

## Observed and Performed Error Signals in Auditory Lexical Decisions

Duru G. Özkan,<sup>a,b,\*</sup> Mirjam Broersma,<sup>c</sup> Harold Bekkering<sup>b</sup> and Sybrine Bultena<sup>b,c</sup>

<sup>a</sup> Department of Psychology, Sapienza University of Rome, Italy

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

<sup>c</sup> Centre for Language Studies, Radboud University Nijmegen, Netherlands

**Abstract**—This study investigates the error processing components in the EEG signal of Performers and Observers using an auditory lexical decision task, in which participants heard spoken items and decided for each item if it was a real word or not. Pairs of participants were tested in both the role of the Performer and the Observer. In the literature, an Error Related Negativity (ERN)-Error Positivity (Pe) complex has been identified for performed (ERN-Pe) and observed (oERN-oPe) errors. While these effects have been widely studied for performance errors in speeded decision tasks relying on visual input, relatively little is known about the performance monitoring signatures in observed language processing based on auditory input. In the lexical decision task, native Dutch speakers listened to real Dutch Words, Non-Words, and crucially, long Pseudowords that resembled words until the final syllable and were shown to be error-prone in a pilot study, because they were responded to too soon. We hypothesised that the errors in the task would result in a response locked ERN-Pe pattern both for the Performer and for the Observer. Our hypothesis regarding the ERN was not supported, however a Pe-like effect, as well as a P300 were present. Analyses to disentangle lexical and error processing similarly indicated a P300 for errors, and the results furthermore pointed to differences between responses before and after word offset. The findings are interpreted as marking attention during error processing during auditory word recognition. © 2021 The Author(s). Published by Elsevier Ltd on behalf of IBRO. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

**Key words:** auditory lexical decision, error processing, P300, performance monitoring, observation.

### INTRODUCTION

Detection of errors is crucial for adaptive behaviours (Falkenstein et al., 1991; Gehring, et al., 1993; Cavanagh and Frank, 2014; Koban et al., 2012; Luu et al., 2004; Ullsperger et al., 2014a). The performance monitoring system thought to be responsible for error detection has been suggested to be social by nature; it does not only detect errors of our own goal directed actions, but also those of others (e.g., De Bruijn et al., 2007, 2011; Navarro-Cebrian et al., 2016). Much work on performed and observed errors is based on evidence from visual stimuli, where errors are easily observable for Performers and Observers alike (e.g., Miltner et al., 2004; Panasiti et al., 2016; Joch et al., 2017). However, little is known about monitoring the errors of our own and of others within the auditory domain. In the present study, we focus on error signals during auditory decision making, specifically in a lexical decision task. Auditory processing differs in its nature from visual processing, as the signal unfolds over time, which means that a prediction on accuracy is more difficult and may need to be

updated during word processing. Detection of errors may therefore also be more challenging than in tasks that rely on visual cues only. In this light, we address the question to what extent error processing signatures are present for the auditory modality for Performers and Observers during a lexical decision task.

Subprocesses of one's own performance monitoring have been studied extensively using EEG (De Bruijn and Von Rhein, 2012; Luu et al., 2004; Ullsperger et al., 2014b). In the time domain, two event related potentials (ERPs) have been identified to be related to error monitoring processes: The Error Related Negativity (ERN) and Error Positivity (Pe; Luck and Kappenman, 2011). The ERN is a negative deflection peaking fronto-centrally around ~100 ms after an error has been committed. It is thought to be generated in the anterior cingulate cortex (ACC), which is involved in cognitive control and adaptive functions (Holroyd and Coles, 2002). The Pe is a positive deflection that usually follows the ERN, with maximal amplitude over the central-parietal area (Shalgi et al., 2009; Wessel, 2012; Pezzetta et al., 2018), and is thought to reflect error awareness, and context updating (Nieuwenhuis et al., 2001). Typically, the ERN seems to be present for errors made during speeded choice tasks, such as in the Flanker and Go-No Go that usually require

\*Corresponding author.

E-mail address: [duru.ozkan@isikun.edu.tr](mailto:duru.ozkan@isikun.edu.tr) (D. G. Özkan).

<https://doi.org/10.1016/j.neuroscience.2021.02.001>

0306-4522/© 2021 The Author(s). Published by Elsevier Ltd on behalf of IBRO.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

a button-press, as well as reach-to grasp tasks with embodied virtual avatars (Riesel et al., 2013; Moreau et al., 2020; Pavone et al., 2016), which involve the detection of a visual mismatch between the correct and incorrect response. Yet, the ERN effect is not limited to visually detected errors involving motor cues. Although relatively few ERN studies have been conducted outside the action domain, it has been shown that the ERN and Pe can also be elicited for internal monitoring of one's own speech (Arnstein et al., 2011; Ganushchak et al., 2011; Masaki et al., 2001; Riès et al., 2011). Similar neural signatures have been observed for decisions based on auditory processing; Sebastian-Gallés et al. (2006) have demonstrated the occurrence of an ERN for erroneous responses to an auditory lexical decision task, which included non-words that differed from words by only a subtle acoustic feature.

Monitoring of others' actions has been widely studied in the context of motor simulation (Keysers and Perrett, 2004; Rizzolatti and Craighero, 2004; Kilner et al., 2007). The process of motor simulation involves understanding the goal and intention of an action, and whether it is performed correctly (Panasiti et al., 2017). Processing of others' actions shows more motor activation in case of errors compared to correct responses (Aglioti et al., 2008; Koelewijn et al., 2008). In keeping with the motor simulation account, processing during the observation of others' motor errors yields electrophysiological markers in the brain that are similar to internal error monitoring, known as the observational ERN (oERN) and observational Pe (oPe; Bediou et al., 2012; Koban and Pourtois, 2014). The oERN has been shown to have a longer latency and a smaller amplitude compared to the classical ERN, for example during the observation of Flanker task performance (Van Schie et al., 2004). In the Van Schie paradigm, which showed an oERN effect, the Observers were given explicit information regarding the correct response on a display separate from the Performers, in order to avoid complicating the task of the Observer. This may, however, not resemble authentic error observation. Observers are not typically provided with explicit information of others' performance, yet are able to detect a mismatch between accurate and faulty behaviour by relying on their internal monitoring system. In an effort to control for this, Bates et al. (2005) conducted a Go/No Go task where the Observers could see the same visual display as the performer. The oERN in this study showed little delay in comparison to the ERN of Performers, which the authors explained by the fact that both participants were able to see the same and thus shared the same mind-set.

It is important to note that observations or erroneous actions do not always elicit oERN or oPe components. A study by De Bruijn et al. (2007) showed participants sequences of pictures that depicted everyday action errors; when the final picture violated the expected pattern, a P300 rather than an oERN was observed, possibly due to the study's design that did not involve an explicit forced choice decision. The occurrence of a typical attention component can be explained by the idea that general monitoring of unexpected events requires high levels of attention (Polich, 2007).

In line with the idea that motor simulation may underlie the observed error monitoring (Bates et al., 2005), in the studies conducted so far, visual processing was key to detecting the error. The present study examines observational error processing in the auditory domain. We created a setting to address the performance monitoring processes involved in observation of language processing. Language is a useful tool to address the question whether observed errors are marked with the typical error processing ERPs, fitting with the call to test performance monitoring in more realistic settings (Wessel, 2014). We investigated errors in auditory word recognition by means of a lexical decision task, in which participants decided whether a presented stimulus is a real word or not (Goldinger, 1996). Crucially, in auditory perception, the listener has to wait for the stimulus to unfold. Detecting an auditory mistake is therefore typically contingent on hearing the auditory presentation of the stimulus in full or, at least, up until the word uniqueness point. Each word has a uniqueness point from which they can unambiguously be recognised; hearing more of the stimulus implies that other options, so-called lexical competitors that are activated upon hearing part of the word, are ruled out and a single lexical candidate remains (Norris and McQueen, 2008).

In this study, we created an auditory lexical decision task that required the listener to make a speeded decision on whether the stimuli they heard were Words or Non-Words, and the Observer to be attentive to this procedure. Typically, a lexical decision task is a relatively simple task; to make the task more challenging, the stimuli included not only real Words and obvious Non-Words, but also Pseudowords that were particularly intended to elicit errors. All Non-Words and Pseudowords conformed to Dutch phonotactic requirements, i.e., they were possible but non-existing Dutch words. The obvious Non-Words were short and were expected to elicit few errors (e.g. *blij* 'happy' → *blooi*). The Pseudowords were long, and were not revealed as Non-Words until the end of the auditory presentation, after the listener would have predicted the stimulus to be a real word (e.g. *koffiezetapparaat* 'coffee maker' → *koffiezetapparaatas*). The late deviation point was thus thought to lead word prediction processes astray. Pseudowords were therefore expected to result in some responses being given early due to the prediction of it being a real word, and some to be given after the completion of the item (Early: response given before stimulus completion; Late: response given after stimulus completion). We applied time pressure to enhance the occurrence of earlier, and thus likely erroneous, responses. Therefore, an analysis of Late responses provided an overview of when the items were fully heard, while an Early response analysis focused on the errors that were made due to incorrect lexical completion (Astheimer and Sanders, 2011; MacGregor et al., 2012; Gagnepain et al., 2012; Baart and Samuel, 2015). Furthermore, the ratio of Words to Non-Words (the latter including Pseudowords) in the experiment was 2:1, to bias participants towards giving yes-

responses, thus enhancing the occurrence of errors on the Pseudowords.

As per previous evidence (Sebastian-Gallés et al., 2006), we hypothesised that the erroneous responses (yes-responses to Pseudowords) in the present auditory lexical decision task should be marked with a response-locked ERN for Performers; additionally, we expected an oERN for Observers. We further expected the ERN for errors to be followed by stronger Pe and oPe for errors marking the subject's awareness of the error. We thus hypothesised that the oERN and oPe should be present when the task requires the Observer to do the task themselves internally, while also observing their partner perform the auditory lexical decision task, similar to Bates et al. (2005). This would show the error monitoring mechanism is also shared for the errors that are committed during observation in auditory lexical decision. Indirectly, if oERN and oPE are present, the role of motor simulation transpires for decision making in the auditory modality. However, errors committed on the Pseudowords may require more attention because they take longer to process because of their resemblance to real words up until near word offset, especially when responses are made under time pressure. Therefore, as an alternative to observational error monitoring, the dual task requirement for the Observers could result in more attention-driven components such as the P300 (akin to De Bruijn et al., 2007).

## EXPERIMENTAL PROCEDURES

### Participants

Twenty pairs of participants were tested, whose only native language was Dutch. Participants were between the ages of 18–35 ( $M_{Age}$ : 22.4,  $SD$ : 4.2; 28 females), and recruited through the SONA system; 12 pairs knew each other before the experiment. All forty participants were right-handed, had normal or corrected-to-normal vision, and did not have hearing problems, colour blindness, dyslexia or neurological issues.

### Materials

The lexical decision task included items of three different types: Words (240), Non-Words (90), both of which served as filler items, and word-like Pseudowords (30) that served as target items, for the reason of being error-prone. Given the two-part nature of the task, in which participants once performed the task themselves and once observed the performance of their partner, two comparable sets of 360 experimental items were created.

In total, 480 Dutch Words of varying word classes (74% nouns, 11% verbs, and 15% adjectives) were selected. A further 180 Non-words and 60 Pseudowords conforming to Dutch phonotactic requirements were created, which resulted in a total of 720 items for two lists. Selected Words could either be short (1–6 syllables) or long (7–10 syllables). Non-Words varied in length between 1 and 6 syllables, and Pseudowords measured 7–10 syllables. All Pseudowords were Nouns,

in agreement with the majority of Word items (more than 70%). The items were divided over two lists. Numbers of items per list are shown in Table 1.

Words had high lexical frequencies according to the Dutch CELEX corpus (Baayen et al., 1995) and SUBTLEX-NL (Keuleers, Brysbaert, and New, 2010). None had active derivational suffixes (-heid, -teit, -isch, -lijk, -ig) or were inflected word forms (e.g., plurals), because these complicate the prediction of the word endings. Furthermore, none were generic compounds, reduced forms, names, taboo words, or emotionally loaded words (Moors et al., 2013).

Pseudowords were created by changing the final syllable of long words, such that they were comparable to real words until the last syllable, to allow the predictive process to complete the word before the end was known. To do so, we either added a syllable at the end of the word (e.g., *koffiezetapparaat* 'coffee maker' → *koffiezetapparaatas*), added one or more consonants at the end of the last syllable (e.g., *verantwoordelijkheidsgevoel*, 'feeling of responsibility' → *verantwoordelijkheidsgevoelt*), and/or replaced the vowel or final consonant(s) in the last syllable (*scheikundelaboratorium* 'chemistry laboratory' → *scheikundelaboratoriuuf*).

Obvious Non-Words were not identical to any existing Dutch word and deviated from existing words early on. They were created by changing a single sound (e.g., *spek* 'bacon' → *spak*), replacing several syllables or combining parts from multiple words, resulting in nonsense words (e.g., '*kleuzelschetter*' or '*abanteurenmaroon*'). Words with strong phonological similarity to German or English were also avoided, as these were likely to be known by our participants. An additional 8 Words were selected, and 4 Non-Words created to form 6-item practice parts for both lists.

All Words, Pseudowords and Non-Words were recorded by a female native Dutch speaker in a soundproof booth. Recordings were segmented at word boundaries to create audio files per item, which were cut at the zero crossings right before stimulus onset and after stimulus offset respectively (Avg. length 1100 ms; Min. 342 ms, Max. 2302 ms) using PRAAT software (Boersma and Weenink, 2019).

### Procedure

Participants were asked to come to the lab in pairs and were seated in a comfortable chair. After having signed the consent forms, they were explained the procedure and mounted with the EEG caps simultaneously by two researchers. The session was composed of two tasks, both of which involved an Observation and a Performance part. The order of the Observer/Performer roles was assigned randomly. Upon completion of the first task, the participants changed roles, and were presented with the second experimental list. After completion of both tasks, they were asked to fill out a post-experiment questionnaire. The full experiment lasted 2 h. Each participant was compensated with €20 for their time.

**Table 1.** Overview of Item Types per List

# Of Syllables	Words	Non-words	Pseudowords	Word Length
1	30	15	–	Short
2	30	15	–	
3	30	15	–	
4	30	15	–	
5	30	15	–	
6	30	15	–	
7	30	–	15	Long
8	10	–	5	
9	10	–	5	
10	10	–	5	
<b>TOTAL</b>	<b>240</b>	<b>90</b>	<b>30</b>	

### Experimental Set-up

Participants sat facing each other, with a table placed between them. The set-up included a touchscreen that was embedded in a table, on which two response buttons were presented, such that the Observer could easily see the hands of the Performer. The performing participant had both of their hands on planks placed on either side over the touch screen, with their index fingers hovering over the response buttons. To give a response, they touched a green box on the right side of the touch screen, if they thought what they had heard was a real word, or a red box on the left side for a Non-Word. Participants were encouraged to respond as soon as possible, stressing that they did not need to wait until the end of the item had been heard.

The task was run with Presentation® software (Version 20.0, Neurobehavioral Systems, Inc., Berkeley, CA, [www.neurobs.com](http://www.neurobs.com)). The auditory stimuli were delivered via speakers and presented in a pseudorandomized order for each list. Same item types (Words, Non-words, Pseudowords) were not repeated more than five times in a row, and no consecutive items had the same number of syllables. The task lasted ~ 25 min delivered in five blocks, interrupted by short breaks. Each block included 72 items. In order not to create noticeable patterns, the number of different item types per block was not counterbalanced.

### Auditory Lexical Decision Task

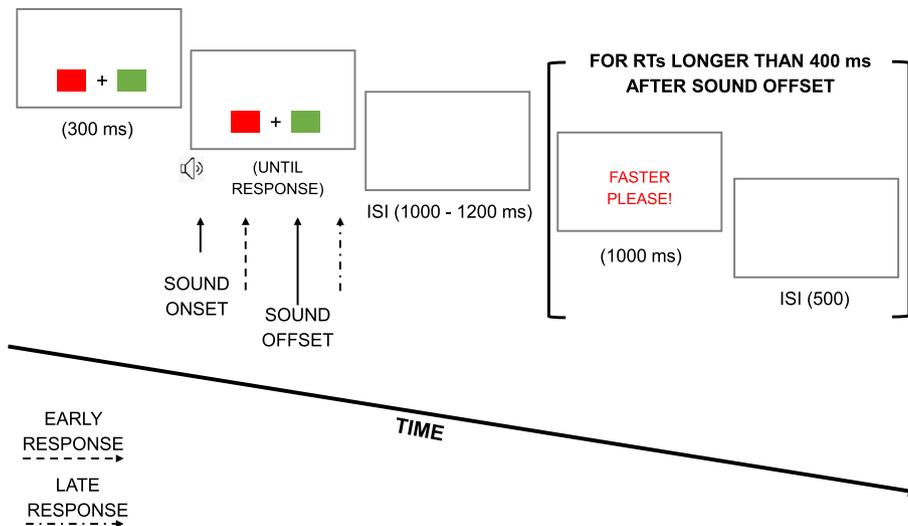
The Performer's task was to decide whether the item they heard was a real word or not and respond accordingly using the buttons presented on the touchscreen. The Observer's task was to pay attention to the auditory item, as well as the Performer's response on the touch screen (green for 'word', red for 'not a word'). Additionally, both partners had to keep track of the number of incorrect responses to ensure the errors were being attended to. After each block, participants were asked to write down their estimates on how many errors they thought they had observed or committed in the preceding block, depending on their role; and not to worry about the exact number. This was done to ensure the Observer's attention throughout the session. Response hands were kept the same across rounds to make it simpler for the Observer to discriminate the Performer's response on the touch screen.

Each trial started with a 300 ms fixation cross and the response buttons presented on a screen. Then, an auditory item was presented via speakers. Participants could respond during or after the item had been completed. In an attempt to increase the number of errors, a tight stimulus-dependent response deadline was set. As determined by behavioural pilot sessions (see [Suppl. Materials: Fig. 1](#)), the time out duration was set to 400 ms after the offset of the auditory stimulus (see e.g., [Ernestus and Cutler, 2015](#)). For responses within the time limit, a white inter-stimulus-interval screen was presented for 1000–1200 ms before the next sound file was played. If the response was given more than 400 ms after the offset of the item, the ISI was followed by a message ('Sneller A.U.B.! – *Faster please!*) that told participants to respond faster (see [Fig. 1](#)). Participants were not given feedback regarding their or their partner's performance.

### EEG recordings and preprocessing

The EEG recordings were made using BrainVision Recorder and two Brain Products ActiCaps. Each cap had 32 electrodes including the external ocular and reference electrodes. The signals were acquired from 27 channels: FP2, F4, F8, FC6, C4, T8, CP2, CP6, O2, P4, Cz, Pz, Oz, O1, P3, CP5, CP1, T7, C3, FC1, FCz, FC2, Fz, FC5, F7, F3, Fp1. Reference electrodes were applied to the left (digital reference) and right (physical reference) mastoids. All electrodes were algebraically re-referenced offline to the average of both mastoid electrodes. The Horizontal Electro-Oculogram (HEOG) was recorded bilaterally, and the Vertical (VEOG) with electrodes positioned above and below the left eye. The signal was recorded with an online band-pass filter (0.01–250 Hz) with a sampling rate of 500 Hz. Impedances were maintained under 5 K $\Omega$  by applying gel (ElectroGel) to all electrodes.

All preprocessing of the EEG data was carried out using FieldTrip, an open source toolbox (Donders Institute, Nijmegen; [Oostenveld et al., 2011](#)) in MATLAB (The MathWorks, Inc.). Removal of visual artefacts was done in two steps. First, a blind source separation method, the Independent Component Analysis (ICA; [Jung et al., 2000](#)), was applied to remove the components

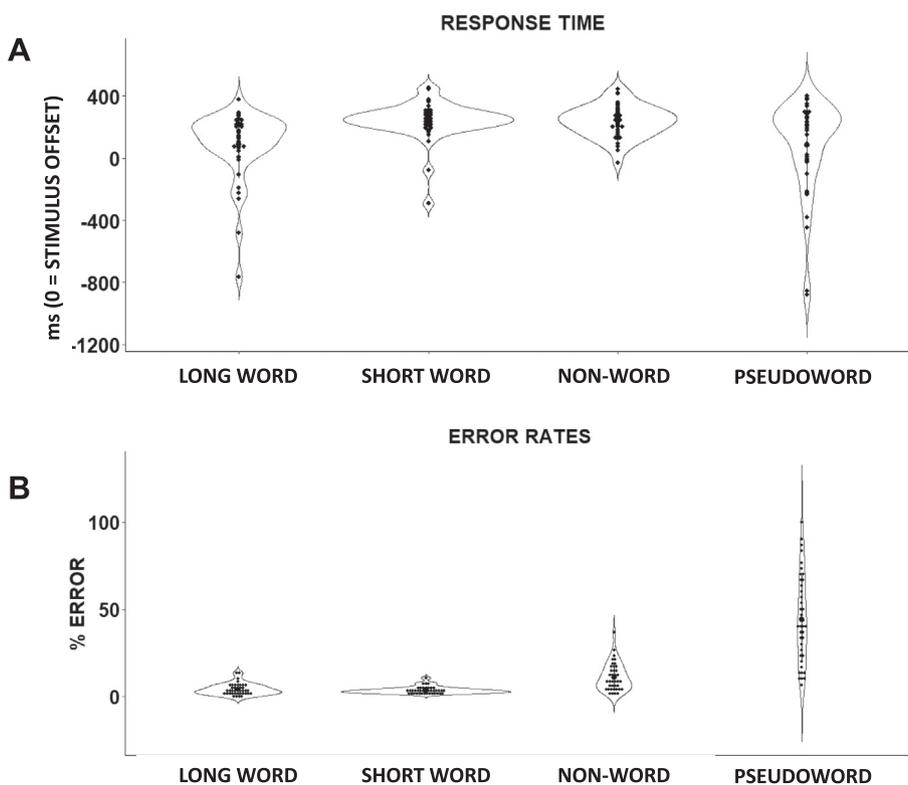


**Fig. 1.** Sequence of one trial. Each trial started with a presentation of the buttons and fixation cross for 300 ms, followed by the sound presentation. Responses could be given either during (Early response) or after (Late response) the sound. Upon response, the buttons disappeared from the screen, and a blank interval screen was presented. Duration of the sound files varied between 394 and 2468 ms. Note: The blank screen after the response was jittered, and so was the interval after the 'Faster Please!' message. When the response was within the time limits and the 'Faster Please!' message was not presented; the first blank screen was the only jitter between trials.

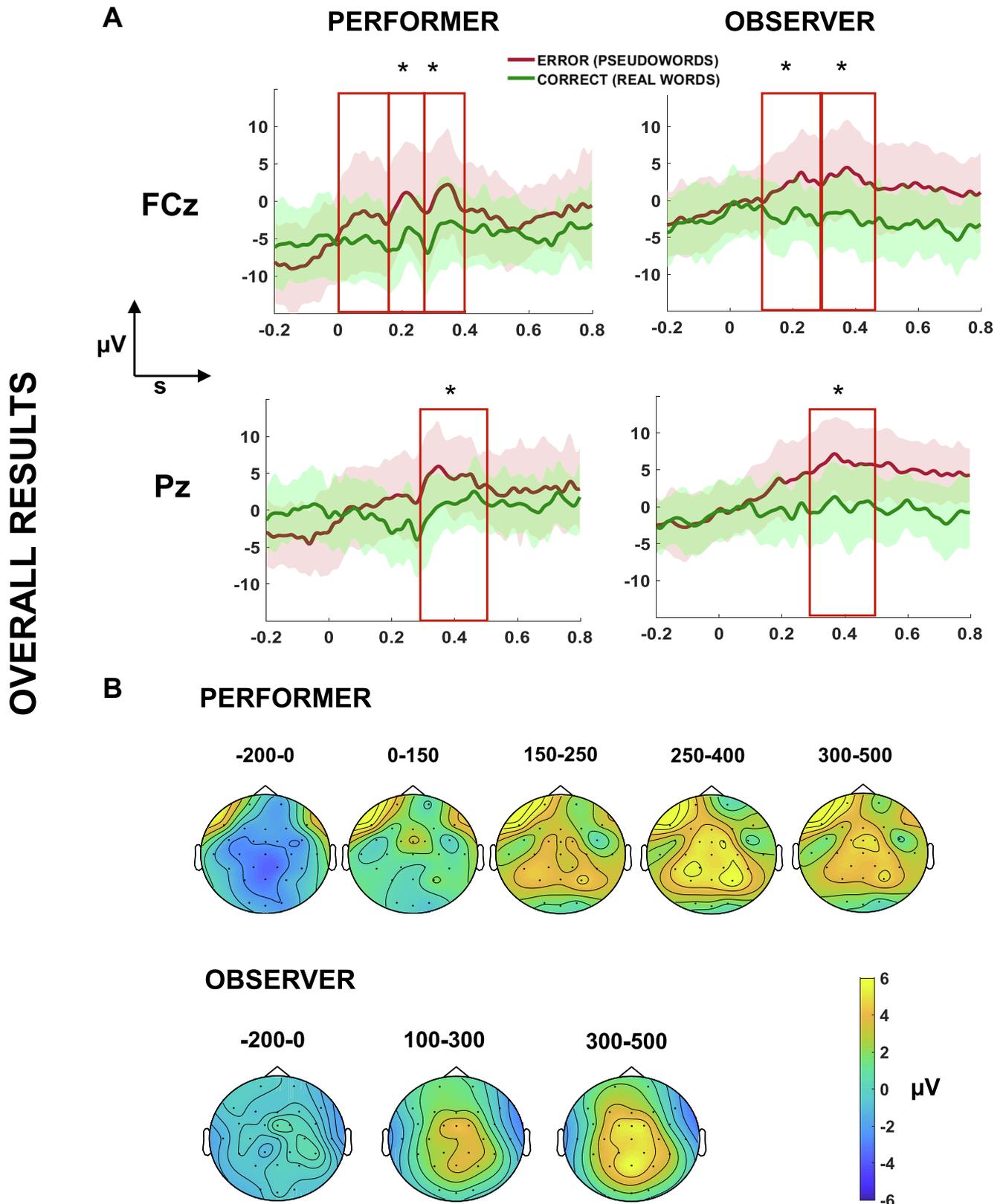
that are related to eye movements. Trials with response times more than two standard deviations away from the mean were removed from both behavioural and EEG analysis, matched between the Performers and the Observers. This removed a further average of 0.9 trials ( $SD = 1.16$ ) from each condition that was analysed. Then, all trials were visually inspected for residual eye blinks and other artefacts, which removed an average of 14.8 trials per participant ( $SD = 8.43$ ).

### Design and analysis

The dependent variables in this study included behavioural and EEG measures based on task performance (described in separate sections below), which were analysed for effects of accuracy (correct/error). Performer's and Observer's data were analysed separately. Across all analyses, the Performer's correct trials (yes-responses to real Words) were compared to the erroneous trials (yes-responses to Pseudowords). No-responses to Words were not used as errors because they might include a differential processing compared to yes-response to Words, which could confound the results if averaged with the yes-response to Pseudowords. Similarly, correct No-responses to Pseudowords were not used in order not to confound the correct responses to Words. Note that Flanker tasks have selectively analysed performance on incongruent trials (e.g., Danielmeier et al., 2009). For each participant, the number of error responses was matched with an equal number of randomly selected correct responses. These correct responses were selected based on being comparable to the erroneous Pseudowords in terms of length. This meant that from the total set of 240 Words of each participant, only a subset of correct responses to the 60 long



**Fig. 2.** Behavioural data: (A) Error rates for each type of stimuli (B) RTs to each type of stimuli. Note that time 0 represents the sound offset. Negative reaction times represent the responses that were given before the sound offset. Long words were between 7 and 10 syllables long, whereas short word ranged between 1 and 6 syllables.



**Fig. 3. (A)** ERP waveforms time-locked to responses for each group; the red line represents error trials on Pseudowords (yes response to Pseudowords), the green line shows correct trials on Words (yes response to Words). The y-axis shows  $-10$  to  $10 \mu\text{V}$ , negative down; the x-axis shows a 200 ms pre-response period; time 0 indicates the response time. Red boxes indicate the time windows of interest for ERN, oERN Pe, oPe, and for the P300. (\*) indicates the statistical significance. Note that the baseline correction was done using a 200 ms interval *before* the Stimulus Onset, therefore the pre-response interval on the figure does not reflect the baseline. **(B)** Scalp distributions: error – correct.

Words (7–10 syllables) was considered for analysis (see Table 1). For sake of the ERP analyses, a minimum of 6 trials were averaged over per condition<sup>1</sup> (Olvet and Hajcak, 2009; Pontifex et al., 2010).

T-tests were used to compare the responses between correct (yes to Real-words) and error trials (yes to Pseudowords). In order for both of the participants in a pair to be assigned to each role within a two-hour session, the number of items was limited to 360 per list. This meant that several other comparisons beyond the main hypothesis (including those between item types) were not possible due to the relatively small number of target items per list. An overview of analysis parameters used for these potentials in the literature can be found in the Suppl. Materials (see Table 8 and 9).

### Behavioural analyses

The behavioural data were analysed in terms of error rates and Reponse Times (RTs). In line with studies on auditory language processing (e.g., Ernestus and Cutler, 2015), we measured RTs from word offset, as the late deviation point in Pseudowords meant that lexical decisions could only be properly made once the item had been presented in full. 7 participants were excluded from analyses due to an insufficient number of erroneous trials (see Suppl. Materials Table 7 for an overview of the number of trials, and trials that were removed due to too long RTs). Furthermore, we have separately analysed responses that were made earlier and later than the sound offset. Earlier responses were made due to our manipulation of the time limit, so that more errors were made.

### ERP analyses

The ERPs were calculated as mean amplitudes, time-locked to the Performer's response. These were obtained by segmenting the signal into epochs of 1000 ms length (–200 ms to +800 ms, relative to the performer's response time; see Fig. 4), which were band-pass filtered offline (0.1–30 Hz). Baseline correction was done using the time window from 200 milliseconds before the presentation of the auditory stimulus, which was well before the response onset (Beyersmann et al., 2019).

Statistical analyses on response-locked ERPs were performed with Matlab. Based on the ERN, Pe and P3a/b literature, analyses are focused on FCz and Pz electrodes (Danielmeier et al., 2009; Clayson et al., 2013; Comerchero and Polich, 1999; De Bruijn et al., 2007). Response-locked negativities were defined as the mean amplitude on FCz (Clayson et al., 2013) in two time windows: the 0–150 ms window for ERN in Performers (Danielmeier et al., 2009; Kaczurkin, 2013; Wessel and Ullsperger, 2011; Endrass et al., 2008;

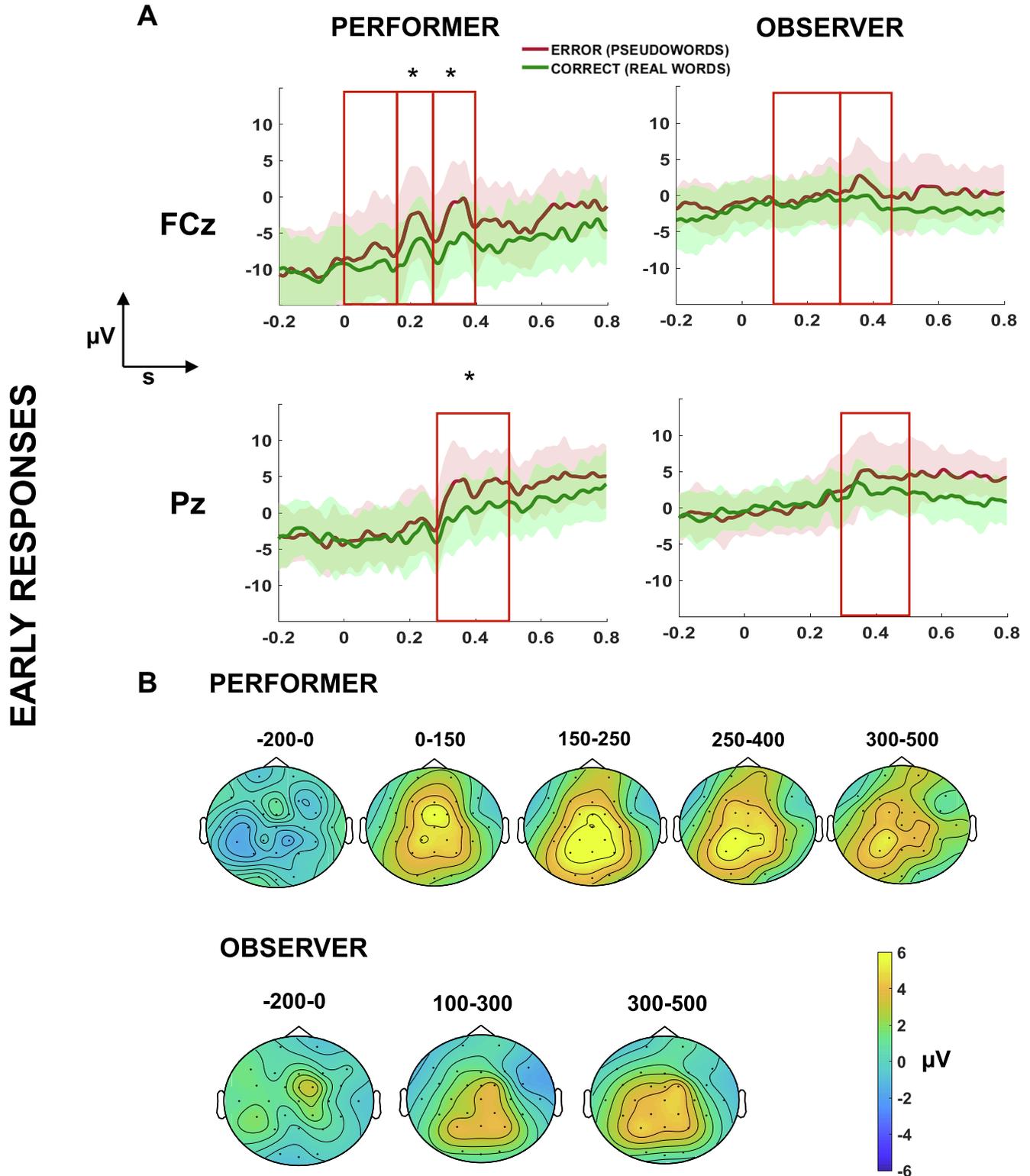
Zambrano-Vazquez and Allen, 2014); and the 100–300 ms window for oERN in Observers (Pavone et al., 2016; Pezzetta et al., 2018). The time window of interest for Pe was defined as the mean amplitude at FCz in the time window immediately following the (o)ERN: 150–250 ms post-response time for Performers (Bediou et al., 2012) and between 300 and 450 ms for Observers. Furthermore, a third time window was analysed at FCz for the Performers between 250 and 400 ms, due to a difference that was visible in the ERP waveform. This exploratory analysis was done to detect a potentially significant effect. Upon the inspection of the topographical distribution, we also analysed the P3b, a subcomponent of P300 and typically a marker for attention allocation, which was determined as the mean amplitude in the 300–500 ms post-response time window at Pz (Comerchero and Polich, 1999; Polich, 2007; Volpe et al., 2007; De Bruijn, 2007). The P300 time window was equal for Performers and Observers; we expected the attention allocation to occur simultaneously for both in relation to response processing, as the participants' task was to pay attention to the response.

## RESULTS

### Overview of task performance (EEG)

Response times were analysed relative to the stimulus offset; negative response times indicated the response was made before the sound offset, and positive response times indicated that the response was made after the sound offset. Overall, participants were 91% accurate on 360 trials. On average, there were ~2 errors ( $SD = 1.95$ ) out of 60 Long Words (4%); ~6 errors ( $SD = 3.55$ ) out of 180 Short Words (3.60%), ~10 errors ( $SD = 7$ ) out of 90 Non-Words (11%), and ~13 errors ( $SD = 7$ ) out of 30 Pseudowords (44%). Error responses ( $M = 137$  ms;  $SD = 205$  ms) were executed faster than correct responses ( $M = 209$  ms;  $SD = 116$  ms ( $t(39) = -3.85$ ,  $p < 0.001$ ). Overall, RTs to Pseudowords ( $M = 74$  ms,  $SD = 313$ ) were comparable with RTs to Long Words ( $M = 88$  ms,  $SD = 218$ ). Correct responses to (Long) Words were made after word offset at 64.65 ms ( $SD = 249.75$ ), while erroneous responses to Pseudowords were typically made before word offset, at –108.71 ms ( $SD = 251.70$ ). Participants' guess accuracy difference scores for own errors were computed by subtracting "own" estimates from the actual numbers of errors, and the partners' guess accuracy difference scores were computed by subtracting "partner" estimates from the actual numbers of errors. These estimates were calculated from the guesses they made at each break. Performers on average estimated their errors more accurately (0.02 more than actual errors  $SDs = 10.60$ ), than Observers (2.64 more than actual errors  $SDs = 14.50$ ; see Suppl. Material Table 1 for full details on the calculation of the difference scores). It must be noted that the average difference score for the Observers reflects variable overestimations and underestimations of actual errors.

<sup>1</sup> During the revisions, we noticed that one participant was included to the analyses with 5 trials, instead of the 6 trial minimum.



**Fig. 4. (A)** ERP waveforms time-locked to Early responses for each group, for trials that were responded *before* the sound offset; red line represents error trials on Pseudowords (yes response to Pseudowords), green line shows correct trials on Words (yes response to Words). The y-axis shows  $-10$  to  $10 \mu\text{V}$ , negative down; the x-axis shows a 200 ms pre-response period; time 0 indicates the response time. Red boxes indicate the time windows of interest for ERN, oERN, Pe, oPe, and for the P300. (\*) indicates the statistical significance. **(B)** Scalp distributions: error - correct.

## Event related potentials

Seven of the 40 participants were excluded from analyses due to having fewer than 5 error trials on Pseudowords. On average, there were 15 error trials ( $SD = 6.64$ ) for Pseudowords, therefore a matching number of correct trials was subselected. Overall, the results showed similar ERP patterns for Performers and Observers, albeit not the typical biphasic ERN-Pe complex (Fig. 4; Suppl. Materials Table 2).

## Performer

No negative peaks were observed in either correct or error trials, and no significant difference between error and correct responses was observed for Performers in the window of the standard ERN peak (0–150 ms:  $t(32) = 1.99$ ,  $p = .055$ ). Contrary to expectation, mean amplitudes indicated that errors yielded numerically smaller negativities than correct responses ( $M_{Error} = -3.74 \mu V$ ,  $SD = 7.89$ ,  $M_{Correct} = -5.82 \mu V$ ,  $SD = 6.52$ ). In the Pe window at FCz, erroneous trials showed significantly less negative amplitudes than correct responses (150–250 ms:  $t(32) = -2.74$ ,  $p = 0.01$ ;  $M_{Error} = -2.21 \mu V$ ,  $SD = 7.18$ ,  $M_{Correct} = -5.59 \mu V$ ,  $SD = 5.71$ ), indicating the expected positivity. Upon visual inspection of the ERP waveforms, an additional window of analysis at the 250–400 ms interval was included, where error trials were significantly less negative in amplitude compared to correct trials in FCz (Performer:  $t(32) = 3.03$ ,  $p = 0.048$ ;  $M_{Error} = -0.18 \mu V$ ,  $SD = 7.44$ ,  $M_{Correct} = -3.44 \mu V$ ,  $SD = 5.95$ ). For the P3b window, the ERP waveform in the Pz electrode was significantly more positive for the error trials than correct trials (300–500 ms:  $t(32) = 3.70$ ,  $p = .0008$ ;  $M_{Error} = 4.45 \mu V$ ,  $SD = 5.65$ ,  $M_{Correct} = 1.11 \mu V$ ,  $SD = 4.16$ ), indicating a P3b effect.

## Observer

In the observed-ERN window, error trials were significantly more positive in amplitude compared to correct responses as can be seen at FCz in Fig. 4, (100–300 ms:  $t(32) = 3.76$ ,  $p = 0.0007$ ;  $M_{Error} = 1.92 \mu V$ ,  $SD = 5.75$ ,  $M_{Correct} = -2.04 \mu V$ ,  $SD = 4.68$ ), opposite to the classic oERN. In the oPe window at FCz, errors showed significantly more positive waveforms compared to correct trials (300–400 ms:  $t(32) = 4.24$ ,  $p = .0002$ ;  $M_{Error} = 2.99 \mu V$ ,  $SD = 6.96$ ,  $M_{Correct} = -2.21 \mu V$ ,  $SD = 5.85$ ). Similarly to the Performers, a significant P3b effect at Pz in erroneous trials was present for the Observers (300–500 ms:  $t(32) = 4.11$ ,  $p = 0.0003$ ;  $M_{Error} = 5.63 \mu V$ ,  $SD = 5.65$ ,  $M_{Correct} = 1.36 \mu V$ ,  $SD = 4.16$ ).

## Early Responses

To further control for the times at which responses were made, post hoc analyses were done on subselected responses that were given earlier and later than the sound offset separately (Fig. 4, Early responses; Fig. 5, Late responses). Participants that had fewer than 5

trials per category were not included in the analyses; this meant that for the Early responses 20/40 and for the Late responses 16/40 participants were included in the analyses. For the Early response analyses on remaining participants, there were approximately 13 error trials on average per Performer (and therefore Observer) ( $SD = 6.81$ ), and for Late responses there were 8 trials ( $SD = 4.32$ ) (Fig. 6).

**Performer:** No significant difference was observed in the standard ERN window (0–150 ms:  $t(19) = 1.40$ ,  $p = .18$ ;  $M_{Error} = -7.71 \mu V$ ,  $SD = 7.98$ ,  $M_{Correct} = -9.55 \mu V$ ,  $SD = 6.12$ ). In the Pe window and beyond, ERP amplitudes of errors were significantly less negative compared to correct trials at FCz (150–250 ms:  $t(19) = -2.25$ ,  $p = .037$ ;  $M_{Error} = -4.24 \mu V$ ,  $SD = 8.21$ ,  $M_{Correct} = -7.65 \mu V$ ,  $SD = 6.76$ ; 250–400 ms:  $t(19) = 3.60$ ,  $p = .002$ ;  $M_{Error} = -3.01 \mu V$ ,  $SD = 6.48$ ,  $M_{Correct} = -6.80 \mu V$ ,  $SD = 6.08$ ). In the P3b window at Pz, errors in Pseudowords had a significantly more positive amplitude compared to correct trials on Words (300–500 ms:  $t(19) = 2.89$ ,  $p = 0.009$ ;  $M_{Error} = 3.68 \mu V$ ,  $SD = 5.59$ ,  $M_{Correct} = 0.33 \mu V$ ,  $SD = 5.13$ ).

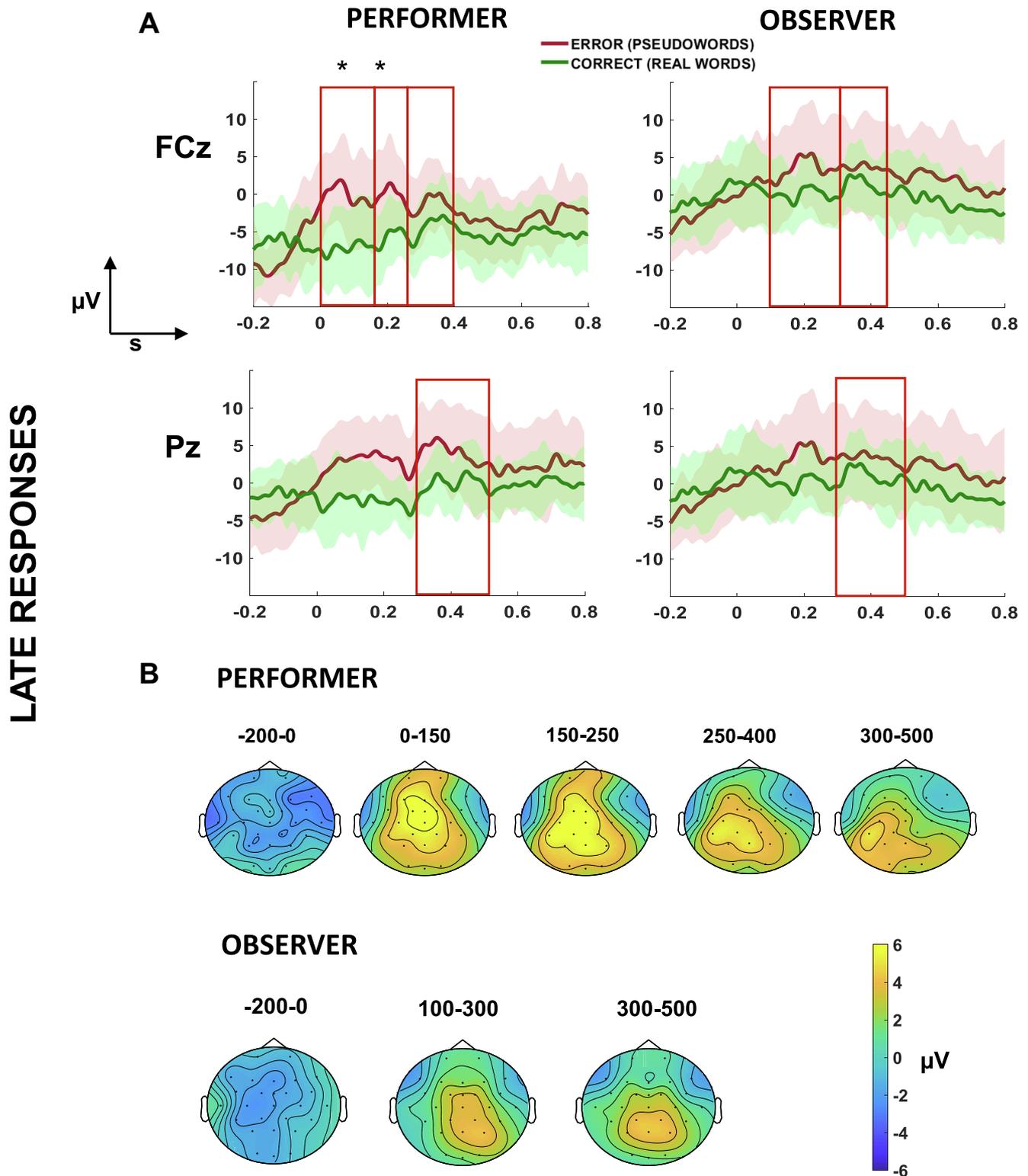
**Observer:** There were no significant differences in the oERN window, in the oPe window, or in the P3b window (FCz, 100–300 ms:  $t(19) = 0.769$ ,  $p = 0.45$ ;  $M_{Error} = -0.33 \mu V$ ,  $SD = 4.35$ ,  $M_{Correct} = -0.89 \mu V$ ,  $SD = 4.17$ ; 300–400 ms:  $t(19) = 1.74$ ,  $p = .09$ ;  $M_{Error} = 1.16 \mu V$ ,  $SD = 5.99$ ,  $M_{Correct} = -0.65 \mu V$ ,  $SD = 3.99$ ; Pz, 300–500 ms:  $t(19) = 1.39$ ,  $p = .11$ ;  $M_{Error} = 4.11 \mu V$ ,  $SD = 4.11$ ,  $M_{Correct} = 2.39 \mu V$ ,  $SD = 3.46$ ). An overview for the Early Responses can be found in Suppl. Material Table 3.

## Late responses

**Performer:** Error amplitudes were significantly less negative compared to correct trials on FCz in the ERN and Pe windows (0–150 ms:  $t(15) = 3.86$ ,  $p = 0.002$ ;  $M_{Error} = -0.01 \mu V$ ,  $SD = 6.86$ ,  $M_{Correct} = -7.21 \mu V$ ,  $SD = 6.56$ ; 150–250 ms:  $t(15) = -2.80$ ,  $p = 0.0134$ ;  $M_{Error} = -0.02 \mu V$ ,  $SD = 7.14$ ,  $M_{Correct} = -5.82 \mu V$ ,  $SD = 5.21$ ). No significant differences were found in the 250–400 ms window on FCz ( $t(15) = 1.90$ ,  $p = .08$ ;  $M_{Error} = -1.34 \mu V$ ,  $SD = 5.54$ ,  $M_{Correct} = -4.45 \mu V$ ,  $SD = 5.55$ ) or in the P3b window on Pz (300–500 ms:  $t(15) = 1.47$ ,  $p = .26$ ;  $M_{Error} = 4.18 \mu V$ ,  $SD = 7.02$ ,  $M_{Correct} = 0.07 \mu V$ ,  $SD = 4.5$ ).

**Observer:** Contrary to the Performers, no significant differences were found for the Observers in any of the time windows of interest (FCz, 100–300 ms:  $t(16) = 1.91$ ,  $p = 0.08$ ;  $M_{Error} = 3.74 \mu V$ ,  $SD = 7.39$ ,  $M_{Correct} = 0.07 \mu V$ ,  $SD = 4.80$ ; 300–400 ms  $t(16) = 1.50$ ,  $p = .16$ ;  $M_{Error} = 3.91 \mu V$ ,  $SD = 7.51$ ,  $M_{Correct} = 1.32 \mu V$ ,  $SD = 4.81$ ; Pz, 300–500 ms:  $t(16) = 1.9$ ,  $p = 0.07$ ;  $M_{Error} = 8.19 \mu V$ ,  $SD = 8.14$ ,  $M_{Correct} = 3.79 \mu V$ ,  $SD = 4.85$ ). An overview for the Early Responses can be found in Suppl. Material Table 4.

In sum, the behavioural results showed that Pseudowords yielded most errors, which were predominantly due to premature responses and usually

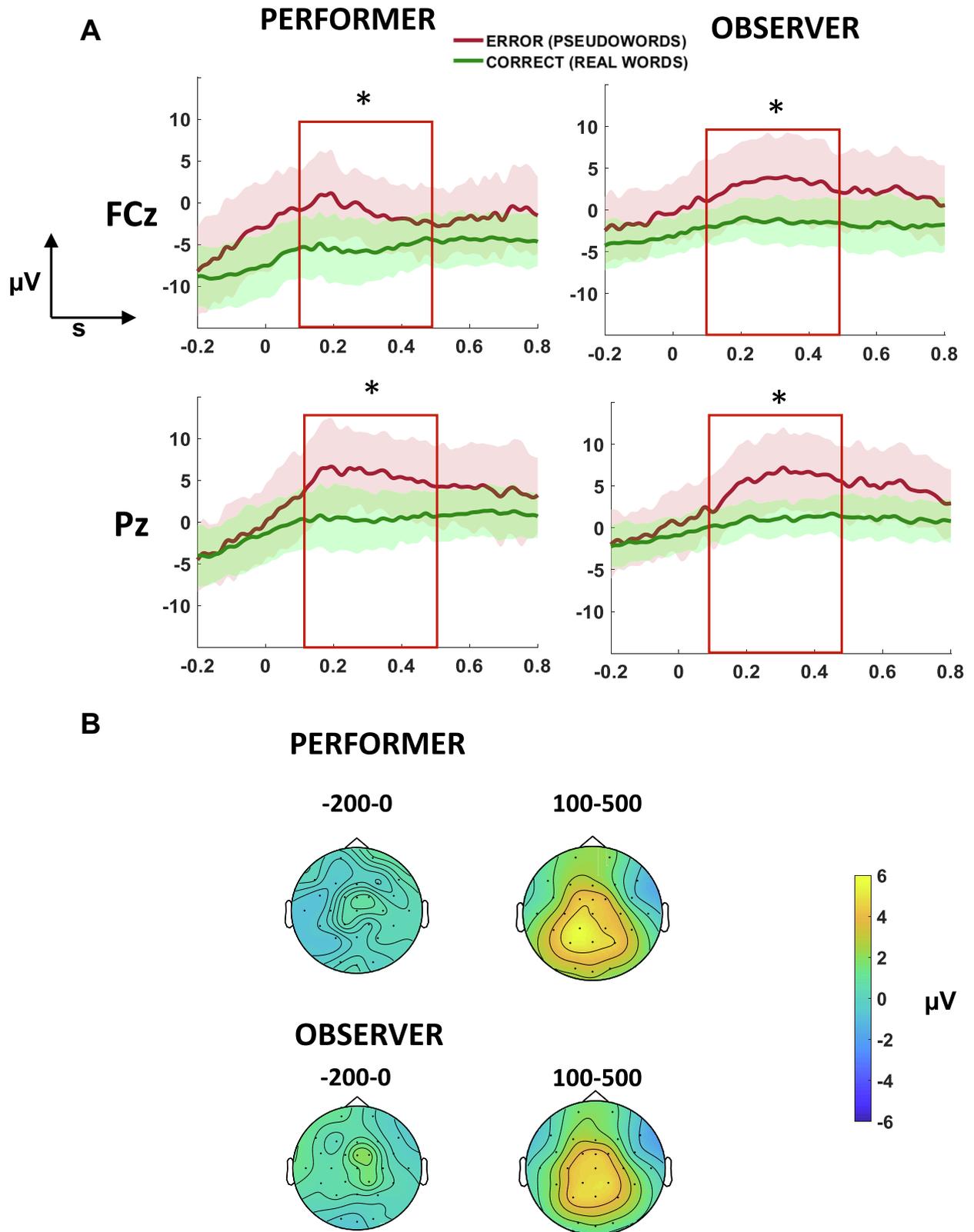


**Fig. 5. (A)** ERP waveforms time-locked to Late responses for each group, for trials that were responded *after* the sound offset; red line represents error trials on Pseudowords (yes response to Pseudowords), green line shows correct trials on Words (yes response to Words). The y-axis shows  $-10$  to  $10 \mu\text{V}$ , negative down; the x-axis shows a 200 ms pre-response period; time 0 indicates the response time. Red boxes indicate the time windows of interest for ERN, oERN Pe, oPe, and for the P300. (\*) indicates the statistical significance. **(B)** Scalp distributions: error – correct.

noticed by Performers and, to lesser extent, Observers. The overall ERP analyses on mean amplitudes showed

less negative waveforms for errors compared to correct responses, which were significantly different for

## DEVIATION ANALYSES



**Fig. 6. (A)** ERP waveforms time-locked to the deviation point for Pseudowords, and  $-200$  ms before the sound offset for the Words for each group; red line for error trials in pseudowords (yes response to pseudowords), green line for correct trials in Words (yes response to Words). The y-axis:  $-10$  to  $10$   $\mu\text{V}$ , negative down; the x-axis shows  $-200$  pre-response period; time 0 indicates the response time. Red boxes indicate the time windows of interest for ERN, oERN Pe, oPe, and for the P300. (\*) indicates the statistical significance. **(B)** Scalp distributions: error – correct.

Performers in the Pe and a subsequent window, and for both ERN and Pe windows for Observers. A P3b effect for errors was visible for both Performers and Observers. When the data were split by response latency, early responses for Performers showed smaller negativities for errors in the Pe window and beyond, in combination with a P3b. Late responses for Performers showed smaller negativities in the ERN and Pe windows, but no subsequent effects or a P3b. Late responses in the Observer data yielded no effects whatsoever. Overall, it can be noted that effects in the window directly following the response (ERN) were generally absent; such effects only surfaced when measured in slightly later time windows that allowed for more word processing, such as the oERN or the ERN for late responses. Later error monitoring responses (Pe and beyond) and P3b were present across the board, albeit not for Observers when the responses were split in early and late responses. Based on these results, it seems as if lexical processing may have partly overlapped with error monitoring. To look into this issue in more depth, we have subsequently performed analyses that took the point at which Pseudowords deviated from real words into account, in an effort to disentangle lexical and error processing.

### Deviation analysis

In order to have an overview of the influence of the word processing itself, the analyses were repeated from the deviation point as the time0. This was done to have an indication of whether the response-locked ERP waveform (reflecting error processing) overlapped with lexical processing. We repeated the analysis from the deviation point of the Pseudowords in the present paradigm (Friedrich et al., 2006). For a comparison with correct responses, the deviation point from the Pseudowords was then matched to an arbitrary deviation point 200 ms before the sound offset in Words. The deviation points of the Pseudowords varied due to the number of different manipulations that were added in the creation of these items.

The topography shows a significant positivity centred in the mid-parietal regions for errors in Pseudowords compared to the correct responses to Words at the 200–500 ms interval in both groups (Performer: FCz,  $t(29) = 6.3$ ,  $p = .0000007$ , Pz,  $t(29) = 7.15$ ,  $p = .00000007$ ; Observer: FCz,  $t(31) = 5.31$ ,  $p = .000009$ , Pz,  $t(31) = 5.49$ ,  $p = 0.000001$ ). We only found a positive peak following the Pseudowords but not any negative peaks. An overview of the deviation analyses can be found in [Suppl. Material Table 5](#).

## DISCUSSION

In this study, we aimed to detect error-related EEG activity in an auditory lexical decision task, both for performance errors (i.e. ERN-Pe, in the Performer role), and for observed errors (i.e. oERN-oPe, in the Observer role). Innovatively, we investigated whether similarities in the error processing signatures between performed

and observed errors exist in the auditory domain. The task was designed in a way to allow us to compare correct response signals (yes response to Words) to error signals elicited by the incorrect prediction (yes response to Pseudowords) of the stimuli. We manipulated the target stimuli (Pseudowords), which included a deviation from regular long words at the end of the item and expected that the incorrect prediction of the auditorily presented item would lead to erroneous responses. The manipulation worked: most errors were present on Pseudowords. We also observed similar ERP responses for Performers and Observers. Contrary to previous findings, our hypothesis that the ERN should be present in the ERP waveform in performed and observed errors was not supported, however, we found a Pe-like waveform, as well as a P300 that can either reflect attention to response errors or attention to the deviant stimuli.

For the Performers and Observers, we observed similar patterns in the ERP waveform in response to both error and correct trials. The overall results showed no ERN for either Performers, or Observers in the auditory modality, contrary to classic ERN studies (Holroyd and Coles, 2002). While Performers showed no effect in the ERN time window, Observers showed the opposite pattern to what was expected, i.e., less negative amplitudes to the errors. This is in contrast to the ERN effect in a previous auditory lexical decision task, based on correct and incorrect vowel pronunciations (Sebastian-Gallés et al., 2006). Similarly, we did not find an oERN for the Observers, contrary to Van Schie et al. (2004), but again observed that error trials were significantly more positive than correct trials.

The absence of ERN effects may be due to the way we designed the task, and the timing of the response deadline in particular. The response deadline was set at 400 ms following stimulus offset, as this manipulation was necessary to ensure erroneous responses. This deadline is compatible with the time it takes to activate the meaning of a word, as indicated by the renowned N400 component, which peaks around 250–400 after stimulus onset and is considered an index of full lexical access (Kutas and Federmeier, 2011). Although lexical processing should have been completed by the time participants had to give their response, processing in the auditory modality may have been more demanding, especially in the case of long word-like stimuli. However, an additional post-hoc analysis we carried out to check whether ERN effects were present for filler items that were short (1–6 syllables) and which were typically responded to after word offset ( $M = 239.41$  ms,  $SD = 102.30$ ) suggests that word length, or the extent to which lexical items were processed in full, cannot explain the absence of error monitoring effects. This comparison between error trials on Non-Words and correct responses to short Words showed a P300 effect for the Performers, but no significant effects in the ERN and Pe time windows were found (see [Suppl. Materials: ii. Control analysis of the short items; Table 6 and Fig. 2](#)). Nevertheless, these findings could tentatively be interpreted to suggest a lack of similarity between error processing in

the visual and auditory domains. Previous studies that compared language processing in the auditory modality to the visual modality have shown that the N400 typically lasts longer in the former, which has been explained by the fact that presentation of an auditory signal unfolds over time (Kutas and Van Petten, 1994; Kutas and Federmeier, 2011). Yet, previous evidence for error monitoring effects in an auditory lexical decision task (Sebastian-Gallés et al., 2006) suggests auditory lexical decisions, as such, are not responsible for the pattern of results observed in the present study. We do note that error percentages were much higher in the study by Sebastián-Gallés et al. (more than 60% errors on a total of 120 target Non-Words), which means that the relatively small number of trials in our task may have prevented the detection of error related activity in an auditory setting. Regardless of the power issues, we believe that the lack of significant differences in the present study in the (o) ERN window between correct and error trials, for long target items as well as short filler items, is more likely to have been due to limited processing time allowed in our task, for both the stimuli itself, and the observation of performance. The absence of an early error detection component (i.e., ERN) could suggest that lexical processing was still ongoing as the response was being processed. Such an interpretation implies that the two processes involved, lexical processing of the auditory stimulus and error processing of the response, cannot be properly disentangled in the present design. Fig. 3.

Although we did not find an ERN-Pe complex in its canonical form, we did find a positive peak in the 150–250 ms interval similar to Pe for the Performers' errors, suggesting that error monitoring may have been present at a slightly later stage. In the error monitoring literature, the Pe in the visual domain has previously been observed in the absence of the negative ERN peak (Gibbons et al., 2011; Di Gregorio et al., 2018), although the process of error detection had typically been complicated in such studies. Another difference to the typical Pe concerns the second positive peak that surfaced in the 250–400 ms window (see Suppl. Materials B. i.), which we have included in the analyses post hoc. We are unaware of other studies that have found a second response-locked positivity, but the Pe has previously been likened to a P3 effect, which in a prior study has been observed to consist of multiple positive peaks (P3i and P3ii) across the midline in stimulus-locked analyses of an Eriksen-Flanker task (Davies, Segalowitz, Dywan, and Pailing, 2001). In the absence of alternative explanations, we tentatively interpret the observed positivity as a Pe-like effect. Pe has been reported to be dependent on the detection of an erroneous action (Panasi et al., 2016), error awareness (Nieuwenhuis et al., 2011), or the conscious evaluation of errors as part of a monitoring system (Di Gregorio et al., 2018). Such an interpretation is fitting with our data. Based on overall guess accuracies regarding the number of errors, we can claim that the errors were well attended to by Performers, which indicates error awareness and could explain the Pe-like waveform. Yet, standard deviations for the guess accuracies were relatively high (see Suppl. Material: Table 1),

which implies that errors for Pseudowords may not always have been attended to. This holds in particular for the guesses by Observers; where guesses by Performers in most cases deviated no more than 5 items from the actual number of errors, guesses by Observers showed relatively large overestimations as well as underestimations: in a majority of cases guesses and actual numbers differed by more than 10 items (on a total of 360 items). Nonetheless, the Pe-like finding is robust in that it consistently appeared in both Early (both groups) and Late (only the Performers) response results.

A central finding in our results was the auditory P3b (Comerchero and Polich, 1999; Polich, 2007; Volpe et al., 2007). As a subcomponent of P300, P3b is elicited by attention allocation to deviant stimuli. Previously, in a study where participants observed everyday action errors, a posterior P300 has been found in the absence of ERN for more lifelike action errors (De Bruijn et al., 2007). Despite substantial differences in task design, we have found a similar pattern of P3b. As argued above, it is possible that the time pressure associated with the response deadline of 400 ms made prompt error detection a challenge. It may have been too short to observe error monitoring signatures in the auditory lexical decision task, or ERP effects may have been masked by an overlap of lexical processing and error processing signals. The prominence of P300 effects in several analyses implies that attentional processing overruled error monitoring. Such an explanation is supported by the notion that language processing in the auditory domain, in contrast to the visual domain, demands full attention of the listener from the beginning until the end of the stimulus, as a participant has no control over the stimulus input. Linguistic processing of visual stimuli allows for more control, as the participant is able to maintain fixations on a deviant word ending for as long as it is presented (Kutas and Federmeier, 2011). Furthermore, the fact that Observers were not explicitly informed about response accuracy meant that Observers were not only (i) internally deciding whether the stimulus was a word or not, but also (ii) whether their partner had made the correct decision. In the present study, The combination of online internal processing of long word-like stimuli combined with the observation of button presses may have resulted in a rather demanding dual task, which arguably could have resulted in a higher cognitive load than for the observation of Flanker task performance, where stimuli remain on screen until the response is given (cf. Bates et al., 2005). We therefore interpret the significant large positive deflection in the parietal region following the erroneous responses of both Performers and Observers to be a response-locked P3b effect, as a sign of attention allocation arising from the fact that the participants were explicitly instructed to keep track of the error responses. However, the deviation point-locked analysis for Pseudowords compared to Words showed a positivity that was temporally widespread for the Pseudowords. Despite the arbitrary nature of this analyses that should be interpreted with caution, we cannot rule out that this P3b effect may additionally have been caused by attention to deviant stimuli. It must be noted that the P3b effect was not present in all condi-

tions; the Early responses for in the Performer group indicated a P3b for errors, similar to the overall results, but no such effect was present for the Observers. In contrast, the Late responses showed no P3b effects in either group. This might be because separation of these two types of responses has less statistical power, so the results of these analyses should be interpreted with caution.

## CONCLUSIONS AND FUTURE DIRECTIONS

In this study, we aimed to integrate the EEG error-related brain responses during action observation (Van Schie et al., 2004; De Bruijn et al., 2011; De Bruijn and Von Rhein, 2012; Pezzetta et al., 2018), with EEG responses associated with auditory lexical error processing (Sebastian-Gallés et al., 2006). As a first attempt in investigating observational error processing in auditory lexical decisions, we showed that the errors in our task result in a Pe-like signal, thought to mark error awareness, despite the absence of an ERN. Furthermore, a stronger P3b effect was elicited for attention allocation to response errors (on Pseudowords) compared to correct trials (on Words). As expected, the Observers' signal has shown similar patterns to that of the Performers', suggesting that error processing of observed errors is processed similarly to one's own errors, also in the auditory domain concerning lexical decisions.

We addressed the question whether the error monitoring processes is present in other modalities than the purely visual paradigm. Our task, admittedly, still involved visual error detection to some extent, but can be considered a first step to addressing the complexities in performance monitoring in during lexical decision making and the observed performance in the auditory domain. A follow up study that additionally allows for a comparison with passively listened to stimuli could help disentangle lexical processing (stimulus-locked) from the error processing signals (response-locked). Furthermore, in an attempt to decrease the assumed cognitive load for the Observer in an auditory task, it would be an option to manipulate the latency of stimuli. They could be presented to the Performers with delay, such that Observers hear the stimuli a little before the Performers, which would allow for more time to process the item and judge their partners' performance on the task. Future studies could also address the question of whether stimulus related effects differ between item categories. More generally, the investigation of error processing in the auditory modality could benefit from time-frequency and source analysis in order to determine the involvement of performance monitoring processes at a more detailed level (Cohen et al., 2008).

Ultimately, this study can be considered as an exploration on how much the error monitoring processes are present in an observed auditory lexical decision task. Importantly though, while the error signals in the performance monitoring literature have a consistent latency, the influence of lexico-semantic processing on error processing mechanisms should be addressed in more detail with further studies.

## FUNDING

This research was supported by a mobility grant for PhD students from Sapienza, University of Rome, awarded to D.Ö.; a Vidi grant from the Dutch Research Council (NWO), grant number 276-89-006, awarded to M.B.; and by the Dutch Research Grant 024.001.006 to the consortium Language in Interaction awarded to H.B.

## ACKNOWLEDGEMENTS

We would like to thank Susanne Brouwer for lending her voice to the recordings, Niklas Frechen for work on the stimulus materials, and Anna Aumeistere and Marta Blasco-Oliver for testing participants.

## REFERENCES

- Aglioti SM, Cesari P, Romani M, Urgesi C (2008) Action anticipation and motor resonance in elite basketball players. *Nat Neurosci* 11 (9):1109–1116. <https://doi.org/10.1038/nn.2182>.
- Arnstein D, Lakey B, Compton RJ, Kleinow J (2011) Preverbal error-monitoring in stutterers and fluent speakers. *Brain Lang* 116 (3):105–115. <https://doi.org/10.1016/j.bandl.2010.12.005>.
- Astheimer LB, Sanders LD (2011) Predictability affects early perceptual processing of word onsets in continuous speech. *Neuropsychologia* 49(12):3512–3516. <https://doi.org/10.1016/j.neuropsychologia.2011.08.014>.
- Baart M, Samuel AG (2015) Early processing of auditory lexical predictions revealed by ERPs. *Neurosci Lett* 585:98–102. <https://doi.org/10.1016/j.neulet.2014.11.044>.
- Baayen RH, Piepenbrock R, Gulikers L, Linguistic Data Consortium (1995) *The CELEX Lexical Database (CD-ROM)*. Philadelphia, PA: University of Pennsylvania.
- Bates AT, Patel TP, Liddle PF (2005) External behavior monitoring mirrors internal behavior monitoring: error-related negativity for observed errors. *J Psychophysiol* 19(4):281–288. <https://doi.org/10.1027/0269-8803.19.4.281>.
- Bediou B, Koban L, Rosset S, Pourtois G, Sander D (2012) Delayed monitoring of accuracy errors compared to commission errors in ACC. *NeuroImage* 60(4):1925–1936. <https://doi.org/10.1016/j.neuroimage.2012.02.010>.
- Beyersmann E, Bolger D, Pattamadilok C, New B, Grainger J, Ziegler JC (2019) Morphological processing without semantics: An ERP study with spoken words. *Cortex* 116:55–73. <https://doi.org/10.1016/j.cortex.2019.02.008>.
- Boersma P, Weenink D. (2019). Praat: Doing phonetics by computer. Version 6.0.52.
- Cavanagh JF, Frank MJ (2014) Frontal theta as a mechanism for cognitive control. *Trends Cogn Sci* 18(8):414–421. <https://doi.org/10.1016/j.tics.2014.04.012>.
- Clayton PE, Baldwin SA, Larson MJ (2013) How does noise affect amplitude and latency measurement of event-related potentials (ERPs)? A methodological critique and simulation study: ERP measurement. *Psychophysiology* 50(2):174–186. <https://doi.org/10.1111/psyp.12001>.
- Cohen MX, Ridderinkhof KR, Haupt S, Elger CE, Fell J (2008) Medial frontal cortex and response conflict: Evidence from human intracranial EEG and medial frontal cortex lesion. *Brain Res* 1238:127–142. <https://doi.org/10.1016/j.brainres.2008.07.114>.
- Comerchero MD, Polich J (1999) P3a and P3b from typical auditory and visual stimuli. *Clin Neurophysiol* 110(1):24–30. [https://doi.org/10.1016/s0168-5597\(98\)00033-1](https://doi.org/10.1016/s0168-5597(98)00033-1).
- Danielmeier C, Wessel JR, Steinhauser M, Ullsperger M (2009) Modulation of the error-related negativity by response conflict. *Psychophysiology* 46(6):1288–1298. <https://doi.org/10.1111/j.1469-8986.2009.00860.x>.
- Davies P, Segalowitz S, Dywan J, Pailing P (2001) Error-negativity and positivity as they related to other ERP indices of attentional

- control and stimulus processing. *Biol Psychol* 56:191–206. [https://doi.org/10.1016/S0301-0511\(01\)00080-1](https://doi.org/10.1016/S0301-0511(01)00080-1).
- De Bruijn ERA, Schubotz RI, Ullsperger M (2007) An event-related potential study on the observation of erroneous everyday actions. *Cogn Affect Behav Neurosci* 7(4):278–285. <https://doi.org/10.3758/cabn.7.4.278>.
- De Bruijn ERA, Miedl SF, Bekkering H (2011) How a co-actor's task affects monitoring of own errors: evidence from a social event-related potential study. *Exp Brain Res* 211(3-4):397–404. <https://doi.org/10.1007/s00221-011-2615-1>.
- De Bruijn ER, Von Rhein DT (2012) Is your error my concern? An event-related potential study on own and observed error detection in cooperation and competition. *Front Neurosci* 6:8. <https://doi.org/10.3389/fnins.2012.00008>.
- Di Gregorio F, Maier ME, 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* 172:427–436. <https://doi.org/10.1016/j.neuroimage.2018.01.081>.
- Endrass T, Klawohn J, Schuster F, Kathmann N (2008) Overactive performance monitoring in obsessive-compulsive disorder: ERP evidence from correct and erroneous reactions. *Neuropsychologia* 46(7):1877–1887. <https://doi.org/10.1016/j.neuropsychologia.2007.12.001>.
- Ernestus M, Cutler A (2015) BALDEY: A database of auditory lexical decisions. *Q J Exp Psychol* 68(8):1469–1488. <https://doi.org/10.1080/17470218.2014.984730>.
- 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. *Electroencephalogr Clin Neurophysiol* 78(6):447–455. [https://doi.org/10.1016/0013-4694\(91\)90062-9](https://doi.org/10.1016/0013-4694(91)90062-9).
- Friedrich CK, Eulitz C, Lahiri A (2006) Not every pseudoword disrupts word recognition: An ERP study. *Behav Brain Funct* 2:36. <https://doi.org/10.1186/1744-9081-2-36>.
- Gagnepain P, Henson R, Davis M (2012) Temporal predictive codes for spoken words in auditory cortex. *Curr Biol* 22(7):615–621. <https://doi.org/10.1016/j.cub.2012.02.015>.
- Ganushchak L, Christoffels I, Schiller NO (2011) The use of electroencephalography in language production research: a review. *Front Psychol* 2:208. <https://doi.org/10.3389/fpsyg.2011.00208>.
- Gehring WJ, Goss B, Coles MG, Meyer DE, Donchin E (1993) A neural system for error detection and compensation. *Psychol Sci* 4(6):385–390. <https://doi.org/10.1111/j.1467-9280.1993.tb00586.x>.
- Gibbons H, Fritzsche A-S, Bienert S, Armbrrecht A-S, Stahl J (2011) Percept-based and object-based error processing: An experimental dissociation of error-related negativity and error positivity. *Clin Neurophysiol* 122(2):299–310. <https://doi.org/10.1016/j.clinph.2010.06.031>.
- Goldinger SD (1996) Auditory Lexical decision. *Lang Cogn Process* 11(6):559–568. <https://doi.org/10.1080/016909696386944>.
- Holroyd CB, Coles MG (2002) The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol Rev* 109(4):679. <https://doi.org/10.1037/0033-295X.109.4.679>.
- Joch M, Hegele M, Maurer H, Müller H, Maurer LK (2017) Brain negativity as an indicator of predictive error processing: the contribution of visual action effect monitoring. *J Neurophysiol* 118(1):486–495. <https://doi.org/10.1152/jn.00036.2017>.
- Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ (2000) Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clin Neurophysiol* 111(10):1745–1758. [https://doi.org/10.1016/s1388-2457\(00\)00386-2](https://doi.org/10.1016/s1388-2457(00)00386-2).
- Kaczurkin AN (2013) The effect of manipulating task difficulty on error-related negativity in individuals with obsessive-compulsive symptoms. *Biol Psychol* 93(1):122–131. <https://doi.org/10.1016/j.biopsycho.2013.01.001>.
- Keuleers E, Brysbaert M, New B (2010) SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behav Res Methods* 42(3):643–650. <https://doi.org/10.3758/BRM.42.3.643>.
- Keyser C, Perrett DI (2004) Demystifying social cognition: a Hebbian perspective. *Trends Cogn Sci* 8(11):501–507. <https://doi.org/10.1016/j.tics.2004.09.005>.
- Kilner JM, Friston KJ, Frith CD (2007) Predictive coding: an account of the mirror neuron system. *Cogn Process* 8(3):159–166. <https://doi.org/10.1007/s10339-007-0170-2>.
- Koban L, Pourtois G, Bediou B, Vuilleumier P (2012) Effects of social context and predictive relevance on action outcome monitoring. *Cogn Affect Behav Neurosci* 12(3):460–478. <https://doi.org/10.3758/s13415-012-0091-0>.
- Koban L, Pourtois G (2014) Brain systems underlying the affective and social monitoring of actions: An integrative review. *Neurosci Biobehav Rev* 46:71–84. <https://doi.org/10.1016/j.neubiorev.2014.02.014>.
- Koelewijn T, van Schie HT, Bekkering H, Oostenveld R, Jensen O (2008) Motor-cortical beta oscillations are modulated by correctness of observed action. *Neuroimage* 40(2):767–775. <https://doi.org/10.1016/j.neuroimage.2007.12.018>.
- Kutas M, Federmeier KD (2011) Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu Rev Psychol* 62(1):621–647. <https://doi.org/10.1146/annurev.psych.093008.131123>.
- Kutas M, Van Petten C (1994) Psycholinguistics electrified: event-related brain potential investigations. In: Gernsbacher MA, editor. *Handbook of psycholinguistics*. San Diego, CA: Academic. p. 83–143.
- Luck SJ, Kappenman ES, editors. *The Oxford handbook of event-related potential components*. Oxford University Press.
- Luu P, Tucker DM, Makeig S (2004) Frontal midline theta and the error-related negativity: neurophysiological mechanisms of action regulation. *Clin Neurophysiol* 115(8):1821–1835. <https://doi.org/10.1016/j.clinph.2004.03.031>.
- Masaki H, Tanaka H, Takasawa N, Yamazaki K (2001) Error-related brain potentials elicited by vocal errors. *Neuroreport* 12(9):1851–1855. <https://doi.org/10.1097/00001756-200107030-00018>.
- MacGregor LJ, Pulvermüller F, Van Casteren M, Shtyrov Y (2012) Ultra-rapid access to words in the brain. *Nat Commun* 3(1):1–7. <https://doi.org/10.1038/ncomms1715>.
- Miltner WHR, Brauer J, Hecht H, Trippe R, Coles MGH (2004) Parallel brain activity for self-generated and observed errors. In: Ullsperger M, Falkenstein M, editors. *Errors, conflicts, and the brain: Current opinions on performance monitoring*. Leipzig, Germany: MPI of Human Cognitive and Brain Sciences. p. 124–129.
- Moors A, De Houwer J, Hermans D, Wanmaker S, van Schie K, Van Harmelen A-L, De Schryver M, De Winne J, Brysbaert M (2013) Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behav Res* 45(1):169–177. <https://doi.org/10.3758/s13428-012-0243-8>.
- Navarro-Cebrian A, Knight RT, Kayser AS (2016) Frontal monitoring and parietal evidence: mechanisms of error correction. *J Cogn Neurosci* 28(8):1166–1177. [https://doi.org/10.1162/jocn\\_a\\_00962](https://doi.org/10.1162/jocn_a_00962).
- Nieuwenhuis S, Ridderinkhof KR, Blom J, Band GP, 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.
- Norris D, McQueen JM (2008) Shortlist B: a Bayesian model of continuous speech recognition. *Psychol Rev* 115(2):357. <https://doi.org/10.1037/0033-295X.115.2.357>.
- Oostenveld R, Fries P, Maris E, Schoffelen J-M (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci* 2011:1–9. <https://doi.org/10.1155/2011/156869>.
- Olvet DM, Hajcak G (2009) The stability of error-related brain activity with increasing trials. *Psychophysiology* 46(5):957–961. <https://doi.org/10.1111/j.1469-8986.2009.00848.x>.

- Panasiti MS, Pavone EF, Aglioti SM (2016) Electrocortical signatures of detecting errors in the actions of others: An EEG study in pianists, non-pianist musicians and musically naïve people. *Neuroscience* 318:104–113. <https://doi.org/10.1016/j.neuroscience.2016.01.023>.
- Panasiti MS, Porciello G, Aglioti SM (2017) The bright and the dark sides of motor simulation. *Neuropsychologia* 105:92–100. <https://doi.org/10.1016/j.neuropsychologia.2017.05.020>.
- Pavone EF, Tieri G, Rizza G, Tidoni E, Grisoni L, Aglioti SM (2016) Embodying others in immersive virtual reality: electro-cortical signatures of monitoring the errors in the actions of an avatar seen from a first-person perspective. *J Neurosci* 36(2):268–279. <https://doi.org/10.1523/JNEUROSCI.0494-15.2016>.
- Pezzetta R, Nicolardi V, Tidoni E, Aglioti SM (2018) Error, rather than its probability, elicits specific electrocortical signatures: a combined EEG-immersive virtual reality study of action observation. *J Neurophysiol* 120(3):1107–1118. <https://doi.org/10.1152/jn.00130.2018>.
- Polich J (2007) Updating P300: an integrative theory of P3a and P3b. *Clin Neurophysiol* 118(10):2128–2148. <https://doi.org/10.1016/j.clinph.2007.04.019>.
- Pontifex MB, Scudder MR, Brown ML, O'Leary KC, Wu CT, Themanson JR, Hillman CH (2010) On the number of trials necessary for stabilization of error-related brain activity across the life span. *Psychophysiology* 47(4):767–773. <https://doi.org/10.1111/j.1469-8986.2010.00974.x>.
- Riesel A, Weinberg A, Endrass T, Meyer A, Hajcak G (2013) The ERN is the ERN is the ERN? Convergent validity of error-related brain activity across different tasks. *Biol Psychol* 93(3):377–385. <https://doi.org/10.1016/j.biopsycho.2013.04.007>.
- Riès S, Janssen N, Dufau S, Alario F-X, Burle B (2011) General-purpose monitoring during speech production. *J Cogn Neurosci* 23(6):1419–1436. <https://doi.org/10.1162/jocn.2010.21467>.
- Rizzolatti G, Craighero L (2004) The mirror-neuron system. *Annu Rev Neurosci* 27(1):169–192. <https://doi.org/10.1146/annurev.neuro.27.070203.144230>.
- Sebastian-Gallés N, Rodríguez-Fornells A, de Diego-Balaguer R, Díaz B (2006) First- and second-language phonological representations in the mental Lexicon. *J Cogn Neurosci* 18(8):1277–1291. <https://doi.org/10.1162/jocn.2006.18.8.1277>.
- Shalgi S, Barkan I, Deouell LY (2009) On the positive side of error processing: error-awareness positivity revisited. *Eur J Neurosci* 29(7):1522–1532. <https://doi.org/10.1111/j.1460-9568.2009.06690.x>.
- Ullsperger M, Danielmeier C, Jochem G (2014a) Neurophysiology of performance monitoring and adaptive behavior. *Physiol Rev* 94(1):35–79. <https://doi.org/10.1152/physrev.00041.2012>.
- Ullsperger M, Fischer AG, Nigbur R, Endrass T (2014b) Neural mechanisms and temporal dynamics of performance monitoring. *Trends Cogn Sci* 18(5):259–267. <https://doi.org/10.1016/j.tics.2014.02.009>.
- van Schie HT, Mars RB, Coles MGH, Bekkering H (2004) Modulation of activity in medial frontal and motor cortices during error observation. *Nat Neurosci* 7(5):549–554. <https://doi.org/10.1038/nn1239>.
- Volpe U, Mucci A, Bucci P, Merlotti E, Galderisi S, Maj M (2007) The cortical generators of P3a and P3b: a LORETA study. *Brain Res Bull* 73(4-6):220–230. <https://doi.org/10.1016/j.brainresbull.2007.03.003>.
- Wessel JR, Ullsperger M (2011) Selection of independent components representing event-related brain potentials: A data-driven approach for greater objectivity. *Neuroimage* 54(3):2105–2115. <https://doi.org/10.1016/j.neuroimage.2010.10.033>.
- Wessel JR (2012) Error awareness and the error-related negativity: evaluating the first decade of evidence. *Front Hum Neurosci* 6:88. <https://doi.org/10.3389/fnhum.2012.00088>.
- Wessel JR (2014) Performance monitoring in realistic environments: can translating neuroscientific insights augment real-world behavioral adaptation? *J Neurosci* 34(27):8934–8936. <https://doi.org/10.1523/JNEUROSCI.1840-14.2014>.
- Zambrano-Vazquez L, Allen JJB (2014) Differential contributions of worry, anxiety, and obsessive compulsive symptoms to ERN amplitudes in response monitoring and reinforcement learning tasks. *Neuropsychologia* 61:197–209. <https://doi.org/10.1016/j.neuropsychologia.2014.06.023>.

## APPENDIX A. SUPPLEMENTARY DATA

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.neuroscience.2021.02.001>.

(Received 21 August 2020, Accepted 1 February 2021)  
(Available online xxxx)