Electrophysiological correlates of monitoring in typing with and without visual feedback

Svetlana Pinet[a] and Nazbanou Nozari[a,b]

[a] Department of Neurology, Johns Hopkins University, 1629 Thames Street Suite 350, Baltimore, MD 21231, USA

[b] Department of Cognitive Science, Johns Hopkins University, 237 Krieger Hall, 3400 N. Charles Street, Baltimore, MD 21218, USA

Corresponding author: Svetlana Pinet, svetlana.pinet@gmail.com

Abstract

New theories of monitoring in language production, regardless of their mechanistic differences, all posit monitoring mechanisms that share general computational principles with action monitoring. This perspective, if accurate, would predict that many electrophysiological signatures of performance monitoring should be recoverable from language production tasks. In this study, we examined both error-related and feedback-related EEG indices of performance monitoring in the context of a typing-to-dictation task. To disentangle the contribution of the external from internal monitoring processes, we created a condition where participants immediately saw the word they typed (the immediate-feedback condition) vs. one in which displaying the word was delayed until the end of the trial (the delayed-feedback condition). The removal of immediate visual feedback prompted a stronger reliance on internal monitoring processes, which resulted in lower correction rates and a clear Error-related Negativity (ERN). Compatible with domain-general monitoring views, a Pe was only recovered under conditions where errors were detected or had a high likelihood of being detected. Examination of the feedback-related indices (feedback-related negativity, FRN, and fronto-central positivity, FCP) revealed a two-stage process of integration of internal and external information. The recovery of a full range of well-established EEG indices of action monitoring in a language production task strongly endorses domain-general views of monitoring. Such indices, in turn, are helpful in understanding how information from different monitoring channels are combined.

Keywords: Language production, error monitoring, ERN, FRN, Pe

1. Introduction

When speakers make speech errors, they are usually able to notice and correct them. The most obvious way to detect an error is to hear the word that was spoken and realize that it does not match the word that was intended. This process involves monitoring of the auditory output of speech, also called monitoring through the *external channel* (e.g., Hartsuiker & Kolk, 2001; Levelt, 1983). The external channel alone, however, is not sufficient to explain fast and efficient repairs often observed in language production, which has led to the proposal of an additional *internal channel* (Levelt, 1983). Over the past decades, various mechanisms have been proposed for how the internal channel functions, including comprehension-based monitors (e.g., Levelt, 1983, 1989), production-based (e.g., Laver, 1980; Nozari, Dell, & Schwartz, 2011; Van Wijk & Kempen, 1987), and forward models (e.g., Hickok, 2012; Pickering & Garrod, 2013; Tourville, Reilly, & Guenther, 2008). Despite their differences, all of these accounts propose mechanisms capable of detecting an error before it becomes an external stimulus for processing by the perceptual systems (see Nozari & Novick, 2017 and Postma, 2000 for reviews).

In parallel, there has been increasing evidence that monitoring in language production involves many of the same brain regions involved in monitoring action in other domains, and perhaps follows the same domain-general computations (Nozari & Novick, 2017). It is, therefore, reasonable to expect generally similar electrophysiological indices for language monitoring and monitoring in other domains. This paper combines the two approaches, by investigating the EEG indices of action monitoring in language production with and without the contribution of the external channel. Typing is chosen instead of spoken production because a) precise time-locking of EEG waveforms to the production of single segments (i.e., keystrokes) is possible in typed, but not in spoken, production, and b) removal of the visual output at the time of production is also possible in typed but not in spoken production. The results provide a first comprehensive view of the EEG correlates of monitoring with and without feedback in typing, and may have implications for monitoring in language production in general.

1.1. Monitoring through the Internal vs. the External Channel

In spoken production, when auditory processing of the production output is blocked by noise, speakers are still capable of detecting a large number of their errors (e.g., Hartsuiker, Kolk, & Martensen, 2005; Lackner & Tuller, 1979; Postma & Kolk, 1992). Some studies have

reported lower correction rates under noise-masked conditions (Oomen & Postma, 2001; Postma & Kolk, 1992), but this finding has not always been replicated (Nooteboom & Quené, 2017). The most striking demonstration of the involvement of the external channel in speech monitoring was provided by a study in which words produced in a Stroop task were occasionally replaced by other words (e.g., "gray" for "green") through the manipulation of auditory feedback (Lind, Hall, Breidegard, Balkenius, & Johansson, 2014). The authors reported that, under ideal playback conditions (short lag between production and playback), participants accepted the manipulated word on 68% of trials, thereby admitting to an error they had not made solely based on the information provided by the external channel. These results suggest an interesting division of labor between the internal and external monitoring channels, but noise masking in spoken production has two problems. First, it does not fully block the processing of speech output through the perceptual system because of bone conduction, unless the noise is very loud. Loud noise, on the other hand, may change monitoring performance for reasons other than simply blocking the auditory input, such as distracting or distressing the speaker. Second, the impaired hearing induced by noise masking activates the Lombard reflex (Lane & Tranel, 1971), causing alterations to speech such as the use of louder voice with higher fundamental frequency and slowed articulation.

Typing, on the other hand, is not subject to the same problem: visual feedback can be easily delayed without adverse effects. Logan & Crump (2010) manipulated visual feedback in the same way as explained above for the Lind et al.'s (2014) study, in a copy-typing task. Similar to Lind et al.'s results, the authors found that typists tended to accept the manipulated output as their own production, pointing to the influence of the external monitoring channel. This acceptance, however, was not complete, and a good proportion of manipulated trials were detected as having been tampered with, when participants were given the option to label them as such. This finding shows that the internal channel is capable of overriding the external channel, and that the artificial manipulation of feedback may induce unusual behavior in monitoring. Simple removal of visual feedback has also been conducted in typing, but the experiments either did not allow for natural corrections (e.g., Snyder, Logan, & Yamaguchi, 2015), or did not have a comparison with typing with visual feedback (Kalfaoğlu & Stafford, 2014). Moreover, all these studies have used copy typing, which does not require retrieving the orthographic word form from memory (Bonin, Méot, Lagarrigue, & Roux, 2015). In the current study, we used the same general logic as noise-masked studies in speech production and compared typing with and without visual feedback, in a spelling to dictation task in which participants were allowed to type (and correct) like they would in everyday life.

It is important to note that the label "external channel" is being used here to refer to the mechanisms that process the final outcome of production; that is the auditory word in spoken production and the visual word in typed production. We do not mean to imply that these are the only perceptual consequences of production available for monitoring. In fact, the mere act of motor production creates movements with perceptual consequences that are processed by the proprioceptive system. Most monitoring accounts, however, maintain that monitoring through proprioception involves an internal mechanism (e.g., a forward model of estimating the perceptual consequences; e.g., Guenther, 2016), making it distinct from processing of the final output. Moreover, blocking noise in the studies reviewed above does not remove the proprioception of speaking the word, yet it has consequences for monitoring, showing the independent contribution of what has been called the external channel. This work is focused on exploring the contribution of this channel to monitoring in typing.

1.2. Neural Correlates of Monitoring

While the classic theory of speech monitoring, the perceptual loop (Levelt, 1983; 1989), posited a special mechanism for the internal channel, i.e., monitoring of inner speech through the comprehension system, the more recent theories of monitoring have shifted more towards a domain-general perspective (Hickok, 2012; Nozari et al., 2011; Pickering & Garrod, 2013; Tourville et al., 2008). Even though these theories differ in the precise mechanism(s) through which monitoring is accomplished, which may, in part, be due to the differences in the scope of the models (e.g., motor production vs. lexical-semantic processing), they all apply the same computations involved in monitoring in the action domain to language production. Part of the motivation for the emergence of these domain-general views of monitoring has been the discovery of the shared neural correlates between language and action monitoring. Some of these areas are the anterior cingulate cortex (ACC; Debener et al., 2005; Gehring & Willoughby, 2004; Herrmann, Römmler, Ehlis, Heidrich, & Fallgatter, 2004), the posterior cingulate cortex (Charles, Van Opstal, Marti, & Dehaene, 2013), the supplementary motor area (SMA; Bonini et al., 2014), and perhaps the prefrontal cortex, insula, cerebellum, thalamus and basal ganglia (Bohland, Bullock, & Guenther, 2010; Bourguignon, 2014; Christoffels, Formisano, & Schiller, 2007; Fedorenko, Duncan, & Kanwisher, 2013; Fiez, 2016; Marvel & Desmond, 2010; Riès, Xie, Haaland, Dronkers, & Knight, 2013).

The involvement of domain-general regions in language monitoring generates the expectation that the ERP signatures of monitoring should be, at least to some extent, shared between language and action monitoring. ERP indices of monitoring fall under the two

general groups of *error-related* and *feedback-related* components. Error-related indices include two major components: the error-related negativity (ERN or Ne; Gehring, Goss, Coles, Meyer, & Donchin, 1993), and the error positivity (Pe; Falkenstein, Hohnsbein, Hoormann, & Blanke, 1991). The feedback-related indices include the feedback-related negativity (FRN; Miltner, Braun, & Coles, 1997) and the more recently described fronto-central positivity (FCP; Butterfield & Mangels, 2003). In keeping with the domain-generality prediction, at least one of these components, the ERN, has been recovered in picture naming tasks (e.g., Ganushchak & Schiller, 2008; Riès, Janssen, Dufau, Alario, & Burle, 2011; see Nozari & Pinet, in press). The rest of the components remain understudied in language production. Below, we briefly review these components and their interpretations in the non-linguistic literature. We then look for the same components in typed production.

1.2.1. ERN

The ERN is a negative component with a fronto-central distribution, arising in a 100ms window around an error. Although the ERN's magnitude can be sensitive to awareness, it is also found on trials without conscious awareness of the errors (Wessel, 2012), suggesting that it is an early and automatic process. Its magnitude is comparable to correct trials when participants are unaware of their errors and classify them as correct with high confidence (Shalgi & Deouell, 2012), suggesting that the ERN arises from a mechanism which estimates the *likelihood* of an error; if such likelihood is estimated (correctly or incorrectly) to be close to zero, no prominent ERN is observed. Especially relevant for our purposes, the ERN has been reported in tasks using linguistic material, such as semantic categorization (Balass, Halderman, Benau, & Perfetti, 2016), lexical decision (Ito & Kitagawa, 2006; Sebastian-Gallés, Rodríguez-Fornells, De Diego-Balaguer, & Díaz, 2006), phoneme monitoring (Ganushchak & Schiller, 2008a), as well as picture naming (Ganushchak & Schiller, 2008b; Riès et al., 2011, 2013) and tongue twister tasks (Acheson & Hagoort, 2014; Möller, Jansma, Rodríguez-Fornells, & Münte, 2007). The ERN is also found in high-conflict conditions, in the absence of overt errors (e.g., Acheson, Ganushchak, Christoffels, & Hagoort, 2012).

All these studies, however, reported an ERN locked to the vocal onset of the word but not to each segment (i.e., phoneme or letter), even though the errors on phonemes or letters may reflect selection problems at the level of segments. To our knowledge, only one study has reported an ERN time-locked to each segment in language production: Kalfaoğlu, Stafford, & Milne (2018) reported higher amplitude of the ERN on corrected compared to uncorrected letters in a sentence copying task. However, they reported an ERN on a

component isolated by ICA (Independent Component Analysis), which was itself selected based on its activity on corrected trials. The choice of the ICA component based on a condition that was later included in the comparison creates a bias that may affect the findings.

1.2.2. Pe

Following the ERN, a centro-parietal component called the error-related positivity (Pe) arises from 100 ms to 300 ms post-error (Falkenstein et al., 1991; Gehring et al., 1993). Unlike the ERN, the Pe does not seem to be generated unless participants are consciously aware of having made an error (Nieuwenhuis, Ridderinkhof, Blom, Band, & Kok, 2001; O'Connell et al., 2007). More recently, Boldt and Yeung (2015) reported a graded modulation of the Pe by confidence ratings in a perceptual decision-making task, which ties the Pe closely to metacognitive judgments, further indicating that it is an index of conscious error detection. Despite their frequent co-occurrence —at least on trials where participants are aware of having made an error— evidence suggests that the ERN and Pe are separable components, which arise from different regions (ERN from SMA and caudal ACC, and Pe from rostral ACC; Herrmann et al., 2004), and may represent different error monitoring processes or systems (Herrmann et al., 2004; Hughes & Yeung, 2011; Nieuwenhuis et al., 2001).

While there are numerous reports of the ERN in linguistic tasks, explicit investigations of the Pe are scarce. Masaki, Tanaka, Takasawa, & Yamazaki (2001) reported a higher positive amplitude following errors than correct trials with a centro-parietal topography but in a later time-window than the usual Pe, while Riès et al. (2011) failed to find such an effect. Kalfaoğlu et al. (2018) reported a positive component of higher amplitude for corrected keystrokes than uncorrected and correct keystrokes in typing. This component followed the ERN on the same ICA-isolated component and could be an early Pe. However, the analysis used by the authors imposed the strict constraint that the two components would have a similar brain localization, while previous evidence rather suggested that the ERN and Pe might arise from different regions. Looking at the full brain activity is thus necessary for evaluating the presence of a centro-parietal Pe. Whether the Pe can index conscious detection of errors in a generative task such as language production in the same way as forced choice tasks remains to be established.

1.2.3. FRN

The FRN is a feedback-locked index of performance monitoring, observed over fronto-central electrodes, with higher amplitude following negative than positive feedback (Miltner et al.,

1997). The amplitude of the FRN is higher the more informative the feedback is (Arbel, Goforth, & Donchin, 2013). In tasks that require explicit associations (e.g., paired-associates learning task), the amplitude of the FRN is positively correlated with future learning, such that items that generated larger FRNs were better recognized later (Arbel et al., 2013; Luft, 2014). The ERN and FRN have many similarities (e.g., Van Schie, Mars, Coles, & Bekkering, 2004; see Holroyd & Coles, 2002 for a review), and they are thought to index internal and external monitoring, respectively. In line with this view, in the course of learning a new task, the amplitude of the ERN increases while the amplitude of the FRN decreases – indicating a shift from reliance on external feedback towards monitoring based on internal information processing (Eppinger, Kray, Mock, & Mecklinger, 2008; Holroyd & Coles, 2002). As far as we know, no studies have looked at the FRN in the context of language production.

1.2.4. FCP

An under-reported component, called the FCP, follows the FRN after feedback presentation and is associated with long-term learning (Arbel et al., 2013; Arbel & Wu, 2016; Butterfield & Mangels, 2003). In a paired-associate learning paradigm, Arbel and Wu (2016) showed that larger FCP amplitude to negative feedback was predictive of better learning at the end of the task. In a two-session semantic knowledge task, Butterfield and Mangels (2003) found that the amplitude of the FCP on the first session was correlated with the rate of correction on the second session, pointing again to the relation between the FCP and learning. Similar to the FRN, we are unaware of any studies that have reported the FCP in language production.

1.3. The Current Study

This study investigates the monitoring of segmental errors, i.e., any kind of error that involves a segment (a letter) in typing. If there had been multiple errors in a word, only the first error was analyzed to avoid contamination from earlier post-error processes on later errors. The Appendix includes a definition of error types with examples, as well as additional information about specific error types in this study. As explained earlier, using typing instead of speaking has the advantage of easily blocking external (i.e., visual) feedback. In addition, it is possible to easily time lock ERPs to individual keystrokes. A third advantage is that we could elicit a fair number of errors in a natural task (typing-to-dictation), which allows us to better extend the findings to processes involved in everyday typing. The results, however, may have implications beyond the typing modality. A large body of research has shown the similarities between production patterns across different output modalities such as spoken, hand-written

and typed production (e.g., Breining, Nozari, & Rapp, 2016; Damian & Freeman, 2008; Muscalu & Smiley, 2018; Nozari, Freund, Breining, Rapp, & Gordon, 2016; Pinet, Ziegler, & Alario, 2016). A recent study in our lab has specifically demonstrated largely similar patterns of lexical and segmental errors in spoken and typed production, despite differences such as sensitivity to phonotactic constraints which is, unsurprisingly, much more prominent in spoken production (Pinet & Nozari, 2018). These similarities suggest similar underlying mechanisms for language production at the stages of lexical selection and segmental encoding across different modalities. Together with the domain-general perspective on monitoring embraced by various accounts of monitoring in language production, the findings of this study could thus shed light not only on monitoring during typing but more broadly on monitoring mechanisms in language production in general.

In a simple typing-to-dictation task, participants typed the words they heard in two conditions. In the immediate-feedback condition, they saw what they typed on the screen in real time. In the delayed-feedback condition, the screen remained blank while they typed and what they had typed appeared with a delay at the end of the trial. The delayed-feedback condition also included a metacognitive judgment, i.e., awareness of errors: after they typed the word but before seeing it on the screen, participants were asked to indicate whether they believed they had made an error or not. Comparison of correction rates in the immediate- and delayed-feedback conditions allows us to evaluate the contribution of the external channel to segmental corrections. Since missed corrections may reflect errors that were not detected, we also compare the rate of corrections to the rate of error awareness in the metacognitive judgment task, in the delayed-feedback condition. A low correction rate despite a higher rate of error awareness would suggest that errors had been detected but not corrected.

EEG analysis investigated the standard indices of action monitoring time-locked to each keystroke. Since the ERN indexes internal monitoring, we expect to find it when the main channel available is the internal channel (i.e., the delayed-feedback condition). Its presence in the immediate-feedback condition might depend on the contribution of internal monitoring when external monitoring is also available. If the Pe indexes conscious detection of an error in language production, similar to non-linguistic tasks, we would expect to observe it when an error is detected. The FRN and FCP can be investigated in the delayed-feedback condition, time-locked to the delayed presentation of feedback. Crucially, this condition allows us to understand how the system combines internal and external feedback.

2. Methods

2.1. Participants

Seventeen right-handed native English speakers (10 females, age = 25.4±4.5 years old) took part in this study. They were consented under a protocol approved by the Institutional Review Board of Johns Hopkins Medicine, and were compensated for their time.

2.2. Stimuli and Apparatus

We selected 352 seven- and eight-letter nouns from the English Lexicon Project database (Balota et al., 2007). Plural and compound words, and words that had homophones, phonological or orthographic neighbors were not included in the material. Words were divided into two lists of 176 words, balanced on psycholinguistic (word frequency, length, number of syllables and phonemes, mean bigram frequency) as well as motoric variables (laterality of the first letter on the keyboard, percentage of bigrams typed with two hands). Words were recorded by a female native speaker of American English. Participants' responses were recorded using a DirectIN PCB v2016 Empirisoft keyboard, which provides millisecond accuracy for keystrokes.
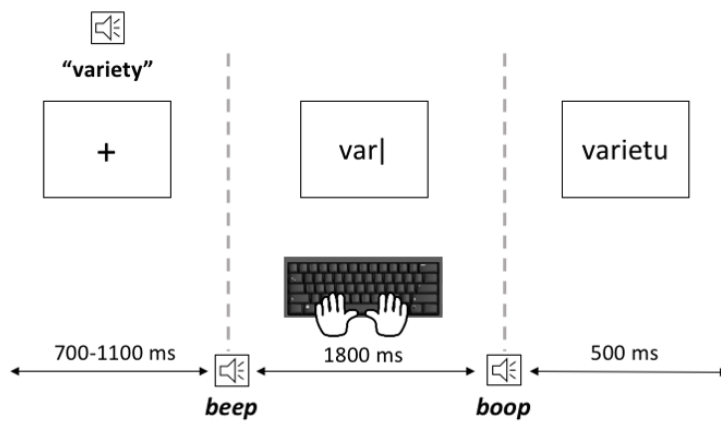
2.3. Procedure

Participants performed a typing-to-dictation task in two conditions (Figure 1). In the immediate-feedback condition, participants saw what they were typing on the screen in real time. In the delayed-feedback condition, a fixation cross was presented on the screen during typing, and what they had typed appeared only at the end of the trial. Conditions were blocked and participants saw them in a counterbalanced order. One of the two balanced lists of stimuli was used for each feedback condition. The list used for each condition was also counterbalanced between participants.

A trial unfolded in a similar way in both conditions. Participants heard a word, followed by a beep (500 Hz, 100 ms). They were instructed to type the word as fast and as accurately as possible when they heard the beep and to finish typing before a second beep (300 Hz, 100 ms) was presented, 1800 ms after the first one. They had 200 ms after the second beep to finish typing their answer, making it a total typing duration of 2000ms from the first beep before the trial was terminated. In the immediate-feedback, what participants had typed stayed on the screen 300 ms after the end of the trial, during which they could not edit their answer anymore. This was done in order to avoid visual potentials elicited by the

change of display from contaminating the signal of interest. Then there was a 1.5 s interval (1000 ms fixation cross, 500 ms blank screen) before the next word was presented. In the delayed-feedback, the screen remained blank during typing. Five hundred milliseconds after the end of the trial, participants were asked to judge the accuracy of their response by pressing a key with their right hand ("L" for a correct response, "J" for an incorrect response). What they had typed during the trial appeared 200 ms later, and stayed on the screen for 1000ms. The inter-trial interval was the same as the immediate-feedback condition. There were three breaks within each block.
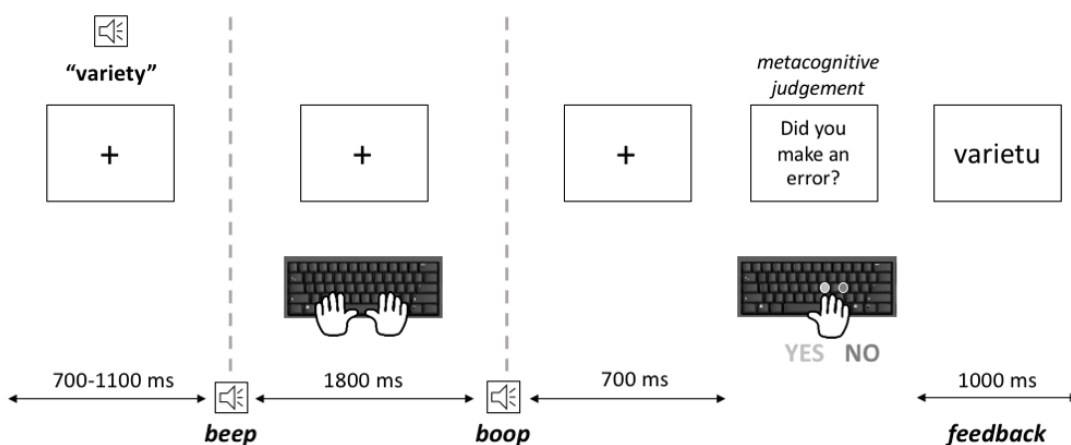

Figure 1. Design of the experimental task. In the immediate-feedback condition, visual feedback was on for the whole duration of the trial. In the delayed-feedback condition, visual feedback was off during typing and only appeared after participants made a metacognitive judgment.

2.4. EEG Recordings and Pre-Processing

EEG data were recorded using a 128-channel HydroCel Geodesic Sensor Net (EGI system), with a vertex reference. Data analysis was performed using Brainstorm (Tadel, Baillet, Mosher, Pantazis, & Leahy, 2011). Participants' performance was closely monitored during the task by the experimenter. As EEG data quality depends on avoiding unnecessary head and eye movements, the experimenter ensured that participants maintained fixation on the screen while typing and did not look back and forth at their hands and the screen. This procedure also ensured that participants did not simply replace one mode of visual feedback (the word on the screen) with another (looking at the keys) in the delayed-feedback condition. Data were filtered offline (0.1-100 Hz) and re-referenced to the averaged mastoids (or one mastoid reference if one of the two mastoid electrodes was too noisy, as was the case for three participants). Electrodes that were noisy for the whole experimental session were rejected and interpolated. Blinks were corrected using Independent Component Analysis (Infomax algorithm, Delorme, Palmer, Onton, Oostenveld, & Makeig, 2012). Residual artefacts were rejected via visual inspection.

Data were segmented in two time-windows of interest time-locked to (1) each keystroke (-500 to +500 ms) and (2) feedback presentation (-300 to +1000 ms, only in delayed-feedback). For the latter window, a baseline was taken from -200 to 0 ms and removed by subtraction. Around keystrokes, we followed another procedure. Data inspection revealed that from the first to the last keystroke of the word, the EEG signal presented a linear positive trend. This means that amplitude was higher (in absolute values) at the end of the word compared to the beginning. In order to collapse data over keystrokes positions, we had to normalize for the position of the keystroke within the word. Because we were interested in the error keystrokes, we subtracted the averaged signal for correct keystrokes from the averaged signal for error keystrokes for each position. This procedure replaced baseline removal for the window time-locked to keystrokes.

2.5. Data Analysis

2.5.1. Behavioral data.

Typed responses were classified as errors if any keystroke was different from the target. Reaction time (RT) was computed as the time between the start beep (i.e., end of auditory stimulus) and the first keystroke. Interkeystroke intervals or IKIs (i.e., the time between two keystrokes) were averaged over each word and taken to index typing speed. Data outside 2.5 SD above or below the mean RT or IKI for each subject were excluded. In total, 2.3% and

2.7% of data were excluded for RTs and IKIs analyses respectively. Accuracy rates, RTs and IKIs were compared using non-parametric statistical tests (Wilcoxon signed rank test) that are robust against the assumptions of parametric tests.

Error awareness was determined based on the metacognitive judgment in the delayed-feedback condition (pressing the "correct" button indicated that the participant had not detected a response as an error, while pressing the "incorrect" button indicated that they had detected their response as an error). *Error correction* was indexed by the use of backspace (which must accompany all corrections in typing regardless of the accuracy of the outcome). To quantify monitoring performance, we used the framework of the signal detection theory (SDT; Clarke, Birdsall, & Tanner, 1959; Galvin, Podd, Drga, & Whitmore, 2003), which has also been successfully applied to explaining both monitoring and selection processes in language production (Nozari et al., 2011; Nozari & Hepner, 2018). In particular, SDT-Type 2 is used for metacognitive decision about one's own performance. Combining the actual accuracy of the response (correct or error trial) with the accuracy of the metacognitive judgments (error detected or not) gives four categories of trials: correct rejection (correct trials judged as correct), hit (error trials judged as error), miss (error trials judged as correct), and false alarm (correct trials judged as error). An efficient monitor is one that has a high hit rate and a low false alarm rate. To capture monitoring performance uncontaminated by response bias (i.e., tendency towards answering "correct" leading to more misses and fewer false alarms or vice versa), we computed d' as z(hit rate) - z(false alarm rate), where z is the z-score of the cumulative normal distribution for the two response types (Macmillan & Creelman, 1991). A lower d' indicates poorer monitoring behavior. Data and analysis scripts are available at: 10.17605/OSF.IO/7EBR4.

2.5.2. EEG data.

Analyses focused on indices of monitoring, namely the ERN, Pe, FRN, and FCP. We started with broad pre-determined windows based on previous studies and refined them using visual observation and Global Field Power (GFP). GFP is an index of the spatial standard deviation of the data and corresponds to the root mean square of the grand average over all electrodes. GFP is used to describe the general time course of the signal, extract peaks of interest and compare conditions (e.g., Python, Fargier, & Laganaro, 2018). Once a GFP peak has been identified, non-parametric tests (Wilcoxon signed rank test, unless otherwise specified) were performed on GFP amplitudes averaged over a 40ms time windows (unless otherwise specified) centered on the peak. Our approach was very similar to an omnibus ANOVA: if an

effect was significant on the GFP amplitudes, then further tests, or "contrasts", were run on all electrodes to uncover the spatial distribution of the effect. To do so, mean ERP amplitudes of all electrodes in the time-window where GFP amplitudes displayed a significant effect were compared with permutation Wilcoxon tests. However, if no peak could be readily identified on GFP, ERP amplitudes were compared point-by-point over the whole time-window. If there was only one condition (e.g., around each keystroke), one-sample t-tests against zero were used. Unless stated otherwise, a threshold of p < .05 was taken for all statistical tests. To provide a measure of the consistency of the effect across subjects, we plotted the proportion of subjects that presented a difference between conditions in the same direction as the grand averages for each electrode and time point, and reported the maximum percentage (Rousselet, Foxe, & Bolam, 2016). Data and analysis scripts are available at: https://figshare.com/projects/Electrophysiological_correlates_of_monitoring_in_typing_with_ and_without_visual_feedback/69770.

3. Results

Participants' ability to type fast and accurately was assessed through an online typing test consisting of copying a short text. Participants that typed faster than 60 words per minute and made less than five mistakes entered the study. After the completion of the study, one participant was excluded because he failed to reach 40% accuracy on the experimental task. On average, participants in the remaining sample typed at a typing speed of 86.4 ± 12 words per minute and made an average of 0.75 ± 0.9 errors in the typing test.

3.1. Behavioral Results

In total, 5632 responses (2816 per condition) were analyzed (see Figure 2). The accuracy rate was not significantly different in the immediate- and delayed-feedback conditions (M(immediate) = 72.5 ± 13%; M(delayed) = 76.5 ± 14%; Z = -1.53, p = .133). RTs were also not significantly different by condition (M(immediate) = 309 ± 35 ms; M(delayed) = 326 ± 41 ms, Z = -1.5, p = .144). Typing speed as measured by mean IKIs was, however, significantly slower in the delayed- than the immediate-feedback condition (M(immediate) = 125 ± 19 ms; M(delayed) = 134 ± 16 ms; Z = 2.64, p = .006).

We then evaluated whether error correction depends on feedback availability (Table 1). While in the immediate-feedback condition, 42% of errors were corrected, only 9% were corrected in the delayed-feedback condition (i.e., 79% change to correction performance; $Z = 3.36, p < .001$). False alarms rates were not significantly different between conditions (delayed: 1.01%, immediate: 1.32%, $Z = 0.98, p = .35$). Average d' was higher in the immediate-feedback (2.06 ± 0.73) compared to delayed-feedback condition (0.89 ± 0.37, $Z = 3.34, p < .001$). Collectively, these results indicate substantially poorer correction performance in the absence of external feedback. This may mean that the error was not detected, or, alternatively, that the error was detected but not corrected. To disentangle the two, we looked at metacognitive judgments, which index error awareness independently of error correction. Results showed that in the delayed-feedback condition, 54% of errors were reported in the metacognitive judgments, significantly higher than the 9% that were corrected, $Z = 3.52, p < .001$ (Table 1). This finding indicates that many more errors in the delayed-feedback condition were detected than corrected.
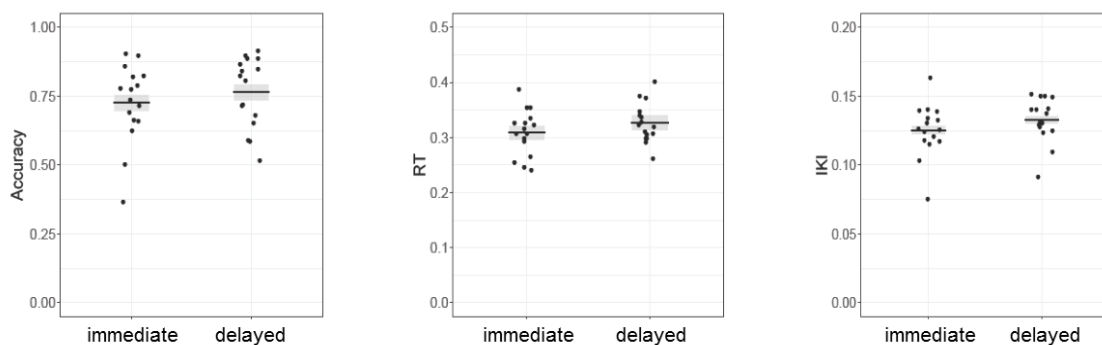

Figure 2. Typing performance, as measured by accuracy rates, reaction times (RT) and typing speed (IKI) compared between immediate- and delayed-feedback. Each dot is a subject, the bold line represents the sample mean, and the shaded area the 95% confidence interval corrected for within-subject design (Morey, 2008).

Table 1. Results of error correction and awareness (metacognitive judgment) for immediate- and delayed-feedback conditions. The legend table specifies what data appear in each cell.

	Not Corrected/Reported	Corrected/Reported
Correct	Correct Rejection	False Alarm
Error	Miss	Hit

	Correction				Awareness	
	Immediate		*Delayed*		*Delayed*	
	No correction	Correction	No correction	Correction	No detection	Detection
Correct	2043 (99%)	11 (1%)	2153 (99.6%)	8 (0.4%)	2065 (96%)	96 (4%)
Error	408 (**54%**)	354 (**46%**)	593 (91%)	62 (**9%**)	300 (**46%**)	355 (**54%**)

3.2. EEG Results

Careful artefact rejection led to the removal on average of 20.7 ± 11% of trials over all time-windows of analysis.

3.2.1. ERN and Pe.

Based on the previous literature, we looked for an ERN from -50 to 100 ms, and for a Pe around 200-400 ms (e.g., Acheson & Hagoort, 2014; Ullsperger, Fischer, Nigbur, & Endrass, 2014) time-locked to each keystroke. Time-locking to each keystroke poses a challenge because keystrokes are separated by only 130 ms on average, thus activity related to one keystroke may influence the activity related to neighboring ones. The only other study that attempted to analyze electrophysiological signal time-locked to individual keystrokes during continuous typing (Kalfaoğlu et al., 2018) did so on ICA components. To avoid the problems with this approach discussed in the earlier sections, we analyzed the full signal without assuming specific components.

Amplitudes were sensitive to the position of the keystroke within the word and increased steadily throughout the word. Therefore, we normalized the ERP signal extracted around error keystrokes. At each keystroke, we subtracted the averaged signal for correct keystrokes from the averaged signal for error keystrokes, and then averaged the normalized signal over all keystroke positions. The resulting normalized signal time-locked to erroneous keystrokes, but independent of keystroke position, is displayed in Figure 3. Since the absence of immediate visual feedback may change processing even on correct trials, we did not directly compare the two feedback conditions.

In the immediate-feedback condition, we were not able to identify a clear negative component resembling an ERN in the predefined time-window (Figure 3AB). There was however a centro-parietal positive component similar to a Pe, starting around 150 ms post-keystroke (mean Cz amplitude = 2.4 ± 3.6 μV (150-250 ms); mean Pz amplitude = 3.2 ± 4.1 μV (300-500 ms); see Figure 3A). Consistent with the previous literature, the effect

developed centrally first (Cz) and then appeared over parietal regions (Pz; see Figure 3A). Point-by-point one-sample t-tests against zero on the ERP amplitudes confirmed an effect from 150 ms to 400 ms, with perfect consistency across subjects (%max = 100%) and widespread over the central electrodes (Fig. 3B).

In the delayed-feedback, we observed a negative fronto-central component similar to an ERN locked to error keystrokes (mean FCz amplitude = -1.2 ± 2.7 μV (0-100 ms); Figure 3CD). Point-by-point one-sample t-tests against zero on the ERP amplitudes confirmed the significance of the fronto-central ERN, in a time-window from -20 to 100 ms, consistent across subjects (%max = 100%), focal in time and centered around the frontal electrodes (Fig. 3D). No clear positive component could be identified following the ERN at the group level, but since the Pe is considered an index of conscious awareness of errors, we further divided errors into reported (i.e., hits; N = 52.7% of trials) and unreported (i.e., misses; N = 47.3% of trials) based on the metacognitive judgments (Figure 3E). Consistent with the theories of the Pe, reported errors (hits) were associated with a centro-parietal component similar to a Pe (mean Cz amplitude = 1.6 ± 6.6 μV (150-250ms)). Unreported errors (misses), on the other hand, showed no clear Pe (mean Cz amplitude = -0.53 ± 3.8 μV (150-250 ms)). However, misses showed a negative fronto-central component similar to an ERN (mean FCz amplitude = -2.0 ± 4.1 μV (0-100 ms)). The contrast between hits and misses was significant in the Pe time window (150-350 ms), over central left electrodes, but not in the ERN time-window.

In summary, we could identify a clear Pe but no clear ERN in the immediate-feedback. In the delayed-feedback condition, on the other hand, an ERN was observed but no clear Pe could be identified when all trials were pooled together. However, breaking down errors by awareness revealed a Pe for hits (errors that participants had consciously detected), but not for misses (those that were not consciously detected).
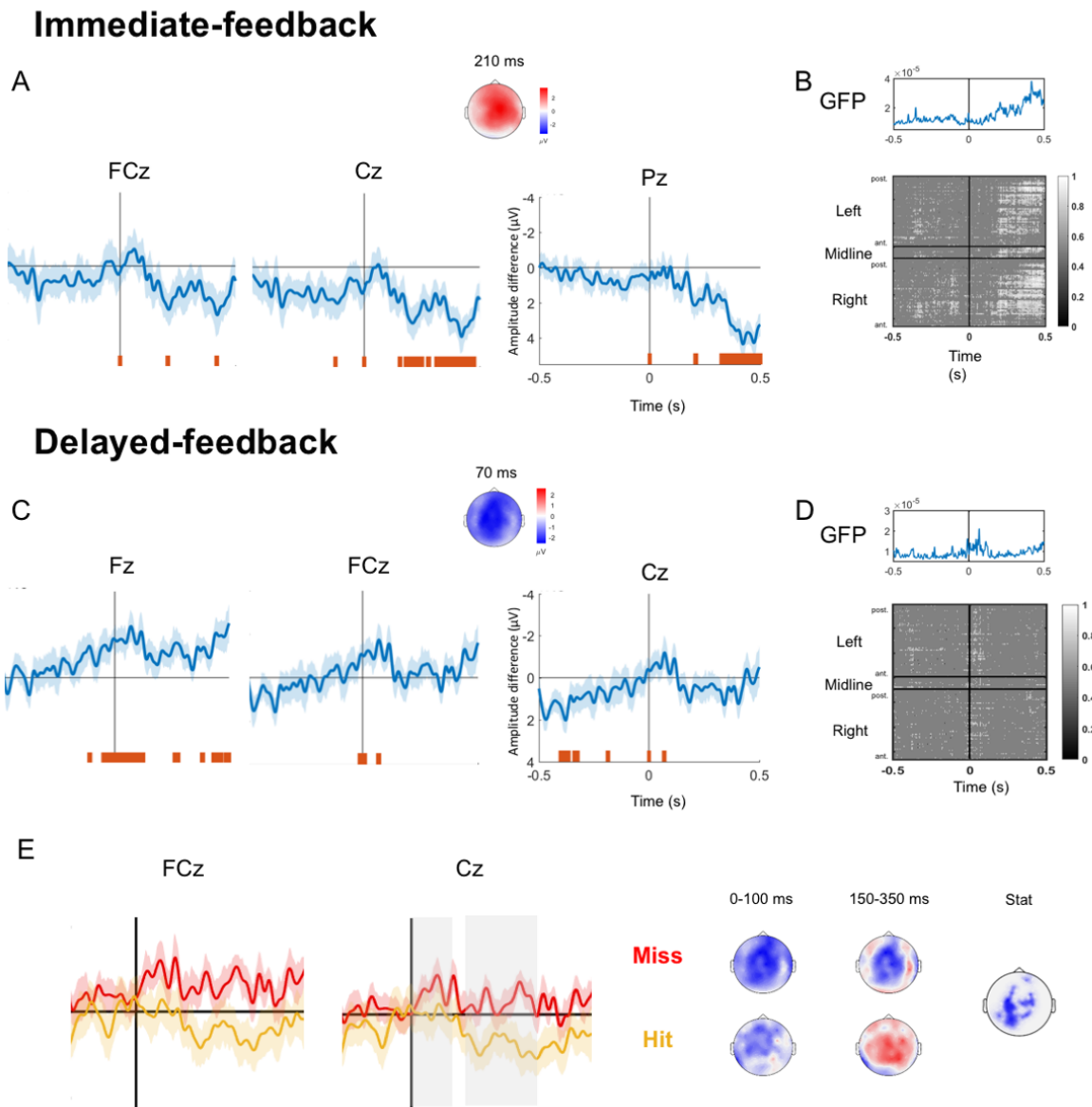
Figure 3. ERPs results for immediate- and delayed-feedback during typing time-locked to error keystrokes. Immediate-feedback: (A) ERP time courses over midline electrodes (FCz, Cz, Pz) for error keystrokes, normalized for position relative to correct keystrokes. Red rugs below the figures indicate significant time points of a one-sample t-test against zero. Shaded area around each line represents the standard error. (B) Global Field Power and individual subject consistency. The proportion of subjects showing an effect in the same direction as the average of subjects is indicated for each time point (x-axis) and electrode (y-axis). Electrodes are ordered from posterior to anterior in left, midline and right hemispheres. White regions correspond to the most consistent effect across subjects. Delayed-feedback: (C) ERP time courses over midline electrodes (Fz, FCz, Cz) for error keystrokes, normalized for position relative to correct keystrokes. (D) Global Field Power and individual subject consistency. (E) ERP time courses for hits (yellow) and misses (red) over midline electrodes (FCz, Cz)

and topographies plotted over 0-100 ms and 150-350 ms time windows (indicated as grey rectangles
on the time courses). The topography of the statistical contrast is plotted at 300 ms.

3.2.2. FRN and FCP.

A clear separation of the external feedback monitoring from the internal monitoring processes
is only possible in the delayed-feedback condition. We focused the analyses on the four types
of responses obtained in the metacognitive judgment task. Based on the previous literature on
visual feedback processing, we chose a predefined time-window of 200-350 ms for the FRN
(Krigolson, 2018) and a predefined time-window of 500-700 ms for the FCP (e.g., Arbel &
Wu, 2016). Figure 4 shows the results. All four trial types presented an early positive
component first, peaking at 220 ms (mean FCz amplitude = 8.7 ± 4.2 µV (correct rejections);
10.0 ± 6.6 µV (hits); 11.2 ± 7.0 µV (misses); 13.6 ± 8.6 µV (false alarms); see Figure AC).
Hits and misses presented a negative component, similar to an FRN peaking at 300 ms (mean
FCz amplitude = 8.3 ± 6.7 µV (hits); 8.5 ± 8.1 µV (misses); see Figure 4A). In contrast, no
such negative component could be identified for false alarms. A later long-lasting positivity
from 400 ms onwards was observed for misses and false alarms with a similar amplitude
(mean FCz amplitude = 13.5 ± 5.8 µV (misses); 14.6 ± 6.0 µV (false alarms)), as well as for
hits albeit with a lower amplitude (mean FCz amplitude = 9.2 ± 7.3 µV).

Based on the time course observed we considered three time-windows of interest for
analysis: the first window examined the positivity that was observed in our data, but had not
been previously reported. The second window reflected the FRN, and the third window the
FCP (Arbel & Wu, 2016; Butterfield & Mangels, 2003; Potts, Martin, Burton, & Montague,
2006). Because of the small number of false alarms that was considerably lower than all other
conditions, statistical comparison of those trials against other trials is problematic. Hits and
misses, however, both had enough number of trials and were comparable to each other. We
thus focused the analyses on these two trial types against correct rejections and against each
other, using mixed linear regressions.

In the early positivity time-window (200-240 ms), statistical analyses on GFP revealed
significant differences in amplitude for all conditions compared to correct rejections (hits: $\beta =$
$15.9, t = 2.1, p = .038$; misses: $\beta = 22.0, t = 2.9, p = .005$; false alarms: $\beta = 37.8, t = 4.9, p$
$< .001$). Importantly, the contrast between hits and misses was not significant, $\beta = 5.9, t =$
$0.79, p = .44$. Permutation tests on ERP amplitudes revealed that significant contrasts with

correct rejections were mostly central, or frontal for hits (Figure 4B) and consistent over subjects (%max = 94% (hits), 94% (misses), 93% (false alarms); Fig. 4D).

The results in the FRN time-window (280-320 ms) were very similar to the earlier time window. Here too the statistical analyses on GFP revealed significant differences in amplitude for all conditions compared to correct rejections (hits: $\beta = 21.5, t = 2.8, p = .008$; misses: $\beta = 22.0, t = 2.8, p = .007$; false alarms: $\beta = 55.1, t = 6.8, p < .001$). Again, the contrast between hits and misses was not significant ($\beta = 0.50, t = 0.07, p = .95$). Permutation tests on ERP amplitudes revealed that significant contrasts with correct rejections were present over central electrodes except for misses (Figure 4B) and consistent across subjects (%max = 93%; Fig. 4D). The between-subject correlation of ERN amplitude measured on FCz (0-100ms) and FRN amplitude averaged over hits and misses was positive and highly significant ($R = .72, p = .002$).

In the FCP time-window (500-700 ms), all trial types were again significantly different from correct rejections on GFP amplitudes (hits: $\beta = 42.5, t = 4.0, p < .001$; misses: $\beta = 63.1, t = 5.9, p < .001$; false alarms: $\beta = 74.6, t = 6.7, p < .001$). In this window, however, misses largely overlapped with false alarms and were marginally different from hits ($\beta = -20.6, t = -1.9, p = .06$). Permutation tests on ERP amplitudes showed that the spatial distribution of the contrast with correct rejections were again central for the three conditions (Figure 4B), and perfectly consistent across subjects (%max = 100%) from 350 ms up to 650 ms for hits, 350 ms to 800 ms for misses and from 200 ms to 800 ms for false alarms (Figure 4D).

In summary, during feedback processing, correct rejections patterned differently from all other trial types in both early and late phases of processing. Interestingly, there was a switch between the pattern of hits and misses from early to late phases of feedback processing: in earlier phases, hits and misses were largely overlapping suggesting similar processing of all objectively incorrect responses. In later stages, however, hits were processed differently from misses, the latter of which patterned closely with false alarms, suggesting a closer grouping of trials based on monitoring performance, as opposed to response accuracy.
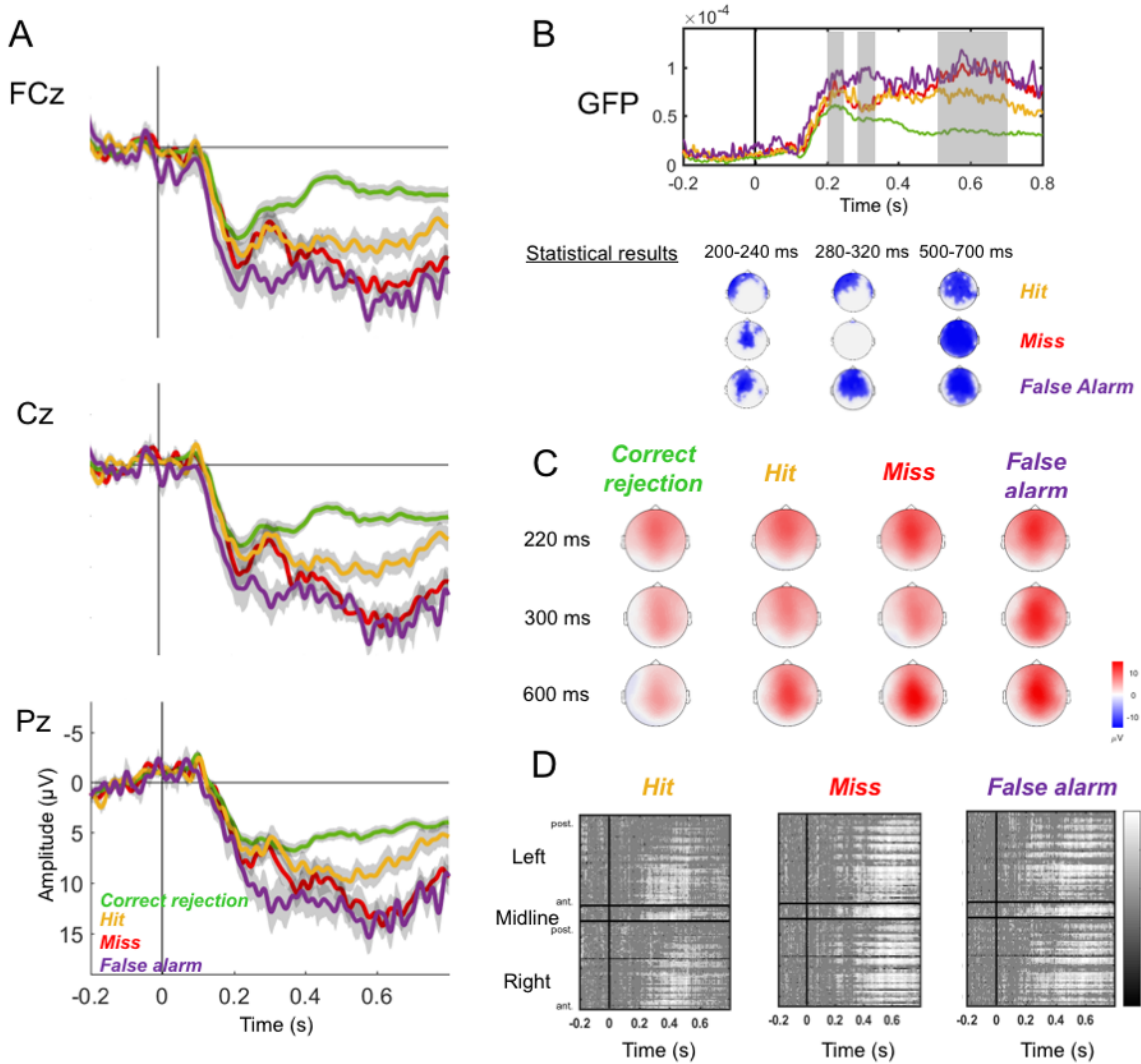
Figure 4. ERPs results time-locked to feedback presentation in the delayed-feedback condition. (A) ERPs time courses over midline electrodes (FCz, Cz, Pz) for correct rejections (green), hits (yellow), misses (red), and false alarms (purple). Shaded area around each line represents standard error. (B) Global Field Power. Grey rectangles correspond to time-windows selected for the statistical analyses. Topographies of significant electrodes for each contrast and time-window are plotted below. (C) Topographies of ERPs are plotted at 220 ms, 300 ms, and 600 ms for each trial type. (D) Individual subject consistency. Proportion of subjects showing an effect in the same direction as the average of subjects is indicated for each time point (x-axis) and electrode (y-axis). Electrodes are ordered from posterior to anterior in left, midline and right hemispheres. White regions correspond to the most consistent effect across subjects.

4. Discussion

4.1 Behavioral Findings

Comparison between typing with and without visual feedback showed that correction rates were significantly lower when external feedback was removed. Metacognitive judgments revealed that participants had actually detected many of the errors they had failed to correct in the delayed-feedback condition. Together, these findings imply a critical role for external feedback in error corrections. Similar to findings in spoken production (e.g., Postma & Kolk, 1992), and unlike the copy-typing task used by Snyder et al. (2015), the removal of visual feedback in our typing-to-dictation task did not increase error rates (although it did slow down typing). This difference shows the reliance of participants on visual stimuli during copy typing, an effect that is preferably to be avoided if true lexical retrieval and segmental encoding are the targeted cognitive processes. In keeping with speech studies, we found a decrease in the rate of corrections when the external channel was blocked, however the magnitude of this decrease was much larger in typing (~80%) compared to speaking (~50% in Postma & Kolk, 1992; ~17% in Oomen, Postma, & Kolk, 2001; and no significant decrease in Nooteboom & Quené, 2017). It is worth noting that a previous study reported higher correction rates than we do in a copy-typing task without visual feedback (Kalfaoğlu & Stafford, 2014), but the task did not include a temporal deadline and instructions specifically emphasized that participants must correct all their errors. Despite this, a correction rate of only 60% was achieved, which could reflect the upper bound of corrections that can be achieved in the absence of the external channel.

The different rates of change in correction rates between spoken and typed production with and without external feedback may reflect the differential contribution of internal vs. external channel to monitoring in these two modalities. However, when artificial errors were inserted in spoken and typed production (Lind et al., 2014; Logan & Crump, 2010), they elicited similar degrees of acceptance, pointing to comparable reliance of participants on the external channel in the two production modalities. Given this finding, the larger drop in correction rates in typing compared to speaking likely reflects the fact that noise masking only partially blocks the external channel during speaking, and removal of that channel may be associated with greater decrease in correction rates in spoken modality as well. Another possibility is that there is a greater need for the kind of information provided by the external channel for correcting errors in typing compared to speaking. While correcting spoken words almost always entails restarting the word, correcting typed errors rarely entails erasing and

retyping the whole word (see Appendix). Thus, information about the identity of the previous letter or the location of the error may be crucial for applying the correct correction. The role of these two factors in efficient corrections should be further explored in future studies.

Interestingly, the much lower correction rates in the absence of visual feedback were *not* accompanied by poor awareness over errors. Participants consciously detected about half of their errors, a rate comparable to previous studies in speech monitoring (Hartsuiker et al., 2005; Nooteboom, 2005; Postma & Noordanus, 1996), and comparable to the rate estimated by the conflict-based model of Nozari et al., (2011). This finding shows that a) error awareness is not sufficient for correction, and b) the role of the external channel is much less prominent in error awareness than in error correction.

To summarize, the behavioral findings of this study showed similar patterns in monitoring of spoken and typed errors, extending previous results that demonstrated the similarities between the error patterns in the two modalities (Pinet & Nozari, 2018). Importantly, the external channel seems to play an important role in repairing errors, but this role is not tied to conscious detection of errors.

Connection to models of typing. The first implementation of a typing model was carried out by Rumelhart and Norman (1982), which consisted of word schema, keystroke schemata and a response system. Once the word schema is activated, the model simulates the process of activating its corresponding keystroke schemata, and ultimately the target positions on the keyboard. This model was the first model to successfully capture the basic properties of typing such as speed and accuracy in skilled typists. A more recent and nuanced model of typing, the hierarchical-processing model, was developed by Logan and Crump (2011) to capture an interesting dissociation observed in Logan and Crump (2010). As described in the Introduction, that study artificially altered a proportion of typed words. The first finding was that typists were willing to accept some of the inserted errors as their original responses, pointing to an influence of the external monitoring channel on monitoring behavior. The second finding was a typical slowing down after self-made errors which was absent for inserted errors that were presumably processed through the external (visual) monitoring channel. This finding suggested reliance on the internal monitoring channel.

Logan and Crump (2011) formalized this dissociation by proposing a model with two informationally-encapsulated loops: the *outer loop* is responsible for visually processing the input in copy-typing, encoding the word to type, and processing the visual feedback. The *inner loop* is responsible for the inner dynamics of typing including mapping the word onto

segments, controlling the sequential activation of keystrokes, navigating the position of fingers on the keyboard, and processing the proprioceptive feedback (see also Yamaguchi, Crump, & Logan, 2013). There are clear parallels between the internal and external monitoring channels that frame the backbone of the current study and Logan and Crump's (2011) inner and outer loops. They are, however, not identical.

One difference is the encapsulation of information between the two loops, which clearly separates lexical and sublexical processes. Pinet and Nozari (2018) provided evidence against such separation by showing that typing errors were affected by the feedback between sublexical segments and lexical representations, similar to interactive models of spoken production (e.g., Dell, 1986). This demonstration is relevant to the current discussion in that it changes the notion of "internal channel" from being confined to the inner loop to encompassing processes that include word retrieval and the interaction between segmental and lexical activation. This, in turn, suggests that internal monitoring may be carried out by multiple mechanisms, examples of which are conflict detection (Nozari et al., 2011), or internal canceling of the activation of perceptual information by forward production commands (Hickok, 2012).

Another potential difference is that the functioning of the external monitoring channel is, to a large extent, outside of the primary processes that contribute to production. For example, regardless of the modality (spoken, hand-written, or typed), language production – broadly defined– entails activating semantic information and competing lexical representations, selecting the target representation among competitors, and mapping that target onto sublexical segments. All of this, according to psycholinguistic models, is carried out within the production system and independently of the external monitoring channel. The outer loop, on the other hand, as suggested by Logan and Crump (2011), plays a critical role in planning parts of the primary process of typing. Now, this could be because the model's primary scope is copy-typing, in which the visual input plays a critical role in initiating the typing process, but our understanding is that the outer loop is also involved in carrying out the processes required to retrieving the target word, in which case, its role is clearly much more extensive that the external monitoring channel.

To summarize, the distinction between the internal/external monitoring channels finds an important parallel in the inner/outer loops in the hierarchical-processing model. This parallel, however, is approximate, and the two labeling systems have differences in how they draw the boundaries that separate stages of processing.

4.2. Error-related indices of monitoring: ERN and Pe

The time-courses of the negative and positive keystroke-locked components that we observed and their fronto-central and centro-parietal topographies were consistent with the ERN and Pe components, respectively, described in the performance monitoring literature (Ullsperger et al., 2014). While a clear ERN was uncovered in the delayed-feedback condition (see also Kalfaoğlu et al., 2018), we could not identify such a component in the immediate-feedback condition. On the other hand, a clear Pe was observed in the immediate-feedback condition when all errors were lumped together and compared against correct trials, but was only evident on a subset of trials in the delayed-feedback condition in which participants were aware of the error.

The presence of an ERN in the delayed-feedback condition was expected, since, despite the debates around its origin (e.g., Hewig, Coles, Trippe, Hecht, & Miltner, 2011; Ridderinkhof, van den Wildenberg, Segalowitz, & Carter, 2004; Yeung, Botvinick, & Cohen, 2004), the ERN has often been implicated as an index of internal monitoring, which this condition taps into. The absence of a clear ERN in the immediate-feedback condition was not predicted a priori, although some modulation of ERN was expected, since both the internal and the external channels are available in the immediate-feedback condition. In past studies, the absence of an ERN has been reported when the stimulus could not be analyzed due to subliminal presentation (Charles et al., 2013), or when no correct response was available (Di Gregorio, Maier, & Steinhauser, 2018). Neither of these apply to the current experiment. Likewise, there are two factors that have been reported to increase the amplitude of the ERN: conscious detection of errors (Hughes & Yeung, 2011; Maier, Steinhauser, & Hübner, 2008), and higher likelihood of corrections (Gehring et al., 1993). Note, however, that error awareness is, if anything, higher in the immediate-feedback condition where participants could immediately see their errors. Similarly, the data show much higher correction rates in the immediate condition. Therefore, neither of these factors could explain the absence of an ERN in the immediate-feedback condition. Finally, a null effect may always arise from lack of statistical power, but a clear ERN was recovered from the delayed-feedback condition with a comparable number of errors, decreasing the probability of a null effect due to statistical problems. We thus cautiously interpret the differential pattern of ERN recovery between the immediate- and delayed-feedback conditions as suggesting relatively less reliance on monitoring through the internal channel when external feedback is available online during performance. Moreover, our results are much better aligned with the ERN being an index of an internal monitoring mechanism, as opposed to a mechanism that brings errors to conscious

awareness or a mechanism that is primarily concerned with the availability of a correction as have been suggested in some studies (Gehring et al., 1993; Hughes & Yeung, 2011; Maier et al., 2008). It remains possible, however, that both of these latter mechanisms could be indirectly related to the dynamics of internal monitoring.

One study before this has explicitly investigated the Pe component in language production. In a copy-typing task, Kalfaoğlu et al. (2018) reported a Pe immediately following ERN on the same ICA component and with the same topography as the ERN. The Pe reported in that study most likely represents the early Pe, not the late Pe typically described in the action monitoring literature. The Pe component uncovered in the current study, however, is precisely what has been described in the action monitoring literature, in terms of timeline, topography, and sensitivity to awareness (see also Ganushchak & Schiller, 2008b; Masaki et al., 2001 for hints of the same components, although not explicitly analyzed). In piano playing, for example, Maidhof (2013) found a Pe (200-350 ms) with a centro-parietal topography, which closely matches the timeline and topography of the Pe in the current study. As described in the Introduction, the Pe has been described as an index of conscious awareness over errors in the non-linguistic tasks (Boldt & Yeung, 2015; O'Connell et al., 2007; Overbeek, Nieuwenhuis, & Ridderinkhof, 2005; Steinhauser & Yeung, 2010). Based on this, we expected the Pe to be most prominent when participants were most likely to be aware of their errors. In keeping with this prediction, a robust Pe was found in the immediate-feedback condition, where the presence of visual feedback during typing provides a strong cue for becoming aware of an error. In the delayed-feedback condition, such a cue was absent. Correspondingly, we only observed a Pe on a subset of trials in which participants reported having been aware of errors. Compatible with our results, Ruiz, Jabusch, & Altenmüller (2009) manipulated the presence of auditory feedback during piano playing and observed that the Pe had a larger amplitude with auditory feedback than without.

Collectively, our findings regarding the ERN and the Pe confirm and extend previous claims that monitoring in language production is carried out by the same domain-general processes as other systems, albeit operating on domain-specific representations (Ganushchak & Schiller, 2008b; Hanley, Cortis, Budd, & Nozari, 2016; Nozari et al., 2011; Nozari & Novick, 2017; Riès et al., 2011). To be specific, the claim is that monitoring mechanisms are based on the operations carried out on domain-specific representations (e.g., linguistic, visual, etc.), and are thus sensitive to domain-specific properties (see Freund & Nozari (2018) for an extension to the implementation of control). Even within a specific domain such as language, monitoring may be sensitive to differences in the nature of representations, an example of

which is the double dissociation between detecting lexical and sublexical errors in individuals with aphasia (Nozari et al., 2011), or the specificity of lexical indices of monitoring for the detection of lexical (but not sublexical) errors in children (Hanley et al., 2016). In the same vein, it is likely that monitoring in typing is sensitive to the specific representations involved in typing, such as those linking segments to target positions on the keyboard (keypress schemata; Rumelhart & Norman, 1982). However, a purely domain-specific monitoring system would not be expected to have the exact same electrophysiological signature across different domains. The report of such a common signature as the ERN in verbal and nonverbal tasks in the past studies (e.g., Riès et al., 2011), complemented by the more comprehensive report of several such signatures (ERN, Pe, FRN, and FCP) in the current paper, argues for a central domain-general component to monitoring mechanisms across various domains, thus ruling out purely domain-specific monitoring accounts.

The dissociations observed in the patterns of ERN and Pe have further implications for monitoring via the internal and external channels: when the external channel is removed, monitoring relies heavily on the internal channel. This generates a clear ERN, but overall lower confidence in having made an error, since no external cues are available to bring errors to conscious attention. When, on the other hand, the external channel is available, it is relied upon by participants for monitoring. This, in turn, decreases the reliance on the internal channel (no clear ERN), but increases error awareness through external cues (clear Pe).

4.3. Feedback-related indices of monitoring: FRN and FCP

The delayed-feedback condition provides the ideal context for examining feedback-related ERPs, as visual feedback is temporally separated from the operation of internal monitoring mechanisms. Moreover, this separation allows us to understand how the system integrates the output of internal monitoring channel with the external channel. We were able to uncover both a reliable FRN and a reliable FCP in this experiment. This is, to our knowledge, the first report of such components in the context of language production. The most striking finding, however, was the discovery of two phases of feedback processing and integration. In the early phase (first and second windows of analysis), i.e., the FRN time window, we did not find a difference in the amplitude of the FRN between hits and misses, suggesting that both detected and undetected errors were treated similarly at this stage, and differently from the majority of correct responses. It is worth mentioning that we found the same pattern that differentiates between error and correct trials in an earlier time-window, which has sometimes been observed, but has not been explicitly reported (Butterfield & Mangels, 2003; Eppinger et al.,

2008; Glazer, Kelley, Pornpattananangkul, Mittal, & Nusslock, 2018). The patterning of hits and misses together may suggest that the first phase of feedback processing concerns the objective status of response (correct/incorrect), with incorrect responses eliciting a different brain response than correct responses.

The above interpretation, however, would predict that false alarms (which are objectively correct responses) would pattern with correct rejections. Contradictory to this prediction, ERPs associated with false alarms showed the greatest positivity of all response types. The problem is that false alarms are rare events in this task and the low number of such trials does not allow for robust comparisons and interpretations. A conservative interpretation of the results that would accommodate this pattern of false alarms would be that the first phase of feedback processing simply reflects the system's sensitivity to response probabilities. Correct rejections are the most common response type, followed by hits and misses that are about equal, with false alarms being the rarest type of response. Experiments with designs that elicit greater rates of false alarms should distinguish between the two interpretations discussed above. The critical point, however, is that in this early phase, the system does not distinguish between hits and misses, and is thus not reflecting the quality of monitoring performance.

In contrast to the early phase, the late phase (the third window of analysis), which falls within the window of the FCP, clearly distinguished between hits and misses, and categorized the latter with false alarms. This pattern shows that the output of internal and external channels has been fully integrated at this point, in order for the system to best adjust itself for performance optimization on future trials: the least positive are correct rejections, which require no adjustments. The next most positive are hits, which require adjustments to the primary production processes, but not to monitoring processes (hits indicate good monitoring). The most positive are misses and false alarms that are failures of the monitoring system, and require a different type of adjustment than hits, such as adjustments to the criterion for detecting an error (Nozari & Hepner, 2018). This pattern of results is compatible with FCP indicating the need for adjustments (i.e., learning) in the system and is compatible with studies in the non-linguistic literature that have linked a higher FCP amplitude to better learning outcomes (Arbel et al., 2013; Arbel & Wu, 2016; Butterfield & Mangels, 2003; Glazer et al., 2018; Mangels, Butterfield, Lamb, Good, & Dweck, 2006).

Finally, we found a strong positive correlation between the amplitude of the ERN and FRN in our study. There is an ongoing debate on whether the ERN and FRN arise from the same or different systems (Gentsch, Ullsperger, & Ullsperger, 2009; Müller, Möller, Rodríguez-Fornells, & Münte, 2005; Nieuwenhuis, Yeung, Holroyd, Schurger, & Cohen,

2004). Potts, Martin, Kamp, & Donchin (2011) showed that the spatial decomposition of both components yielded highly correlated factors, suggesting that they may share a common source in the ACC, although the FRN had a more anterior distribution than the ERN, pointing to the possible existence of an additional frontal source. The finding of a positive correlation between the two components in our study is compatible with a common element between the two. Importantly, it shows that in tasks that are well-learned, e.g., typing of known words, a stronger index of internal monitoring goes hand in hand with a stronger index of feedback monitoring, suggesting that people who monitor their production more closely use both resources to do so. This is in contrast with the findings in learning a new task, in which the magnitude of ERN gradually increases as the magnitude of FRN decreases, suggesting a shift towards internal monitoring (e.g., Eppinger et al., 2008; Nieuwenhuis et al., 2002).

To summarize, the results of our FRN and FCP analyses, similar to the ERN and Pe analyses, point to domain-general processes underlying monitoring language production. They further show that participants use the information from external feedback to check the output of the internal monitoring channel and integrate the two, possibly in order to guide adjustments in the primary production and monitoring processes to make the system more efficient for future use.

4.4. Conclusion

Our study is the first to report both error- (ERN, Pe) and feedback-related (FRN, FCP) domain-general indices of monitoring in a task involving language production. The removal of visual feedback prompted a stronger reliance on internal monitoring processes, which resulted in lower correction rates, a clear ERN, and lower confidence in error detection. Indices of conscious error detection (Pe) were only observed when confidence in having made an error was high, or external feedback was available. Feedback-related indices further showed that external and internal channel outputs are integrated, most likely in order to drive adjustments to the production and monitoring systems, and optimize future performance.

References

Acheson, D. J., Ganushchak, L. Y., Christoffels, I. K., & Hagoort, P. (2012). Conflict monitoring in
    speech production: physiological evidence from bilingual picture naming. *Brain and Language*,
    123(2), 131–6. http://doi.org/10.1016/j.bandl.2012.08.008

Acheson, D. J., & Hagoort, P. (2014). Twisting tongues to test for conflict-monitoring in speech
    production. *Frontiers in Human Neuroscience*, *8*(206), 1–16.
    http://doi.org/10.3389/fnhum.2014.00206

Arbel, Y., Goforth, K., & Donchin, E. (2013). The Good, the Bad, or the Useful? The Examination of
    the Relationship between the Feedback-related Negativity (FRN) and Long-term Learning
    Outcomes. *Journal of Cognitive Neuroscience*, *25*(8), 1249–1260.
    http://doi.org/10.1162/jocn_a_00385

Arbel, Y., & Wu, H. (2016). A Neurophysiological examination of quality of learning in a feedback-
    based learning task. *Neuropsychologia*, *93*(March), 13–20.
    http://doi.org/10.1016/j.neuropsychologia.2016.10.001

Balass, M., Halderman, L. K., Benau, E. M., & Perfetti, C. A. (2016). Semantic processes and
    individual differences detected through error-related negativities. *Journal of Neurolinguistics*,
    37, 82–97. http://doi.org/10.1016/j.jneuroling.2015.08.002

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K., Kessler, B., Loftis, B., … Treiman, R. (2007).
    The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459.
    http://doi.org/10.3758/BF03193014

Bohland, J. W., Bullock, D., & Guenther, F. H. (2010). Neural representations and mechanisms for the
    performance of simple speech sequences. *Journal of Cognitive Neuroscience*, *22*(7), 1504–1529.
    http://doi.org/10.1162/jocn.2009.21306

Boldt, A., & Yeung, N. (2015). Shared Neural Markers of Decision Confidence and Error Detection.
    Journal of Neuroscience, *35*(8), 3478–3484. http://doi.org/10.1523/JNEUROSCI.0797-14.2015

Bonin, P., Méot, A., Lagarrigue, A., & Roux, S. (2015). Written object naming, spelling to dictation,
    and immediate copying: Different tasks, different pathways? *Quarterly Journal of Experimental
    Psychology (2006)*, *68*(7), 1268–1294. http://doi.org/10.1080/17470218.2014.978877

Bonini, F., Burle, B., Liégeois-Chauvel, C., Regis, J., Chauvel, P., Vidal, F., … Vidal, F. (2014).
    Action monitoring and medial frontal cortex: leading role of supplementary motor area. *Science*,
    343(6173), 888–91. http://doi.org/10.1126/science.1247412

Bourguignon, N. J. (2014). A rostro-caudal axis for language in the frontal lobe: The role of executive
    control in speech production. *Neuroscience and Biobehavioral Reviews*, *47*, 431–444.
    http://doi.org/10.1016/j.neubiorev.2014.09.008

Breining, B., Nozari, N., & Rapp, B. (2016). Does segmental overlap help or hurt? Evidence from
    blocked cyclic naming in spoken and written production. *Psychonomic Bulletin & Review*.

http://doi.org/10.3758/s13423-015-0900-x

Butterfield, B., & Mangels, J. A. (2003). Neural correlates of error detection and correction in a semantic retrieval task. *Cognitive Brain Research*, *17*(3), 793–817. http://doi.org/10.1016/S0926-6410(03)00203-9

Charles, L., Van Opstal, F., Marti, S., & Dehaene, S. (2013). Distinct brain mechanisms for conscious versus subliminal error detection. *NeuroImage*, *73*, 80–94. http://doi.org/10.1016/j.neuroimage.2013.01.054

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*(9), 868–879. http://doi.org/10.1002/hbm.20315

Clarke, F. R., Birdsall, T. G., & Tanner, W. P. (1959). Two Types of ROC Curves and Definitions of Parameters. *The Journal of the Acoustical Society of America*, *31*(5), 629–630. http://doi.org/10.1121/1.1907764

Damian, M. F., & Freeman, N. H. (2008). Flexible and inflexible response components: a Stroop study with typewritten output. *Acta Psychologica*, *128*(1), 91–101. http://doi.org/10.1016/j.actpsy.2007.10.002

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., von Cramon, D. Y., & Engel, A. K. (2005). Trial-by-Trial Coupling of Concurrent Electroencephalogram and Functional Magnetic Resonance Imaging Identifies the Dynamics of Performance Monitoring. *Journal of Neuroscience*, *25*(50), 11730–11737. http://doi.org/10.1523/JNEUROSCI.3286-05.2005

Delorme, A., Palmer, J., Onton, J., Oostenveld, R., & Makeig, S. (2012). Independent EEG sources are dipolar. *PloS One*, *7*(2), e30135. http://doi.org/10.1371/journal.pone.0030135

Di Gregorio, F., Maier, M. E., & Steinhauser, M. (2018). Errors can elicit an error positivity in the absence of an error negativity: Evidence for independent systems of human error monitoring. *NeuroImage*, *172*(February), 427–436. http://doi.org/10.1016/j.neuroimage.2018.01.081

Eppinger, B., Kray, J., Mock, B., & Mecklinger, A. (2008). Better or worse than expected? Aging, learning, and the ERN. *Neuropsychologia*, *46*(2), 521–539. http://doi.org/10.1016/j.neuropsychologia.2007.09.001

Falkenstein, M., Hohnsbein, J., Hoormann, 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*(6), 447–455. http://doi.org/10.1016/0013-4694(91)90062-9

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(41), 16616–21. http://doi.org/10.1073/pnas.1315235110

Fiez, J. A. (2016). The cerebellum and language: Persistent themes and findings. *Brain and Language*, *161*, 1–3. http://doi.org/10.1016/j.bandl.2016.09.004

Freund, M., & Nozari, N. (2018). Is adaptive control in language production mediated by learning? *Cognition*, *176*(June 2017), 107–130. http://doi.org/10.1016/j.cognition.2018.03.009

Galvin, S. J., Podd, J. V., Drga, V., & Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: Discrimination between correct and incorrect decisions. *Psychonomic Bulletin & Review*, *10*(4), 843–876. http://doi.org/10.3758/BF03196546

Ganushchak, L. Y., & Schiller, N. O. (2008a). Brain error-monitoring activity is affected by semantic relatedness: an event-related brain potentials study. *Journal of Cognitive Neuroscience*, *20*(5), 927–940. http://doi.org/10.1162/jocn.2008.20514

Ganushchak, L. Y., & Schiller, N. O. (2008b). Motivation and semantic context affect brain error-monitoring activity: An event-related brain potentials study. *NeuroImage*, *39*(1), 395–405. http://doi.org/10.1016/j.neuroimage.2007.09.001

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*(6), 385–390. http://doi.org/10.1111/j.1467-9280.1993.tb00586.x

Gehring, W. J., & Willoughby, A. R. (2004). Are all medial frontal negativities created equal? Toward a richer empirical basis for theories of action monitoring. *Errors, Conflicts, and the Brain. Current Opinions on Performance Monitoring*, 14–20.

Gentsch, A., Ullsperger, P., & Ullsperger, M. (2009). Dissociable medial frontal negativities from a common monitoring system for self- and externally caused failure of goal achievement. *NeuroImage*, *47*(4), 2023–2030. http://doi.org/10.1016/j.neuroimage.2009.05.064

Glazer, J. E., Kelley, N. J., Pornpattananangkul, N., Mittal, V. A., & Nusslock, R. (2018). Beyond the FRN: Broadening the time-course of EEG and ERP components implicated in reward processing. *International Journal of Psychophysiology*, *132*(August 2017), 184–202. http://doi.org/10.1016/j.ijpsycho.2018.02.002

Guenther, F. H. (2016). *Neural Control of Speech*. MIT Press.

Hanley, J. R., Cortis, C., Budd, M. J., & Nozari, N. (2016). Did I say dog or cat? A study of semantic error detection and correction in children. *Journal of Experimental Child Psychology*, *142*, 36–47. http://doi.org/10.1016/j.jecp.2015.09.008

Hartsuiker, R., & Kolk, H. H. J. (2001). Error monitoring in speech production: a computational test of the perceptual loop theory. *Cognitive Psychology*, *42*, 113–157. http://doi.org/10.1006/cogp.2000.0744

Hartsuiker, R., Kolk, H. H. J., & Martensen, H. (2005). The division of labor between internal and external speech monitoring. In *Phonological encoding and monitoring in normal and pathological speech* (p. 187). Psychology Press.

Herrmann, M. J., Römmler, J., Ehlis, A. C., Heidrich, A., & Fallgatter, A. J. (2004). Source localization (LORETA) of the error-related-negativity (ERN/Ne) and positivity (Pe). *Cognitive Brain Research*, *20*(2), 294–299. http://doi.org/10.1016/j.cogbrainres.2004.02.013

Hewig, J., Coles, M. G. H., Trippe, R. H., Hecht, H., & Miltner, W. H. R. (2011). Dissociation of Pe and ERN/Ne in the conscious recognition of an error. *Psychophysiology*, *48*(10), 1390–1396. http://doi.org/10.1111/j.1469-8986.2011.01209.x

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews. Neuroscience*, *13*(2), 135–45. http://doi.org/10.1038/nrn3158

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*(4), 679–709. http://doi.org/10.1037//0033-295X.109.4.679

Hughes, G., & Yeung, N. (2011). Dissociable correlates of response conflict and error awareness in error-related brain activity. *Neuropsychologia*, *49*(3), 405–415. http://doi.org/10.1016/j.neuropsychologia.2010.11.036

Ito, J., & Kitagawa, J. (2006). Performance monitoring and error processing during a lexical decision task in patients with Parkinson's disease. *Journal of Geriatric Psychiatry and Neurology*, *19*(1), 46–54. http://doi.org/10.1177/0891988705284716

Kalfaoğlu, Ç., & Stafford, T. (2014). Performance breakdown effects dissociate from error detection effects in typing. *Quarterly Journal of Experimental Psychology (2006)*, *67*(March 2015), 508–24. http://doi.org/10.1080/17470218.2013.820762

Kalfaoğlu, Ç., Stafford, T., & Milne, E. (2018). Frontal theta band oscillations predict error correction and posterror slowing in typing. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(1), 69–88. http://doi.org/10.1037/xhp0000417

Krigolson, O. E. (2018). Event-related brain potentials and the study of reward processing: Methodological considerations. *International Journal of Psychophysiology*, *132*(October 2018), 175–183. http://doi.org/10.1016/j.ijpsycho.2017.11.007

Lackner, J. R., & Tuller, B. H. (1979). Role of efference monitoring in the detection of self-produced speech errors. *Sentence Processing*, 281–294.

Lane, H., & Tranel, B. (1971). The Lombard Sign and the Role of Hearing in Speech. *Journal of Speech, Language, and Hearing Research*, *14*(4), 677–709.

Laver, J. D. M. (1980). Monitoring systems in the neurolinguistic control of speech production. *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand*, 287–305.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41–104. http://doi.org/10.1016/0010-0277(83)90026-4

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

Lind, A., Hall, L., Breidegard, B., Balkenius, C., & Johansson, P. (2014). Speakers' acceptance of real-time speech exchange indicates that we use auditory feedback to specify the meaning of what we say. *Psychological Science*, *25*(6), 1198–1205. http://doi.org/10.1177/0956797614529797

Logan, G. D., & Crump, M. J. C. (2010). Cognitive illusions of authorship reveal hierarchical error

detection in skilled typists. *Science*, *330*(6004), 683–6. http://doi.org/10.1126/science.1190483

Logan, G. D., & Crump, M. J. C. (2011). Hierarchical control of cognitive processes: the case for skilled typewriting. *The Psychology of Learning and Motivation*, *54*, 1–27. http://doi.org/10.1016/B978-0-12-385527-5.00001-2

Luft, C. D. B. (2014). Learning from feedback: The neural mechanisms of feedback processing facilitating better performance. *Behavioural Brain Research*, *261*, 356–368. http://doi.org/10.1016/j.bbr.2013.12.043

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. *New York*. Cambridge: Cambridge University Press.

Maidhof, C. (2013). Predictive error detection in pianists: a combined ERP and motion capture study. *Frontiers in Human Neuroscience*, *7*(September), 1–14. http://doi.org/10.3389/fnhum.2013.00587

Maier, M. E., Steinhauser, M., & Hübner, R. (2008). Is the error-related negativity amplitude related to error detectability? Evidence from effects of different error types. *Journal of Cognitive Neuroscience*, *20*(12), 2263–2273. http://doi.org/10.1162/jocn.2008.20159

Mangels, J. A., Butterfield, B., Lamb, J., Good, C., & Dweck, C. S. (2006). Why do beliefs about intelligence influence learning success? A social cognitive neuroscience model. *Social Cognitive and Affective Neuroscience*, *1*(2), 75–86. http://doi.org/10.1093/scan/nsl013

Marvel, C. L., & Desmond, J. E. (2010). The contributions of cerebro-cerebellar circuitry to executive verbal working memory. *Cortex*, *46*(7), 880–895. http://doi.org/10.1016/j.cortex.2009.08.017

Masaki, H., Tanaka, H., Takasawa, N., & Yamazaki, K. (2001). Error-related brain potentials elicited by vocal errors. *Neuroreport*, *12*(9), 1851–1855. http://doi.org/10.1097/00001756-200107030-00018

Miltner, W. H. R., Braun, C. H., & Coles, M. G. H. (1997). Event-Related Brain Potentials Following Incorrect Feedback in a Time-Estimation Task : Evidence for a " Generic " Neural System for Error Detection. *Journal of Cognitive Neuroscience*, *9*(6), 788–798. http://doi.org/10.1162/jocn.1997.9.6.788

Möller, J., Jansma, B. M., Rodríguez-Fornells, A., & Münte, T. F. (2007). What the brain does before the tongue slips. *Cerebral Cortex*, *17*(5), 1173–1178. http://doi.org/10.1093/cercor/bhl028

Morey, R. D. (2008). Confidence Intervals from Normalized Data : A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *4*(2), 61–64.

Müller, S. V., Möller, J., Rodríguez-Fornells, A., & Münte, T. F. (2005). Brain potentials related to self-generated and external information used for performance monitoring. *Clinical Neurophysiology*, *116*(1), 63–74. http://doi.org/10.1016/j.clinph.2004.07.009

Muscalu, L. M., & Smiley, P. A. (2018). The illusory benefit of cognates: Lexical facilitation followed by sublexical interference in a word typing task. *Bilingualism: Language and Cognition*, 1–18.

http://doi.org/10.1017/S1366728918000792

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: evidence from an antisaccade task. *Psychophysiology*, *38*(5), 752–760. http://doi.org/10.1111/1469-8986.3850752

Nieuwenhuis, S., Ridderinkhof, K. R., Talsma, D., Coles, M. G. H., Holroyd, C. B., Kok, A., & Van Der Molen, M. W. (2002). A computational account of altered error processing in older age: Dopamine and the error-related negativity. *Cognitive, Affective and Behavioral Neuroscience*, *2*(1), 19–36. http://doi.org/10.3758/CABN.2.1.19

Nieuwenhuis, S., Yeung, N., Holroyd, C. B., Schurger, A., & Cohen, J. D. (2004). Sensitivity of electrophysiological activity from medial frontal cortex to utilitarian and performance feedback. *Cerebral Cortex*, *14*(7), 741–747. http://doi.org/10.1093/cercor/bhh034

Nooteboom, S. G. (2005). Lexical bias revisited: Detecting, rejecting and repairing speech errors in inner speech. *Speech Communication*, *47*(1–2), 43–58. http://doi.org/10.1016/j.specom.2005.02.003

Nooteboom, S. G., & Quené, H. (2017). Self-monitoring for speech errors: Two-stage detection and repair with and without auditory feedback. *Journal of Memory and Language*, *95*, 19–35. http://doi.org/10.1016/j.jml.2017.01.007

Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. *Cognitive Psychology*, *63*(1), 1–33. http://doi.org/10.1016/j.cogpsych.2011.05.001

Nozari, N., Freund, M., Breining, B., Rapp, B., & Gordon, B. (2016). Cognitive control during selection and repair in word production. *Language, Cognition and Neuroscience*, *31*(7), 886–903. http://doi.org/10.1080/23273798.2016.1157194

Nozari, N., & Hepner, C. R. (2018). To select or to wait? The importance of criterion setting in debates of competitive lexical selection. *Cognitive Neuropsychology*, 1–15. http://doi.org/10.1080/02643294.2018.1476335

Nozari, N., & Novick, J. (2017). Monitoring and Control in Language Production. *Current Directions in Psychological Science*, *26*(5), 403–410. http://doi.org/10.1177/0963721417702419

Nozari, N., & Pinet, S. (2020). A critical review of the behavioral, neuroimaging, and electrophysiological studies of co-activation of representations during word production. *Journal of Neurolinguistics*, *53*(February 2020), 100875. http://doi.org/10.1016/j.jneuroling.2019.100875

O'Connell, R. G., Dockree, P. M., Bellgrove, M. A., Kelly, S. P., Hester, R., Garavan, H., … Foxe, J. J. (2007). The role of cingulate cortex in the detection of errors with and without awareness: A high-density electrical mapping study. *European Journal of Neuroscience*, *25*(8), 2571–2579. http://doi.org/10.1111/j.1460-9568.2007.05477.x

Oomen, C. C. E., & Postma, A. (2001). Effects of time pressure on mechanisms of speech production and self-monitoring. *Journal of Psycholinguistic Research*, *30*(2), 163–184.

http://doi.org/10.1023/A:1010377828778

Oomen, C. C. E., Postma, A., & Kolk, H. H. J. (2001). Prearticulatory and Postarticulatory Self-Monitoring in Broca's Aphasia. *Cortex*, *37*(5), 627–641. http://doi.org/10.1016/S0010-9452(08)70610-5

Overbeek, T. J. M., Nieuwenhuis, S., & Ridderinkhof, K. R. (2005). Dissociable components of error processing: On the functional significance of the Pe vis-à-vis the ERN/Ne. *Journal of Psychophysiology*, *19*(4), 319–329. http://doi.org/10.1027/0269-8803.19.4.319

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–47. http://doi.org/10.1017/S0140525X12001495

Pinet, S., & Nozari, N. (2018). "Twisting fingers": The case for interactivity in typed language production. *Psychonomic Bulletin and Review*, *25*(4), 1449–1457. http://doi.org/10.3758/s13423-018-1452-7

Pinet, S., Ziegler, J. C., & Alario, F.-X. (2016). Typing is writing: Linguistic properties modulate typing execution. *Psychonomic Bulletin and Review*, *23*(6), 1898–1906. http://doi.org/10.3758/s13423-016-1044-3

Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition*, *77*(2), 97–131. http://doi.org/10.1016/S0010-0277(00)00090-1

Postma, A., & Kolk, H. H. J. (1992). The Effects of Noise Masking and Required Accuracy on Speech Errors, Disfluencies, and Self-Repairs. *Journal of Speech, Language, and Hearing Research*, *35*(3), 537–544. http://doi.org/10.1044/jshr.3503.537

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*(4), 375–392. http://doi.org/10.1177/002383099603900403

Potts, G. F., Martin, L. E., Burton, P., & Montague, P. R. (2006). When Things Are Better or Worse than Expected: The Medial Frontal Cortex and the Allocation of Processing Resources. *Journal of Cognitive Neuroscience*, *18*(7), 1112–1119. http://doi.org/10.1162/jocn.2006.18.7.1112

Potts, G. F., Martin, L. E., Kamp, S. M., & Donchin, E. (2011). Neural response to action and reward prediction errors: Comparing the error-related negativity to behavioral errors and the feedback-related negativity to reward prediction violations. *Psychophysiology*, *48*(2), 218–228. http://doi.org/10.1111/j.1469-8986.2010.01049.x

Python, G., Fargier, R., & Laganaro, M. (2018). ERP evidence of distinct processes underlying semantic facilitation and interference in word production. *Cortex*, *99*, 1–12. http://doi.org/10.1016/j.cortex.2017.09.008

Ridderinkhof, K. R., van den Wildenberg, W. P. M., Segalowitz, S. J., & Carter, C. S. (2004). Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain and Cognition*,

56(2), 129–40. http://doi.org/10.1016/j.bandc.2004.09.016

Riès, S. K., Janssen, N., Dufau, S., Alario, F.-X., & Burle, B. (2011). General-purpose monitoring during speech production. *Journal of Cognitive Neuroscience*, *23*(6), 1419–36. http://doi.org/10.1162/jocn.2010.21467

Riès, S. K., Xie, K., Haaland, K. Y., Dronkers, N. F., & Knight, R. T. (2013). Role of the lateral prefrontal cortex in speech monitoring. *Frontiers in Human Neuroscience*, *7*(October), 703. http://doi.org/10.3389/fnhum.2013.00703

Rousselet, G. A., Foxe, J. J., & Bolam, J. P. (2016). A few simple steps to improve the description of group results in neuroscience. *European Journal of Neuroscience*, *44*(9), 2647–2651. http://doi.org/10.1111/ejn.13400

Ruiz, M. H., Jabusch, H.-C., & Altenmüller, E. (2009). Detecting Wrong Notes in Advance: Neuronal Correlates of Error Monitoring in Pianists. *Cerebral Cortex*, *19*(11), 2625–2639. http://doi.org/10.1093/cercor/bhp021

Rumelhart, D. E., & Norman, D. A. (1982). Simulating a Skilled Typist : A Study of Skilled Cognitive-Motor Performance. *Cognitive Science*, *6*, 1–36.

Sebastian-Gallés, N., Rodríguez-Fornells, A., De Diego-Balaguer, R., & Díaz, B. (2006). First- and second-language phonological representations in the mental lexicon. *Journal of Cognitive Neuroscience*, *18*(8), 1277–1291. http://doi.org/10.1162/jocn.2006.18.8.1277

Shalgi, S., & Deouell, L. Y. (2012). Is any awareness necessary for an Ne? *Frontiers in Human Neuroscience*, *6*(May), 1–15. http://doi.org/10.3389/fnhum.2012.00124

Snyder, K. M., Logan, G. D., & Yamaguchi, M. (2015). Watch what you type: The role of visual feedback from the screen and hands in skilled typewriting. *Attention, Perception, & Psychophysics*, *77*(1), 282–292. http://doi.org/10.3758/s13414-014-0756-6

Steinhauser, M., & Yeung, N. (2010). Decision Processes in Human Performance Monitoring. *Journal of Neuroscience*, *30*(46), 15643–15653. http://doi.org/10.1523/JNEUROSCI.1899-10.2010

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: a user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience*, *2011*, 879716. http://doi.org/10.1155/2011/879716

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, *39*(3), 1429–1443. http://doi.org/10.1016/j.neuroimage.2007.09.054

Ullsperger, M., Fischer, A. G., Nigbur, R., & Endrass, T. (2014). Neural mechanisms and temporal dynamics of performance monitoring. *Trends in Cognitive Sciences*, *18*(5), 259–267. http://doi.org/10.1016/j.tics.2014.02.009

Van Schie, H. T., Mars, R. B., Coles, M. G. H., & Bekkering, H. (2004). Modulation of activity in medial frontal and motor cortices during error observation. *Nature Neuroscience*, *7*(5), 549–554. http://doi.org/10.1038/nn1239

Van Wijk, C., & Kempen, G. (1987). A dual system for producing self-repairs in spontaneous speech: Evidence from experimentally elicited corrections. *Cognitive Psychology*, *19*(4), 403–440.

Wessel, J. R. (2012). Error awareness and the error-related negativity: evaluating the first decade of evidence. *Frontiers in Human Neuroscience*, *6*(April), 1–16. http://doi.org/10.3389/fnhum.2012.00088

Yamaguchi, M., Crump, M. J. C., & Logan, G. D. (2013). Speed-accuracy trade-off in skilled typewriting: decomposing the contributions of hierarchical control loops. *Journal of Experimental Psychology. Human Perception and Performance*, *39*(3), 678–99. http://doi.org/10.1037/a0030512

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*(4), 931–959. http://doi.org/10.1037/0033-295X.111.4.939

APPENDIX

Error types

We coded errors according to their error types, namely addition, deletion and substitution of a letter (see Table 1). Only the first error of the word was classified.

Table S1. Subtypes of segmental errors analyzed in the present study.

TARGET	TANK
ADDITION	tankg
DELETION	tan
SUBSTITUTION	tang
EXCHANGE	takn

Metacognitive reports and corrections by error types

Table S2. Proportion of each error type that is corrected or reported in each condition.

	CORRECTIONS		REPORTS
	immediate	*delayed*	*delayed*
ADDITION	43.6%	2.7%	55.4%
DELETION	58.6%	12.2%	42.7%
SUBSTITUTION	53.4%	13.0%	52.6%

Note that for the purpose of the current experiment, i.e. analyzing EEG data, only the first error in the word has been coded. Therefore, these data are not representative of the proportion of all errors in the sample.

Correction by distance from the error

In the graph below we see that about 75% of corrections are implemented right after or one keystroke after the incorrect keystroke (similarly in both conditions).
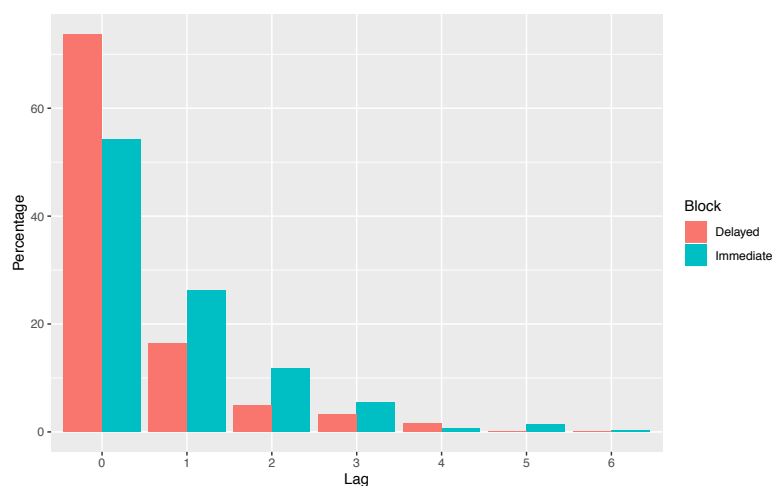

Figure S1. Proportion of corrections according to the distance from the error. Percentages are computed by condition separately