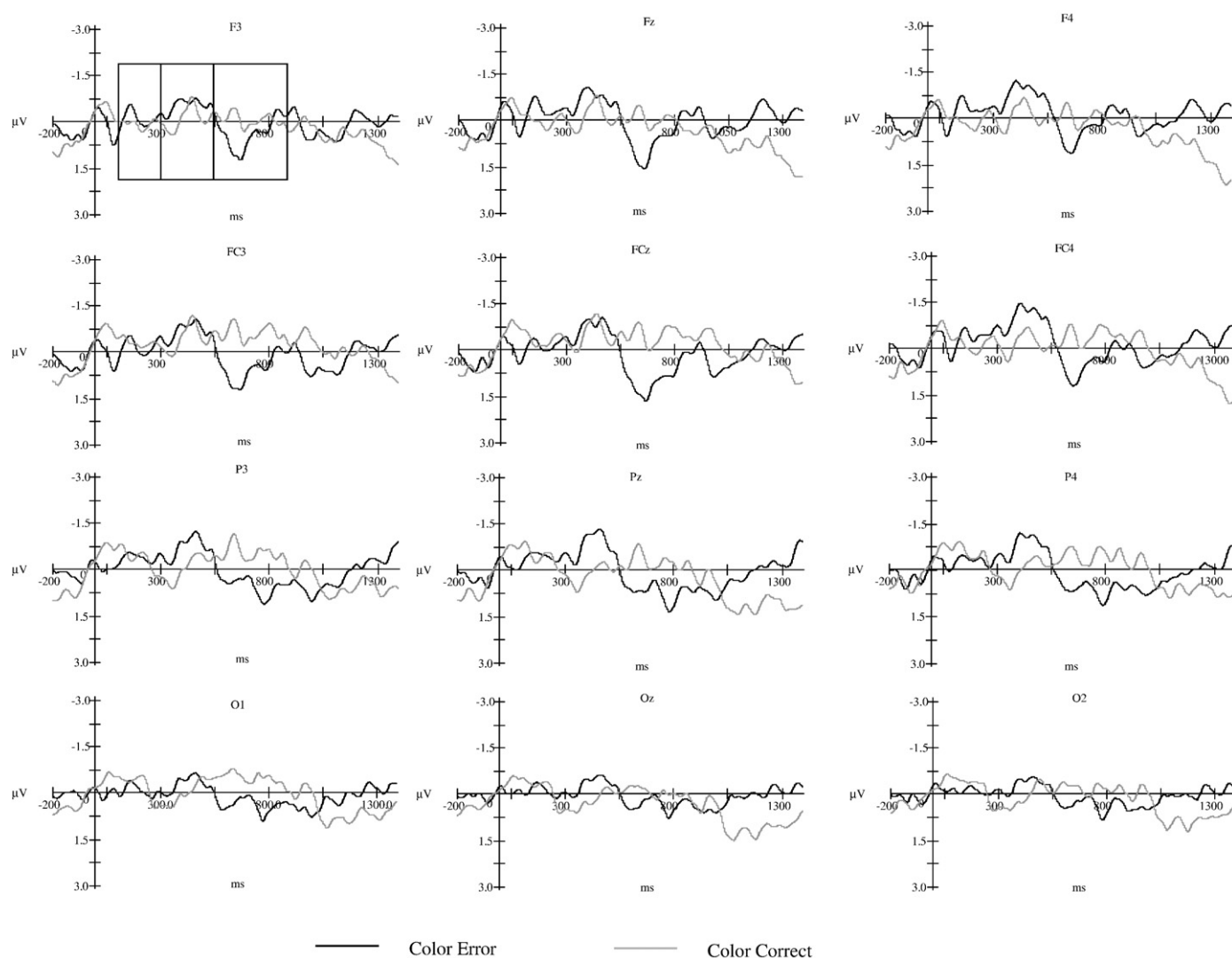


Fig. 3. Grand average event-related potentials for color error (black lines) and color correct (gray lines) trials. The time windows 100–300, 300–550, and 550–900 ms are framed.

trials ($-0.68 \mu\text{V}$, $\text{SD}=1.78$) and correct trials ($0.37 \mu\text{V}$, $\text{SD}=1.13$; $F(1, 25)=7.99$, $\text{MS}_e=7.16$, $p<.01$), while for determiner trials there was no difference between correct and error trials ($F<1$). At occipital sites, error trials ($-0.002 \mu\text{V}$, $\text{SD}=0.93$) were more positive than correct trials ($-0.53 \mu\text{V}$, $\text{SD}=1.44$; $F(1, 25)=4.66$, $\text{MS}_e=6.15$, $p<.05$). The interactions with Hemisphere were not significant (all $F_s<1$).

Discussion

The goal of the present study was to investigate how our cognitive system processes speech errors made by others. Participants listened to descriptions of networks and were required to press a button when they perceived a speech error. There were two types of errors, i.e., color and determiner errors. Participants detected color errors more frequently than determiner errors. Color errors are semantic, whereas determiner errors are grammatical in nature, and as a consequence both error types may differ in saliency. Participants may have paid more attention to semantic than to grammatical errors. This is in line with the results of a study by Oomen and Postma (2002) which showed that participants repaired grammatical errors less frequently than lexical errors. Semantic errors have an immediate effect on the meaning of the description while that may not necessarily be the case for grammatical errors. There are studies demonstrating that people can shift their attention to different aspects of speech and thus

intercept certain types of errors more frequently, depending on what is in the focus of their attention (Motley, 1980; Motley et al., 1981).

In the earliest time window (100–300 ms), we observed a phonological mismatch negativity, which was larger for error trials than correct trials but did not differentiate between two types of errors. This is in accordance with the findings reported in Schiller et al. (2009). Similar PMNs for color and determiner error suggests that the correct word forms have been predicted for both error types. However, this processing similarity is not sustained in later processing steps as reflected in the P600 where color errors and not determiner errors significantly differed from correct trials.

In the 300- to 550-ms time window, we obtained enhanced amplitudes for error trials compared to correct trials. However, contrary to our expectations, determiner and color errors did not differ from each other. It is possible, that this effect reflects an anterior N400 effect. In the present study, we obtained an effect of description correctness in the more anterior electrodes than a typical N400, which usually has a central-parietal maximum. However, in our study, participants saw the entire network, which consisted of various simple line drawings. In the previous literature, it has been shown that pictorial stimuli elicit more anterior N400 effects compared to analogous effects elicited by printed words (for a review, see Van Petten and Luka, 2006). Furthermore, in the study by Schiller et al. (2009), participants were asked in one experiment to listen to descriptions of pictures and perform a language-irrelevant task. In a

second experiment, participants were presented with the identical stimuli, but asked to press a button when they detected a verbal error. Schiller and his colleagues showed that the locus of the N400 effect changed from a more posterior distribution in the first experiment to a more anterior one in the second experiment. Suggestively, such a shift in localization of the N400 effect can be attributed to the fact that language comprehension and production are related to activity in frontal cortex, particularly when a semantic judgment is of relevance (Bookheimer, 2002; Gernsbacher and Kaschak, 2003; Schiller et al., 2009). Moreover, it has previously been shown that listeners can anticipate upcoming words, and this anticipation affects the N400 effect (DeLong et al., 2005; Van Berkum et al., 2005). In the present study, presenting the entire network at once furthermore allowed participants to anticipate the words they were possibly going to hear with the picture by translating the visual context into an expected phonological form for the upcoming verbal description, which in turn could account for the relatively early onset of our N400 effect (see Van Petten et al., 1999). Hearing a determiner or color adjective, which is discrepant with what was expected, might cause difficulties to integrate information and hence elicit an N400.

Alternatively, this frontal effect may be considered a LAN. In the previous literature, it has been shown that LAN is elicited by syntactic violations (e.g., Coulson et al., 1998; Koester et al., 2004). In the present study, effects in the 300- to 550-ms window were present for both color and determiner errors. Considering that color errors were unlikely to be perceived as syntactic violations and that the observed effect was not localized in the left hemisphere – if anything amplitudes were larger at the right hemisphere (see Figs. 2 and 3) – we do not believe that the effect we found in the present study should be classified as a LAN. Interestingly, however, it might be that participants perceived both types of errors as an interruption of the natural flow of speech, which could have resulted in so-called right anterior negativity (RAN). A RAN is usually observed in response to prosodic and musical violations (e.g., Eckstein and Friederici, 2005; Koelsch, 2009).

In the latest time window, i.e., 550–900 ms, we also obtained an effect of correctness of the utterance, which can be interpreted as a P600 effect. Error trials were more positive than correct trials. However, this was only the case for color but not determiner trials. Surprisingly, color errors had a more positive deflection when compared to determiner errors. There was no difference in amplitude of ERP waveforms between correct color and determiner trials.

As stated above, we showed that color errors had larger P600 than determiner errors. Originally, the P600 effect was associated with syntactic and semantic violations. Recently, it has been suggested that a monitoring process in language comprehension may also be reflected by a P600 (Kolk et al., 2003; Van Herten et al., 2005; Vissers et al., 2006, 2008). According to this account, the P600 reflects a process of re-analysis in language perception. The purpose of this re-analysis is to prevent false information from entering the discourse representation. The verbal monitor detects conflict between different analyses of the same linguistic string, and this conflict in turn triggers the re-analysis of the utterance. In other words, the activation of two incompatible interpretations leads to a conflict. This conflict triggers a process of re-analysis and hence a P600 is generated reflecting a check for the possibility of a processing error. The function of the re-analysis is to resolve conflict by monitoring the input for possible processing errors. In the present study, it may be possible that the P600 effect was triggered by a conflict between the predicted representation based on the presented network and the presented description of the network. Furthermore, the P600 may reflect monitoring processes not only at the sentence but also at the word level (Vissers et al., 2006).

In the present study, we demonstrated that color errors elicited P600, while determiner errors did not. Color errors could have been detected by combining visual and auditory information, while detection of determiner errors relied purely on availability of auditory

information. Therefore, by seeing the networks, participants could have formed expectations about what color they could expect to hear, while expectations about determiner forms would have required retrieval of concept of the word and its corresponding gender. This is in accordance with our behavioral results, i.e., determiner errors were missed more often than color errors. When an erroneous description of the color was presented, it resulted in a greater mismatch between an expected word and the actual word that was presented. Based on the monitoring account of the P600, the less an event is expected, the higher the chance that this event occurred due to an error (Vissers et al., 2006, 2008). Thus, violation of a more expected outcome (i.e., color) led to higher conflict between expected and presented representation and thus to a greater P600 effect.

One potential problem of our study may be that participants were required to press a button with their right hand when they detected a speech error, but not when a correct description was presented. One may argue that this may be a potential confound in the sense that observed effects on incorrect descriptions may also reflect participants' motor response preparation to generate the button press. However, one should note that in the present study, button-press latencies were very long (about 2000 ms for color errors and even longer for determiner errors; see Table 1). In contrast, the ERP effects we report here occurred 1000 ms or more before the button presses. Furthermore, the differences between the ERP responses to the two error types cannot be explained by motor response preparation since response preparation should be the same for both error types. Therefore, it is highly unlikely that the ERP effects observed in the present study were confounded with the preparation of the motor response.

To conclude, we showed that listening to erroneous responses elicits different brain potentials than listening to the correct utterances. This suggests that the speech is monitored for errors not only during speech production but during speech comprehension as well. Whereas there are several published studies demonstrating monitoring processes at the level of speech perception (i.e., reading; Kolk et al., 2003; Vissers et al., 2006, 2008), the current study is among the first to demonstrate the elicitation of an ERP component reflecting error monitoring in auditory language comprehension processes on the word level during spontaneously occurring errors.

Acknowledgments

The work presented in this manuscript was supported by NWO grant no. 453-02-006 to Niels O. Schiller. The authors thank Lemke Horemans and Maya Visser for the data acquisition at Maastricht University and two anonymous reviewers for their helpful comments on an earlier version of this manuscript. Correspondence concerning this article should be addressed to Lesya Ganushchak, Department of Cognitive Psychology, Leiden University-Institute for Psychological Research (LU-IPR), Wassenaarseweg 52, 2333 AK Leiden, The Netherlands; Email: lganushchak@fsw.leidenuniv.nl.

References

- Bookheimer, S., 2002. Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annu. Rev. Neurosci.* 25, 151–188.
- Christoffels, I.K., Firk, C., Schiller, N.O., 2007. Bilingual language control: an event-related brain potential study. *Brain Res.* 1147, 192–208.
- Coulson, S., King, J.W., Kutas, M., 1998. Expect the unexpected: event-related brain response to morphosyntactic violations. *Lang. Cogn. Processes* 13, 21–58.
- Connolly, J.F., Phillips, N.A., 1994. Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *J. Cogn. Neurosci.* 6, 256–266.
- Connolly, J.F., Byrne, J.M., Dywan, C.A., 1995. Assessing adult receptive vocabulary with event-related potentials: an investigation of cross-modal and cross-form priming. *J. Clin. Exp. Neuropsychol.* 17, 548–565.
- D'Arcy, R.C.N., Connolly, J.F., Service, E., Hawco, C.S., Houlihan, M.E., 2004. Separating phonological and semantic processing in auditory sentence processing: a high resolution event-related brain potential study. *Hum. Brain Mapp.* 22, 40–51.

- DeLong, K.A., Urbach, T.P., Kutas, M., 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nat. Neurosci.* 8, 1117–1121.
- Deutsch, A., Bentin, S., 2001. Syntactic and semantic factors in processing gender agreement in Hebrew: evidence from ERPs and eye movements. *J. Mem. Lang.* 45, 200–224.
- Eckstein, K., Friederici, A.D., 2005. Late interaction of syntactic and prosodic processes in sentence comprehension as revealed by ERPs. *Cogn. Brain Res.* 25, 130–143.
- Federmeier, K., Kutas, M., 1999. A rose by any other name: long-term memory structure and sentence processing. *J. Mem. Lang.* 41, 469–495.
- Gernsbacher, M.A., Kaschak, M.P., 2003. Neuroimaging studies of language production and comprehension. *Annu. Rev. Psychol.* 54, 91–114.
- Gunter, T.C., Friederici, A.D., Schriefers, H., 2000. Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *J. Cogn. Neurosci.* 12, 556–568.
- Hagoort, P., Brown, C.M., Groothusen, J., 1993. The syntactic positive shift (SPS) as a measure of syntactic processing. *Lang. Cogn. Processes* 8, 439–483.
- Hartsuiker, R.J., Kolk, H.H.J., 2001. Error monitoring in speech production: a computational test of the perceptual loop theory. *Cogn. Psychol.* 42, 113–157.
- Kluender, R., Kutas, M., 1993. Bridging the gap: evidence from ERPs on the processing of unbounded dependencies. *J. Cogn. Neurosci.* 5, 196–214.
- Koelsch, S., 2009. Music-syntactic processing and auditory memory: similarities and differences between ERAN and MMN. *Psychophysiology* 46, 179–190.
- Koester, D., Gunter, T.h.C., Wagner, S., Friederici, A.D., 2004. Morphosyntax, prosody, and linking elements: the auditory processing of German nominal compounds. *J. Cogn. Neurosci.* 16, 1647–1668.
- Koester, D., Gunter, T.h.C., Wagner, S., 2007. The morphosyntactic decomposition and semantic composition of German compound words investigated by ERPs. *Brain Lang.* 102, 64–79.
- Kolk, H.H.J., Chwilla, D.J., Van Herten, M., Oor, P.J., 2003. Structure and limited capacity in verbal working memory: a study with event-related potentials. *Brain Lang.* 85, 1–36.
- Kutas, M., Hillyard, S.A., 1980. Reading senseless sentences: brain potentials reflect semantic anomaly. *Science* 207, 203–205.
- Levelt, W.J.M., 1983. Monitoring and self-repair in speech. *Cognition* 14, 41–104.
- Levelt, W.J.M., Roelofs, A., Meyer, A., 1999. A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–75.
- Motley, M.T., 1980. Verification of “Freudian slips” and semantic prearticulatory editing via laboratory-induced spoonerisms. In: Fromkin, V.A. (Ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand*. Academic Press, New York, pp. 133–147.
- Motley, M.T., Baars, B.J., Camden, C.T., 1981. Syntactic criteria in prearticulatory editing: Evidence from laboratory-induced slips of the tongue. *J. Psycholinguist. Res.* 10, 503–522.
- Oomen, C.C.E., Postma, A., 2001. Effects of time pressure on mechanisms of speech production and self-monitoring. *J. Psycholinguist. Res.* 30, 163–184.
- Oomen, C.C.E., Postma, A., 2002. Limitations in processing resources and speech monitoring. *Lang. Cogn. Processes* 17, 163–184.
- Osterhout, L., Holcomb, P.J., 1992. Event-related brain potentials elicited by syntactic anomaly. *J. Mem. Lang.* 31, 785–806.
- Penke, M., Weyerts, H., Gross, M., Zander, E., Münte, T.F., Clahsen, H., 1997. How the brain processes complex words: an event-related potential study of German verb inflections. *Cogn. Brain Res.* 6, 37–52.
- Postma, A., 2000. Detection of errors during speech production: a review of speech monitoring models. *Cognition* 77, 97–131.
- Schiller, N.O., 2005. Verbal self-monitoring. In: Cutler, A. (Ed.), *Twenty-First Century Psycholinguistics: Four Cornerstones*. Mahwah, NJ, Lawrence Erlbaum Associates, pp. 245–261.
- Schiller, N.O., Horemans, I., Ganushchak, L., Koester, D., 2009. Event-related potentials during the monitoring of speech errors. *NeuroImage* 44, 520–530.
- Van Berkum, J.J.A., Brown, C.M., Zwitserlood, P., Kooijman, V., Hagoort, P., 2005. Anticipating upcoming words in discourse: evidence from ERPs and reading times. *J. Exper. Psychol., Learn., Mem., Cogn.* 31, 443–467.
- Van Herten, M., Kolk, H.H.J., Chwilla, D.J., 2005. An ERP study of P600 effects elicited by semantic anomalies. *Cogn. Brain Res.* 22, 241–255.
- Van Petten, C., Luka, B.J., 2006. Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain Lang.* 97, 279–293.
- Van Petten, C., Coulson, S., Rubin, S., Plante, S., Parks, M., 1999. Time course of word identification and semantic integration in spoken language. *J. Exper. Psychol., Learn., Mem., Cogn.* 25, 394–417.
- Vissers, C.T.h.W.M., Chwilla, D.J., Kolk, H.H.J., 2006. Monitoring in language perception: the effects of misspellings of words in highly constrained sentences. *Brain Res.* 1106, 150–163.
- Vissers, C.T.h.W.M., Kolk, H.H.J., Van de Meerendonk, N., Chwilla, D.J., 2008. Monitoring in language perception: evidence from ERPs in a picture-sentence matching task. *Neuropsychologia* 46, 967–982.