A robust temporal map of speech monitoring from planning to articulation.

Dorokhova Lydia [1], Morillon Benjamin[2], Baus Cristina[3], Belin Pascal[4], Dubarry Anne-Sophie[5], Alario François-Xavier[6], Runnqvist Elin[1]

[1] Aix-Marseille Université, CNRS, LPL, Laboratoire Parole et Langage, Aix-en-Provence, France, [2]Aix Marseille Université, Inserm, INS, Institut de Neurosciences des Systèmes, Marseille, France, [3] Department of Cognition, Development and Educational Psychology, University of Barcelona,Barcelona, Spain, [4]Aix-Marseille Université, CNRS, INT,Institut de Neurosciences de la Timone, Marseille, France, [5] Aix Marseille Univ, CNRS, LNC, Marseille, France, [6]Aix-Marseille Université, CNRS, LPC, Marseille, France

Abstract

Speakers continuously monitor their own speech to optimize fluent production. However, the precise timing and underlying variables influencing speech monitoring remain insufficiently understood. This study aimed to provide a comprehensive temporal map of monitoring processes ranging from speech planning to articulation. Two closely resembling experiments were conducted, focusing on effects that consistently emerged across both. Participants engaged in a speeded language production task designed to elicit speech errors of either a lexical or articulatory-phonetic origin, while their EEG activity was recorded. On correctly produced utterances, we explored error probability at different levels of processing (lexical and articulatory-phonetic) and we also compared errors with correct trials to capture the potential diversity of response conflict and monitoring processes. Concerning the effects of error probability on correct trials, differences driven by the lexical status of a competing response were observed during initial stages of speech planning, while differences related to articulatory phonetically driven response competition emerged during speech motor preparation. In contrast, errors showed differences with correct utterances in both early and late speech motor preparation and during articulation. Taken together, these findings suggest that (a) response conflict on ultimately correct trials does not persist during articulation ; (b) the timecourse of response conflict is restricted to the time window during which a given linguistic level is task relevant (early on for response appropriateness related variables and later for articulation relevant variables) ; and (c) monitoring during the response seems to be primarily triggered by pre-response monitoring failure.

Key words : error monitoring ; response-conflict ; language production ; EEG, MVPA

Introduction

Speaking is a complex process that engages both cognitive and motor components, encompassing semantic and lexical retrieval as well as articulatory programming and execution. Extensive research has provided evidence that both cognitive and motor aspects of speech are continuously monitored to optimize fluent production. For instance, naturally occurring and laboratory induced speech errors show patterns suggesting the anticipation of potential undesired outcomes during speech planning. In particular, contextually inappropriate responses such as taboo words or non-lexical speech errors occur below chance even in controlled error protocols (SEVERENS et al. 2011; HARTSUIKER, CORLEY et al. 2005). Monitoring is also evident in speakers' behavior in response to their own speech errors, including accurate self-reporting of errors in various environments (POSTMA et al. 1996; GAUVIN et al. 2016); post-error increases in response latencies (GANUSHCHAK et al. 2006); and self-repairs (LEVELT 1983) . It has been observed that certain speech error repairs occur too rapidly to be attributed to the interception and planning of corrections after the error is produced, suggesting that errors are intercepted before becoming overt (LEVELT 1983; HARTSUIKER et KOLK 2001). Furthermore, studies involving modulated speech feedback have demonstrated error monitoring during articulation, as participants adapt their speech production (pitch or formants) to compensate for perceived distortions in feedback (e.g. SAVARIAUX et al. 1995, NIZIOLEK et al. 2013). Somatosensory feedback has also been altered to the same effect (e.g., TREMBLAY et al. 2003). Overall, speech error patterns reveal the coexistence of both cognitive and motor dimensions in speech monitoring. Concerning the temporal dynamics of monitoring, it has been shown that error-to-cutoff times display a bimodal distribution, with an interruption of an erroneous segment occurring either shortly after the error or around 500 ms later (NOOTEBOOM et al. 2017). This implies the existence of at least two distinct time points during which monitoring processes occur or interact with the production process. Employing time-sensitive electroencephalograpic recordings (EEG), prior literature has revealed three relevant time points to observe EEG activity in speech-monitoring tasks : speech planning in it's initial stages, speech motor planning, and speech articulation. In the context of *initial stage speech planning*, several studies have examined the EEG signal following stimulus presentation and preceding motor response preparation. In a speech production task designed to prime errors, trials resulting in errors showed an increased negativity between 350 ms and 600 ms after the appearance of a written word pair to be pronounced aloud (MOLLER et al. 2006). Additionally, semantic response conflict on correct trials in a phoneme detection task elicited a negativity around 450 ms after stimulus presentation (GANUSHCHAK et al. 2008a). Concerning *speech motor planning*, previous EEG studies have analyzed the signal following the presentation of a speech-cue, immediately before the response.

In one study, within the 50-150 ms and 230-300 ms time windows erroneous trials exhibited more negative potentials compared to correct productions (MOLLER et al. 2006). In another study, correct trials primed to result in taboo word errors resulted in an increased negativity in the 550-625 ms time window compared to correct trials primed to result in neutral errors (e.g. SEVERENS et al. 2011). Lastly, concerning *speech articulation*, previous studies analyzed the EEG signal following a response (button press or verbal). An error-related negativity (ERN) was observed following a button press in a phoneme detection task, where false alarms generated a larger ERN compared to correct hits (GANUSHCHAK et al. 2008a). Furthermore, the magnitude of this ERN was modulated by the semantic relatedness of the auditory distractor, being greater for semantically related distractors. Additionally, during picture naming, incorrectly named pictures resulted in a larger ERN compared to correctly named pictures, with the ERN also being influenced by semantic naming context (GANUSHCHAK et al. 2008b; MASAKI et al. 2001; RIÈS, JANSSEN et al. 2011; BAUS et al. 2020).

Despite these valuable insights, a comprehensive understanding of the temporal dynamics of monitoring is hindered by the focus on specific variables and time frames of production in each study. The majority of previous studies have targeted monitoring through comparisons involving overt errors, lacking insight into how monitoring operates in contexts where errors are probable but ultimately avoided (but see SEVERENS et al. 2011). As such, it remains unclear whether monitoring occurs at multiple time points for overt errors only or also for correct trials where errors are likely. Additionally, while some previous research has explored the impact of meaning-related variables (e.g., semantic relatedness, taboo status) on speech monitoring, the influence of other linguistic variables susceptible of producing response conflict, such as lexical or articulatory-phonetic variables, remains underexplored, and to our knowledge no study has explored more than one linguistic variable in the same study. Thus, the extent to which the time course of monitoring is similar for all aspects of speech or varies based on the level of representation remains an open question. Some hints to the answer to these open questions can be obtained through the results of two recent fMRI studies that examined correct utterances produced in contexts of high lexically or articualtory phonetic driven error probability respectively, and that also examined erroneous as opposed to correct utterances (RUNNQVIST, CHANOINE et al. 2021 and TODOROVIĆ et al. 2023). It was observed that cerebellar structures (Crus I) related to predictive internal modeling were active for both monitoring of correct but error prone utterances and for overt errors compared to correct utterances across lexical and articulatory phonetic variables. In contrast, additional frontal and medial frontal structures were recruited for error prone utterances at the articulatory phonetic level and for overt errors, indicating that distinct mechanisms are at play in function of distance from articulation. Finally, across both studies, the anterior cingulate cortex was only differentially activated for the contrasts involving overt errors, supporting the involvment of a different feedback control related mechanism for overt errors. While the observed differential brain activity in these studies provides evidence supporting dissociations in monitoring mechanisms depending on whether errors are probable or overt and depending on linguistic variables (see also RUNNQVIST 2023; TEGHIPCO et al. 2023; OKADA et al. 2018; HANSEN et al. 2019a; HANSEN et al. 2019b; VOLFART et al. 2022), it does not offer information about the specific timing of when these mechanisms are engaged during speech production. Doing so was the objective of the current study.

To this end, we conducted two closely resembling experiments, focusing on robust effects that replicated across both. Participants engaged in a speeded language production task designed to elicit speech errors of either a lexical or articulatory phonetic origin, while their EEG activity was recorded. The EEG signal was segmented into three distinct epochs (stimulus-locked, speech-cue-locked, and response-locked, see Figure 1) allowing us to cover the entire speech production process as reflected both by externally triggered events (e.g., stimulus and speech cue appearance, see Figure 1) and internally initiated events (e.g., the response). This design allowed us to explore monitoring processes related to correct but error prone production at both the lexical and articulatory - phonetic levels, as well as to explore monitoring related to overt speech production errors. Finally, we performed both event-related potential (ERP) analyses and multivariate pattern (MVPA) analyses on the data. The reason behind this analysis approach was that MVPA can be more sensitive to subtle variations in neural activity. It can detect distributed patterns of neural activation that ERPs might miss, making it especially useful when studying cognitive processes that involve complex and distributed neural networks (e.g., KING et al. 2014).

Methods and Materials

The study received appropriate ethical approval (filed under id "RCB : 2011-A00562-39") at the regional ethical committee "Comité de Protection des Personnes Sud Méditerranée I".

Participants

Experiment I

Twenty-nine right-handed native speakers of French (22 women) with normal or corrected-to-normal vision took part in the experiment in exchange for monetary compensation (mean age = 21, min = 19, max = 23). No participant reported any history of language or neurological disorders. One participant was excluded from the analyses because she had participated in another study using the same task only a few weeks before the experimental session. Four participants were excluded from further analyses because of excessive noise or EEG data loss. Thus, 24 participants were included in the final analysis.

Experiment II

Fourty-four right-handed native speakers of French (39 women) with normal or corrected-to-normal vision took part in the experiment in exchange for monetary compensation (mean age = 23, min = 19, max = 31). No participant reported any history of language or neurological disorders. Thirteen participants were excluded from the analyses due to different issues : behavioral (5 participants had an error rate outside of the criteria : 1%<n<50%), problems during EEG recordings (8 participants, excessive noise or EEG data loss). Thus, 31 participants were included in the final analysis.

Stimuli

Experiment I

Target stimuli consisted of 160 printed French nouns (those used in RUNNQVIST, BONNARD et al. 2016) to be presented in pairs. For illustrative purposes, the examples in the text are given in English. To control for differences due to auditory stimulation, motor activity, or articulator specific modulations of the signal (e.g., SZIRTES et al. 1977), the same words were to be produced across conditions (albeit combined differently to prime **lexical** and **non-lexical errors**). Thus, across participants, each word was used twice in combination with another word (e.g., *mole sail*, *mole fence*). Exchanging the first letters of these combinations would result in a new word pair in one case (*sole mail*, lexical error outcome) and in a non-word pair in the other case (*fole mence*, non-lexical error outcome). An orthographic criterion was used for selecting stimuli, but even when applying a phonological criterion post-hoc only 7/160 non-lexical items resulted in real words for one of the words in a pair when changing orthography (which sometimes also entailed a change in wordclass , e.g., for the pair *caverne bouton* the primed noun error *couton* does not exist but *coûtons* is a conjugated form of the verb *couter*). All combinations were used in both possible orders across participants (e.g., *mole sail* and *sail mole*). Further, all combinations for which the exchange of initial phonemes resulted in new word-pairs (*mole sail*) were used also in reversed order (*sole mail*). The words in the target pairs were selected with the criterion that they should be unrelated. Despite this effort, for 9/320 word pairs (4 lexical and 5 non-lexical) there was some form of semantic relationship between the two words (e.g., *sick sinus. chicken tavern*). A given participant was only presented with one combination for each word (lexical or non-lexical outcome), and was only presented with one of the words differing in only the first sound (*mole* or *sole*). This resulted in the creation of eight experimental lists with 80 word pairs (40 lexical and 40 non-lexical error outcome) counterbalanced across participants. Finally, all word pairs were coded for the degree of shared phonetic features (place, manner of articulation and voicing) of initial consonants of words in a pair being assigned a number from 0 (**AP distant words**, e.g., [m] is labial, nasal and voiced and [s] is dental, fricative and voiceless) to 2 (**AP close words**, e.g., [m] and [b] both being labial and voiced). Of the 80 word pairs, on average across the 8 lists 25.5 word pairs did not share any features, 40.25 word pairs shared one feature and 14.25 shared 2 features. As a first step, to control for a possible confound between our lexical and articulatory phonetic variables, we controlled whether the stimuli across the lexical and non-lexical conditions differed in the average amount of shared features (SF) and this was not the case (lexical 0.9 shared features vs. non-lexical 0.8 shared features).The two-tailed independent samples t-test comparing average SF values between the lexical and non-lexical conditions ($n = 8$ for both groups) yielded a non-significant result ($p = 0.11$). As a subsequent step, we introduced Articulatory-Phonetic Proximity (AP) as a three-level factor (No SF, 1 SF and 2 SF) during the analysis of errors and reaction times. This allowed us to investigate potential interactions between this factor and the lexical status, as detailed in the Analysis (Section) and Results (Section) sections. More importantly for the current purposes, coding the phonetic proximity between our word pairs also allowed us to assess the impact of this articulatory - phonetic variable known to modulate speech error rates (e.g., NOOTEBOOM et al. 2008; OPPENHEIM et al. 2008) on participants' electrophysiological recordings.

During the experiment, three priming word pairs preceded each target word pair. The first two shared the initial consonants, and the third pair had further phonological overlap with the error being primed (*sun mall – sand mouth – soap mate – mole sail*). Note that, to induce errors, the order of the two initial consonants (/s/ and /m/) is different for the primes and the target. Participants were also presented with 153 filler pairs that had no specific relationship to their corresponding target pairs. One to three such filler pairs were presented to participants before each prime and target sequence. Thus, each participant was presented with 473 unique word combinations (80 targets, 240 primes and 153 fillers). Each list contained three blocks in which these 473 words were repeated three times in different orders. Participants were instructed to read all target word pairs aloud, all prime pairs silently, 35% of the filler pairs aloud and 65% of the filler pairs silently. Prior to the commencement of the experimental trials, participants underwent a task familiarization phase. This entailed exposure to a concise task sample, comprising 10 instances, under the direct supervision of the experimenter.

Experiment II

240 French monosyllabic (120) and bisyllabic (120) nouns were presented in pairs. Pairs were constant; there was no cross-combination of words as in Experiment I. Just as in Experiment I, exchanging the initial consonants of words in a pair gave a **lexical** or **non-lexical** outcome. We applied a phonological criterion for selecting the stimuli. In this stimuli set we manipulated the number of shared phonetic features of the onsets of words in pairs : half of the stimuli were **AP close** (2 shared feature among 3 possible : place, manner and voicing) and half were **AP distant** (no shared features), AP close and AP distant pairs were distributed equally across the lexical status conditions. Thus, here were no intermediate values (1 feature in common) as in Experiment I, because we aimed to maximize the effect by using the extremes. Words in pairs were always presented in the same order. Thus, the list was composed of 120 words pairs, where 60 were bisyllabic, 60 - monosyllabic, each syllabic condition contained 30 lexical and 30 non-lexical outcome pairs, and each lexical condition contained 15 AP close and 15 AP distant pairs. As in Experiment I, each target pair was preceded by 3 primes with the inverted order of onsets of words in pairs. Participants also saw 234 filler pairs, as in Experiment I, 1 to 3 of such pairs were presented before each sequence of primes and targets. The list was repeated 2 times with different order of sequences and primes. Participants were instructed to read all target word pairs aloud, all prime pairs silently, 49% of the filler pairs aloud and 51% of the filler pairs silently. Participants underwent pre-experiment task familiarization, involving a supervised exposure to a 10-instance task sample.

Procedure

Experiment I & Experiment II

The experiment was controlled by Eprime 2.0 software (Schneider et al. 2002). Each word pair remained on the screen for 700 ms and words presented for silent reading were followed by a blank screen for 200 ms. All targets and 35% of the filler items in the Experiment I and 50% of the filler items in Experiment II were followed by (a) a question mark remaining on the screen for 500 ms. (b) an exclamation mark presented 500 ms after the presentation of the question mark and remaining on the screen for 1000 ms, (c) a blank screen for 500 ms before the next trial started (see Schematic representaton of the task in Figure 1). Participants wore a microphone attached to the head in Experiment I, the microphone was placed on the table in front of them in Experiment II. They were instructed to silently read the word pairs as they appeared, but to name aloud the last word pair they had seen whenever a question mark was presented, and before the appearance of an exclamation mark. Productions were recorded both through E-prime and the software Audacity®to be processed off-line.

Electrophysiological Recordings

Experiment I & Experiment II

The EEG was recorded from 64 Ag/AgCl Active-Two pre- amplified electrodes (BIOSEMI, Amsterdam; 10–20 system positions). The sampling rate was 1024 Hz for Experiment I (online filters : DC to 208 Hz, 3 db/octave) and 2048 for Experiment II. Two additional electrodes placed close to Cz, the Common Mode Sense (CMS) active electrode and the Driven Right Leg (DRL) passive electrode, were used to form a feedback loop that maintains the average potential of the participant as close as possible to the AD-box reference potential. Two additional electrodes placed over the left and right mastoid were used to re-reference the signals (average mastoids reference). The vertical EOG was obtained by subtracting the signal of C29 (corresponding to FP2) from the signal of an external electrode placed underneath the left eye. The horizontal EOG was recorded with two external electrodes positioned over the two outer canthi.

Analyses

Behavior

Experiment I & Experiment II

Annotation. A person naive to the purpose of the experiment transcribed all spoken productions, then inspected and coded vocal response onsets of all individual recordings using Check-vocal (PROTOPAPAS 2007). Check-vocal is a software that allows for semi-automatic codification of the response accuracy and timing based on two sources of information : the speech waveform and the spectrogram. The transcriptions were scored as correct, disfluencies, partial responses (e.g., only one word produced), full omissions, and erroneous productions. The latter were classified as **priming related errors** or **other errors**. **Priming related errors** included full exchanges (*mill pad* ⇒ *pill mad*), partial exchanges (anticipations, e.g., *mill pad* ⇒ *pill pad*, perseverations, e.g., *mill pad* ⇒ *mill mad*, other partial exchanges, e.g., *mill pad* ⇒ *mill pack*). repaired and interrupted exchanges (*mill pad* ⇒ *pi..mill pad*), full and partial competing errors (*mill pad* ⇒ *pant milk/pant pad*), and other related errors (*mill pad* ⇒ *mad pill*), **Other errors**' included diverse phonological substitutions that were unrelated to the priming manipulation (e.g., *mill pad* ⇒ *chill pant/gri..mill pad/..pant*).

Experiment I

Data overview. Data of 24 participants initially presented 5760 trials, where each of the 80 pairs was repeated three times, resulting in 240 trials per participant. The lexicality condition was equally distributed in halves of the total number of trials, while the Articulatory-Phonetic (AP) condition introduced three levels (as detailed in Stimuli, Section). This configuration yielded 1834 pairs with no shared features, 2898 pairs with one shared feature, and 1028 pairs with two shared features between the initial consonants of the word pairs.

To ensure data quality, an initial filtration step excluded trials featuring full omissions, leading to the removal of 327 trials (5.68%). Subsequently, instances with RTs less than 100 ms or exceeding 1000 ms were identified as outliers and eliminated, accounting for 99 trials (1.72%). The final data set consisted of 5334 trials, distributed across conditions as follows : lexical (2663 trials), non-lexical (2671 trials), 0 shared features (1701 trials), 1 shared feature (2687 trials), and 2 shared features (946 trials).

Prior to statistical analysis, orthogonal contrasts were implemented for the AP condition using Helmert coding via the R built-in function `contr.helmert`. For the lexicality factor and priming_related_errors factor, sum coding (`contr.sum`) was applied (CHAMBERS et al. 1990 through R documentation).

Experiment II

Data overview. Data of 31 participants initially presented 7440 trials, wherein each of the 120 pairs was repeated twice, resulting in 240 trials per participant. The lexicality and AP conditions were evenly distributed in halves of the total trial count (3720 per condition).

Following the same data filtering process as in Experiment I, the exclusion of full omissions led to a reduction in the number of trials to 7344 (1.29% excluded). The removal of RT outliers (those falling outside the range of 100 ms < RT < 1000 ms) accounted for 144 trials (1.44%). The resultant dataset comprised 7200 trials, distributed across conditions as follows : lexical (3579 trials), non-lexical (3621 trials), AP close (3588 trials), and AP distant (3621 trials).

Given the balanced nature of all conditions, sum contrasts were employed for all factors using sum coding (`contr.sum`) (CHAMBERS et al. 1990 via R documentation) prior to conducting the statistical analysis in R CORE TEAM 2022.

Experiment I & Experiment II

Statistical models. The data analysis was conducted using the RStudio R CORE TEAM 2022 and key analytical tasks were performed utilizing specific packages including `'Matrix'` (version 1.5-1, BATES et al. 2021), `'lme4'` (version 1.1-34, BATES et al.

2015), 'lmerTest' (version 3.1-3, KUZNETSOVA et al. 2021), 'multcomp' (version 1.4-25, HOTHORN et al. 2021), dplyr (version 1.1.2, WICKHAM et al. 2021), ggplot2 (version 3.4.3, WICKHAM 2016), each of which facilitated critical statistical and visualization procedures. For the analysis of errors, we employed a mixed-effects logistic regression model using the glmer function of lme4 package in R CORE TEAM 2022. The initially proposed model :

```
glmer(errors ~ lexicality + AP + lexicality * AP +
      (1 +Lexicality| Subject)  (1 +AP| Subject) + (1 | WordPair),
      family = binomial)
```

involved two primary predictor variables, specifically Lexicality and AP, along with their interaction. Furthermore, the model accounted for a diverse set of random effects, encompassing random slopes and intercepts for both Lexicality and AP based on each individual subject in conjunction with a random intercept for the Word pair. However, due to convergence issues encountered during the modeling process, the complexity of the initial model needed to be streamlined. As a result, the model, shown below, aimed to scrutinize the effects of lexicality, AP condition, and their interaction on priming-related errors :

```
glmer(errors ~ lexicality + AP + lexicality * AP +
      (1 | Subject) + (1 | WordPair), family = binomial)
```

This model featured fixed effects for lexicality and AP condition as well as their interaction, while random intercepts for subjects and Word Pairs captured both individual differences and item-specific effects. Additionally, two separate lmer functions were employed. The first aimed to uncover Reaction Time (RT) discrepancies between overt errors and correct productions :

```
lmer(RT ~ priming_related_errors +
     (1 + priming_related_errors | Subject) + (1 | WordPair))
```

This model included random intercepts for both subjects and items, along with a random slope for priming_related_errors within subjects, addressing subject-specific variations in how these errors influenced Reaction Time.

The second lmer model was exclusively applied to correct trials and focused on examining the interplay between Reaction Time (RT) and the variables of Lexicality and Articulatory-Phonetic (AP), along with their interaction.

```
lmer(RT ~ lexicality + AP + lexicality * AP +
     (1 | Subject) + (1 | WordPair))
```

This model integrated random intercepts for subjects and items to accommodate individual differences and item-specific effects.The final model is an outcome of simplifying a more complex model that originally included random slopes and intercepts for both `Lexicality` and `AP` with respect to each individual subject. The decision to simplify was prompted by issues with convergence that were encountered during the modeling process.

EEG signal processing

Experiment I & Experiment II

Preprocessing. The EEG data was processed using the EEGLAB toolbox (DELORME et al. 2004) in MATLAB INC. 2020. Continuous EEG data were filtered offline through a 0.1 Hz to 30 Hz band-pass filter. Activity from the left and right mastoid electrodes were used off-line to re-reference scalp recordings. For the analyses we defined three epochs of interest : ERPs were either (1) time locked to the stimulus and segmented into 800 ms epochs (-100 to 700 ms), (2) locked to the speech cue into 500 ms epochs (-100 to 400 ms) or (3) locked to the response into 1500 ms epochs (-1000 to 500 ms) (Figure 1). Only segments without artifacts (activity ± 75 μV) were included. The epochs were then averaged and referenced to a 100 ms pre-stimulus, pre-speech-cue and pre-response baseline, respectively.

ERP analyses. As a next step we conducted a peak search within the epochs using ERP lab LOPEZ-CALDERON et al. 2014. The different conditions (overt (priming related) errors/correct; AP close/distant and lexical/non-lexical outcome) were averaged prior to this analysis, and we considered only the response peaks that were observed in both Experiments for subsequent analyses. For this, grand average waves of each epoch of both Experiments were inspected visually for the appearance of positive or negative peaks. Then the approximate time interval was given to the peak latency research function in ERP Measurement tool to obtain the exact peak latency value. These values were compared through two Experiments : when the difference in peak latency was less than 15 ms in between two Experiments, the mean value between the two peaks was used, when the difference was bigger, the peaks were not retained for further analyses. Subsequently, these peaks were used as centers of 100 ms time-windows. When it was impossible to use 100 ms time-windows, the largest possible symmetrical window was defined. The sum up of all the considered time-windows is reported in Table 1.

Statistical models. We utilized the same software and packages mentioned in the **??** Behavior "Statistical models" section to analyze EEG signal data. Each window of each Experiment was analyzed with Linear Mixed-Effects Models on 9 fronto-central

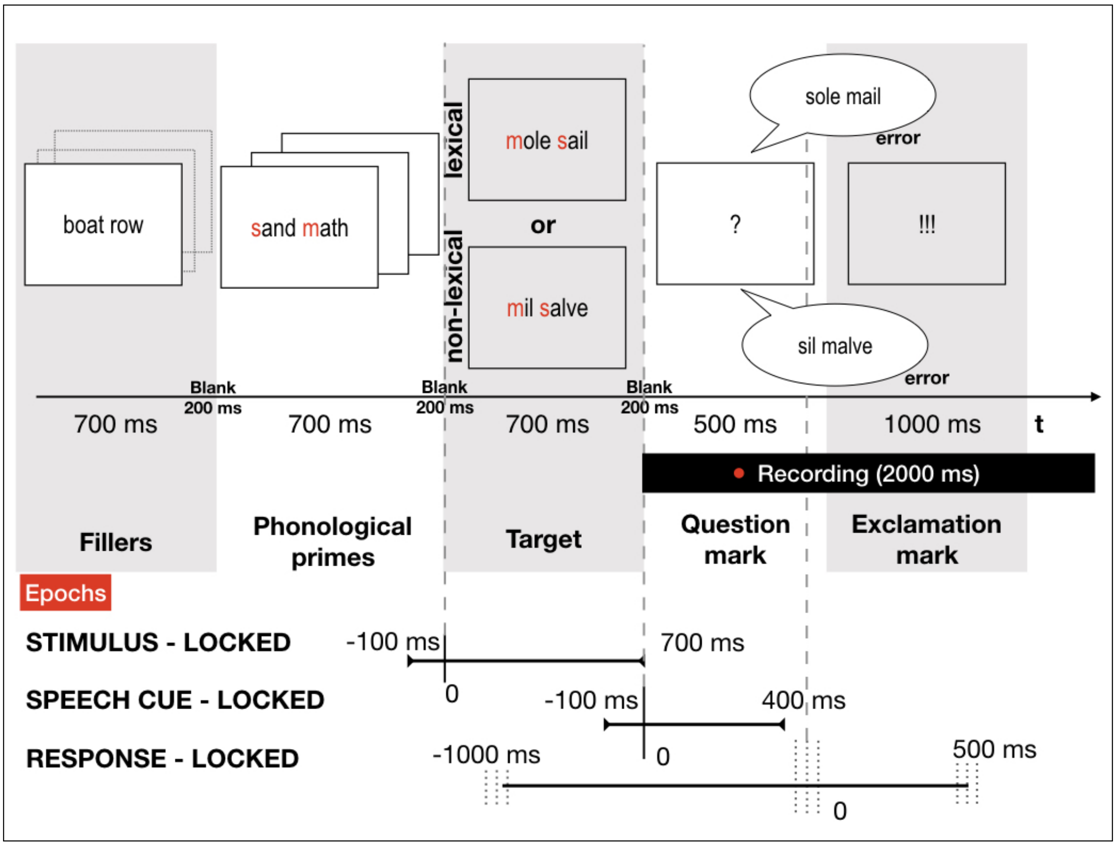

FIGURE 1 – Schematic representation of the procedure and epoching of EEG signal.

electrodes of interest (FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, CP2) same as in GRISONI, MOHR et al. 2019 and on all electrodes (57 after excluding frontal electrodes F7, AF7, Fp1, Fpz, Fp2, and AF8). Separate regressions were applied to each of three conditions : (1) overt errors vs. correct :

```
lmer (Mean amplitude ~ priming_related_errors*Electrode +(1|subject)
```

(2) lexical vs. non-lexical primed error outcome on correct trials :

```
lmer (Mean amplitude ~ lexicality*Electrode +(1|subject)
```

and (3) Articulatory phonetic proximity (AP) (close (2 shared features) vs. distant (no shared features) on correct trials) -only the conditions of 0 and 2 shared features were contrasted in the analysis of AP proximity to ensure comparability across the two experiments :

```
lmer (Mean amplitude ~ AP*Electrode +(1|subject)
```

In summary, each time-window was analysed with three separate models to investigate the effects of errors, lexicality and AP proximity respectively (see below). Even though we conducted two separate experiments and focused on the cross-validated findings, we decided to apply a Bonferroni correction to handle the issue of multiple comparisons. Summarized p-values for both corrected and uncorrected regressions are available in Appendix , specifically in Table S.2 for 9 fronto-central electrodes of interest, and Table S.3 for all 57 electrodes.

MVPA. Multivariate Pattern Analysis (MVPA) was also conducted on both Experiments on the same time-windows as in the ERP analysis, with the sklearn software (PEDREGOSA et al. 2011). We fitted segmented data into a 2D space-time Riemannian manifold to then run a logistic regression to classify across trials each of our three binary variables : we performed binary classification of (1) overt errors vs. correct, (2) lexical vs. non-lexical status and (3) AP close vs. AP distant. We performed nested cross-validations (5-fold) to optimize the regularization strength while preventing overfitting, and to explore generalizability. Splitting of the data was performed using a stratified folding approach, to prevent models from biasing toward the most numerous class. The performance of the selected model was calculated with area under the receiver operating characteristic curve (ROC_AUC). The outer loop of the nested cross-validation was carried out 10 times and averaged per subject. This analysis was repeated for each of the three contrasts and on each time-window. Analyses were performed at the single-subject level and followed by standard parametric one-tail paired t-tests at the group level (distribution of ROC_AUC values across subject compared to chance level (50%)).

Results

Behavior

Experiment I

Errors. Participants made errors in 230 trials (3.99% of all data). After the filtering described in , this number was reduced to 222 (4.16% of filtered data) of which 124 errors were priming related (2.32% of filtered data). More priming related errors were made in the lexical condition (114, 2.14%) than in the non-lexical one (10, 0.18%). The dispatch of errors in articulatory – phonetic proximity groups was as follows : 26 errors out of 1701 trials without shared features AP (1.53%), 75 errors out of 2687 trials with 1 SF (2.79%), and 23 errors out of 946 trials with 2 SF (2.43%). The effect of lexicality was significant ($z = 6.31; p<.001$) on priming-related errors according to the Generalized Linear Mixed-Effects Model (see Analyses for description). There was no effect of the articulatory-phonetic condition (1SF : $z = .02; p = .9$; 2SF : $z = .49; p = .62$) and no interaction with the lexicality condition (1SF : $z = 1.45, p = .14$; 2SF : $z = .18, p = .85$).

RT. Participants were slower in overall error trials (mean $RT_{222} = 612$ ms) that included priming related errors (mean $RT_{124} = 570$ ms)) than in correct trials (mean $RT_{5112} =$

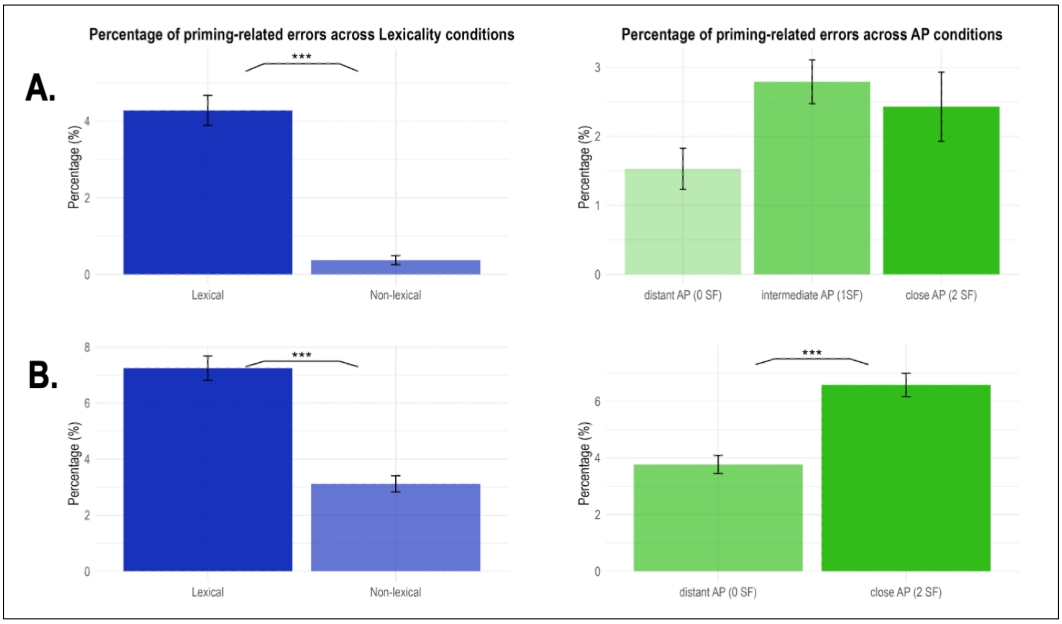

FIGURE 2 – Percentages of priming related errors by Lexicality and AP conditions. Panel A for Experiment I, panel B for Experiment II. Each portion is displayed with respect to the number of trials within the condition (i.e., the potential maximum number of errors).

515 ms). The effect of priming related errors on the RT was significant ($df = 20.59$, $t = -3.03$, p<.01). On correct trials, no significant difference in RT was observed between lexical (mean $RT_{2489} = 515$ ms and non-lexical condition (mean $RT_{2623} = 515$ ms), ($df = 598.8$, $t = .41$, $p = .68$). Participants were gradually slower through the AP condition ranging from mean $RT_{1640} = 508$ ms for 0 SF, $RT_{2568} = 518$ ms for 1 SF to $RT_{904} = 522$ ms for 2 SF. The AP effect was significant (1SF : $df = 580.48 t = 2.73, p = .006$; 2SF : $df = 586.79, t = 2.3, p = .02$), but no interaction with lexicality was observed (1SF : $df = 587.19 t = -1.35, p = .17$; 2SF : $df = 606.73, t = .03, p = .97$).

Experiment II

Errors. Participants made errors in (1100) trials (14.78% of all data), after the filtering described in , this number was reduced to 912 with 372 of priming related errors (5.16% of filtered data). More errors were committed in the lexical condition (259, 3.59%) compared to the non-lexical(113, 1.56%) and in the close articulatory-phonetic condition (236, 3.27%) compared to the distant articulatory-phonetic condition (136, 1.88%). The effects of lexicality and articulatory-phonetic condition were significant (z = 4.12, p<.001 ; z=-3.35, p<.001 respectively) on priming-related errors without interaction of the variables (z=1.19, p=.23).

RT. Participants were slower in overall error trials (mean $RT_{912} = 628$ ms) that included priming related errors (mean $RT_{372} = 633$ ms)) than in correct trials (mean $RT_{6278} = 531$ ms). The effect of priming related errors on the RT was significant ($df = 31.99$, $t =$

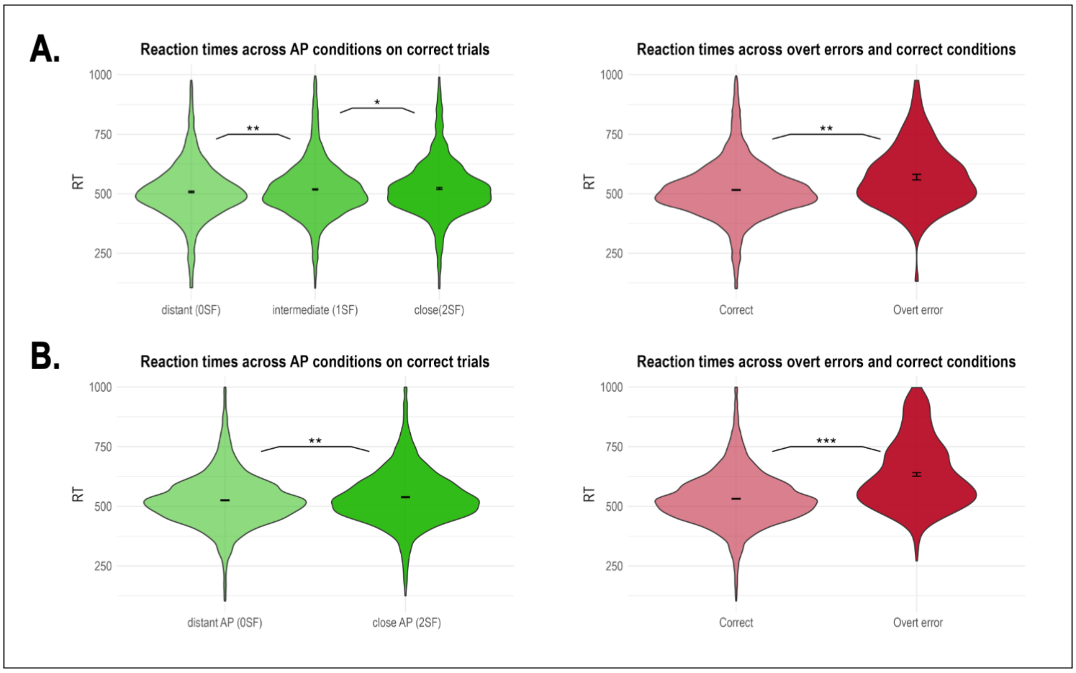

FIGURE 3 – Reaction Time (RT) Distribution by Articulatory phonetic (AP) Condition in Correct Trials (Left) and RT Distribution for Correct Responses and Overt Priming-Related Errors (Right). Panel A corresponds to Experiment I, while Panel B corresponds to Experiment II.

-8.6, p<.001). On correct trials, no significant difference in RT was observed between lexical (mean $RT_{3057} = 532$ ms and non-lexical condition (mean $RT_{3221} = 530$ ms), ($df = 113.53$, $t = .97$, $p = .33$). Participants were slower in the close AP condition (mean $RT_{3028} = 538$ ms) than in the distant AP condition ($RT_{3250} = 525$ ms). The AP effect was significant ($df = 113.51 t = -3.1, p < .01$), but no interaction with lexicality was observed (1SF : $df = 113.52 t = -1, p = .28$).

EEG signal results

At the neural level, we investigated significant effects of lexicality (lexical vs. non-lexical error priming), phonetic articulatory (AP close vs. distant onsets) and error (overt errors vs. correct). We investigated them with both a univariate (ERP) and a multivariate (MVPA) method. Each analysis was performed on the two experiments, the three distinct types of epochs (stimulus-locked, speech-cue-locked, and response-locked) and either all electrodes or a ROI analysis including 9 fronto-central electrodes of interest (see Methods; GRISONI, MOHR et al. 2019). Below we especially focus on the effects that were significant across both experiments, but the effects that did not replicate through Experiments are visualized and marked with asterisk in Figure 4 for response-locked epochs, in Figure 5 for speech cue-locked epochs.

ERP results. Three time windows in the response-locked epoch elicited significant differences for overt errors vs. correct condition in both datasets in the ROI analysis of 9 fronto-central electrodes. The first two time windows occurred before production onset. Firstly, the waveform of correct trials differed significantly from errors in the time window between [-483 to -383 ms] (Figure 6). This effect was followed by a significant difference between errors and correct trials during the pre-response positive drop [-115 ms, -15 ms] (Figure 4), ([-483 ms, -383 ms] and [-115 ms, -15 ms]). A third significant difference between correct trials and errors occurred after the onset of articulation ([62 ms, 162 ms]) (see Figure 6, Figure 4). For the stimulus-locked epochs, no significant effects were observed consistently across both experiments. In the 'all electrodes' analysis, the initial window of the speech cue-locked epoch [164 to 264 ms] exhibited a significant effect of overt errors, albeit not reaching significance after applying Bonferroni correction. The ERP analysis did not reveal any significant effect -cross-validated across Experiments- of the Articulatory-Phonetic (AP) and Lexicality effects. Supplementary tables of p-values can be found in the Appendix .

MVPA results. Both datasets showed significant decoding for the first time window (164 − 264 ms) of the speech cue-locked epoch for AP close vs. distant (Experiment 1 : t =3.10 ; [ROC_AUC] =0.53 ; p<0.01 ; Experiment 2 : t =4.70 ; [ROC_AUC] =0.57 ; p<0.01) and overt errors vs. correct (Experiment 1 : t =2.42 ; [ROC_AUC] =0.52 ; p=0.01 ; Experiment 2 : t =3.29 ; [ROC_AUC] =0.56 ; p<0.01). Furthermore, both datasets showed significant decoding for the response-locked epochs : the first window (-793, -693 ms) revealed a significant lexicality effect (Experiment 1 : t =2.85 ; [ROC_AUC] =0.53 ; p<0.01 ; Experiment 2 : t =5.7 ; [ROC_AUC] =0.61 ; p<0.01) and the second window (-483 ; -383 ms) revealed a significant overt errors vs. correct effect (Experiment 1 : t =3.52 ; [ROC_AUC] =0.53 ; p<0.01 ; Experiment 2 : t = 4.35 ; [ROC_AUC] =0.59 ; p<0.01). No significant effects consistent across both experiments were observed for the stimulus-locked epochs. Additional significant decoding results, not cross-validated between the two experiments, are depicted in Figures 4, 5, and also in Table S.4 and S.9 in the Appendix .

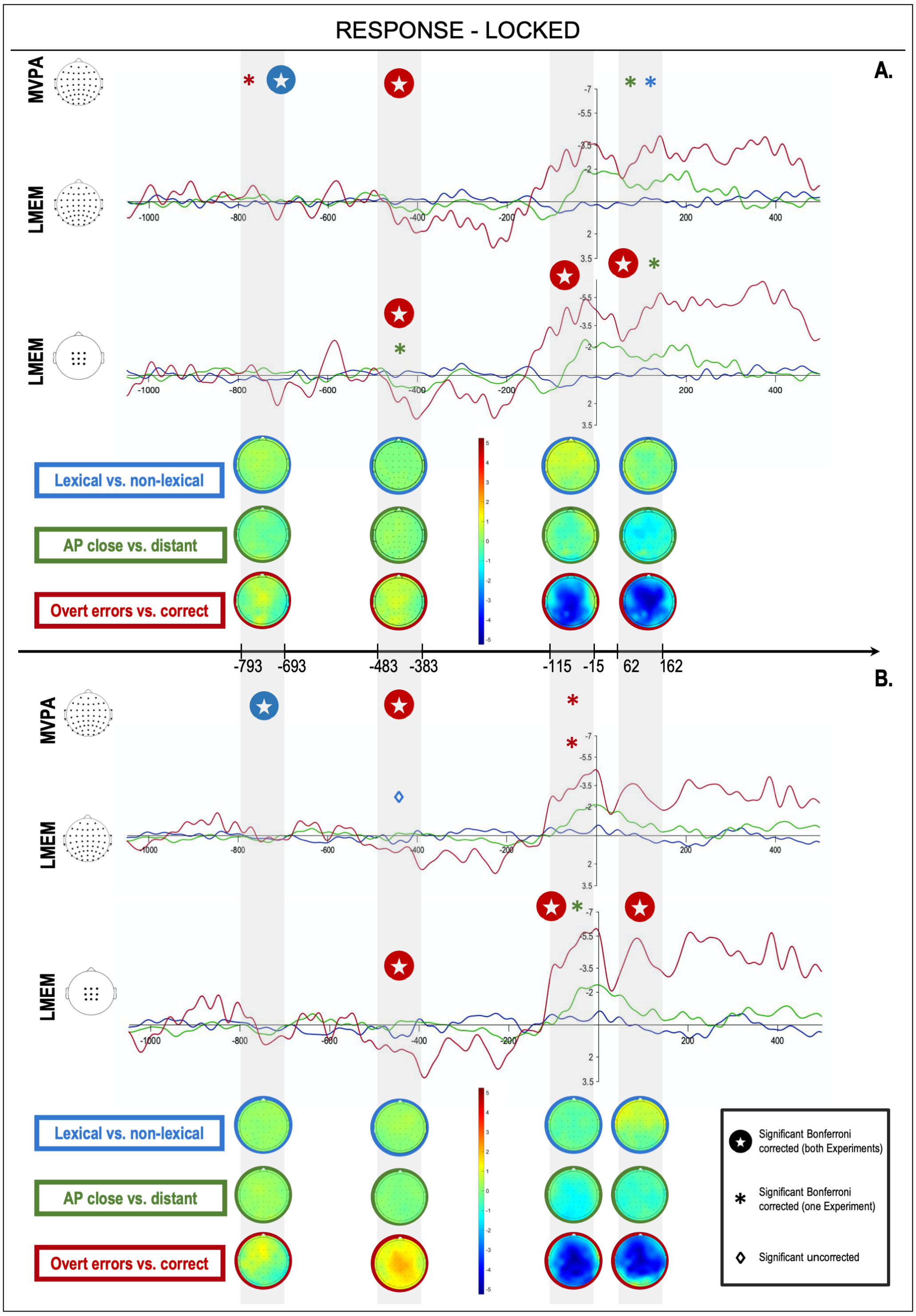

FIGURE 4 – Summary of results of all performed analyses : (top) MVPA, (middle) mean difference waves of all electrodes and (bottom) of the 9 fronto-central electrodes and their topographic maps across Experiment I (panel A.) and Experiment II (panel B.) in response-locked epochs. Color code is used to differentiate the conditions : blue for lexicality, green for articulatory – phonetic proximity and red for overt error vs. correct contrast. Asterisks mark significant bonferroni corrected p-values if observed in one Experiment, empty diamonds mark significant p-values without correction while bold stars mark significant p-values if observed in both Experiments.

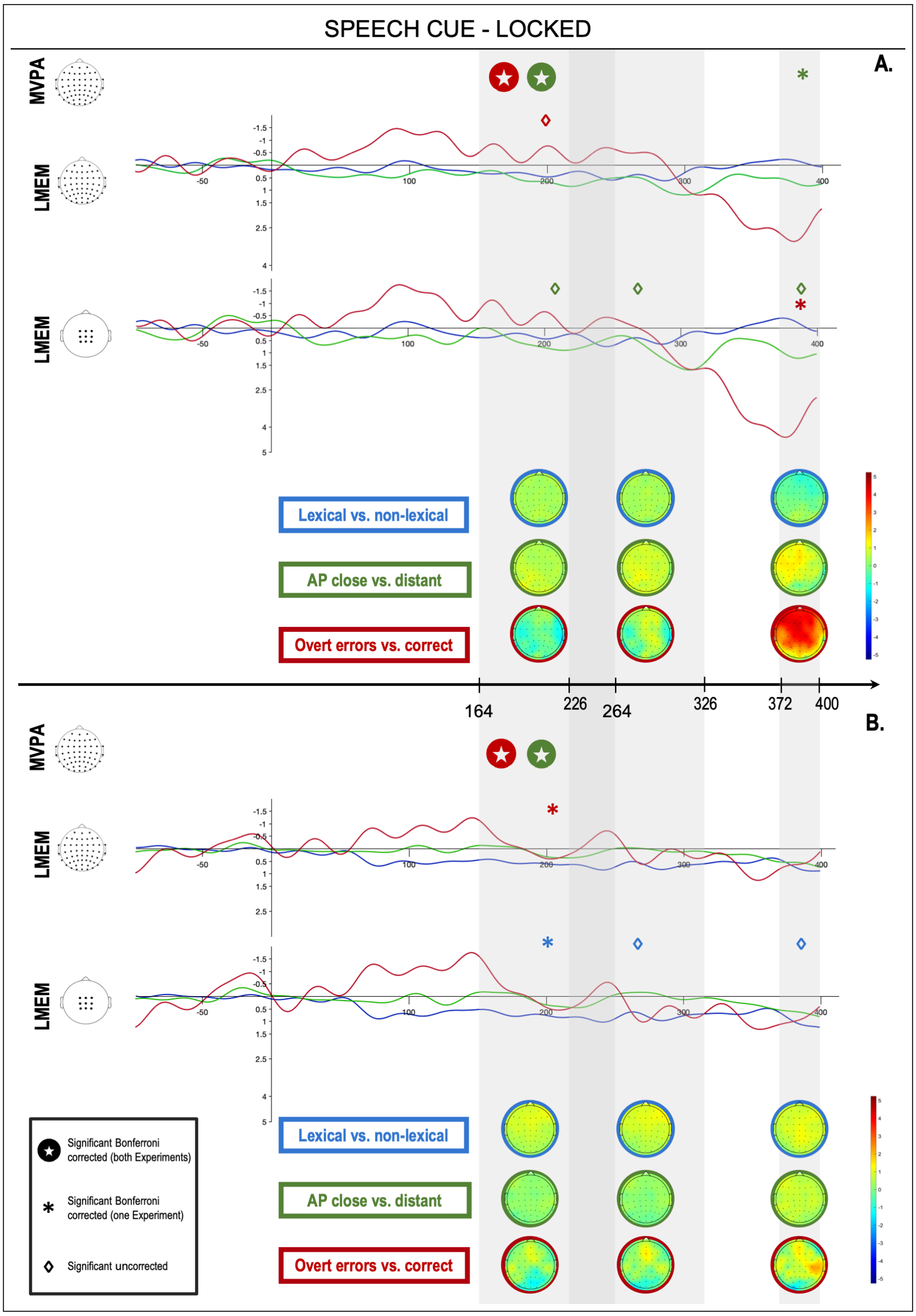

FIGURE 5 – Summary of results of all performed analyses : (top) MVPA, (middle) mean difference waves of all electrodes and (bottom) of the 9 fronto-central electrodes and their topographic maps across Experiment I (panel A.) and Experiment II (panel B.) in speech cue-locked epochs. Color code is used to differentiate the conditions : blue for lexicality, green for phonetic articulatory – phonetic proximity and red for overt error vs. correct contrast. Asterisks mark significant bonferroni corrected p-values if observed in one Experiment, empty diamonds mark significant p-values without correction while bold stars mark significant p-values if observed in both Experiments.

		latency (ms)		difference (ms)	window		
		Exp.1	Exp.2		mean	start	end
Stimulus - locked	negative	93	97	-4	95	45	145
	positive	204	213	-9	209	159	259
	negative	249	270	-21			
	negative	434	445	-11	439	389	489
Speech-cue - locked	negative	28	59	-31			
	positive	212	216	-4	214	164	264
	negative	269	272	-3	270	220	320
	positive	386	386	0	386	336	436
Response - locked	negative	-737	-740	3	-739	-789	-689
	positive	-667	-628	-39			
	negative	-437	-430	-7	-433	-483	-383
	negative	-56	-68	12	-62	-112	-12
	negative		51				
	positive		69				
	negative		119				
	positive		201				

TABLE 1 – Summary table of common peaks found in 2 Experiments

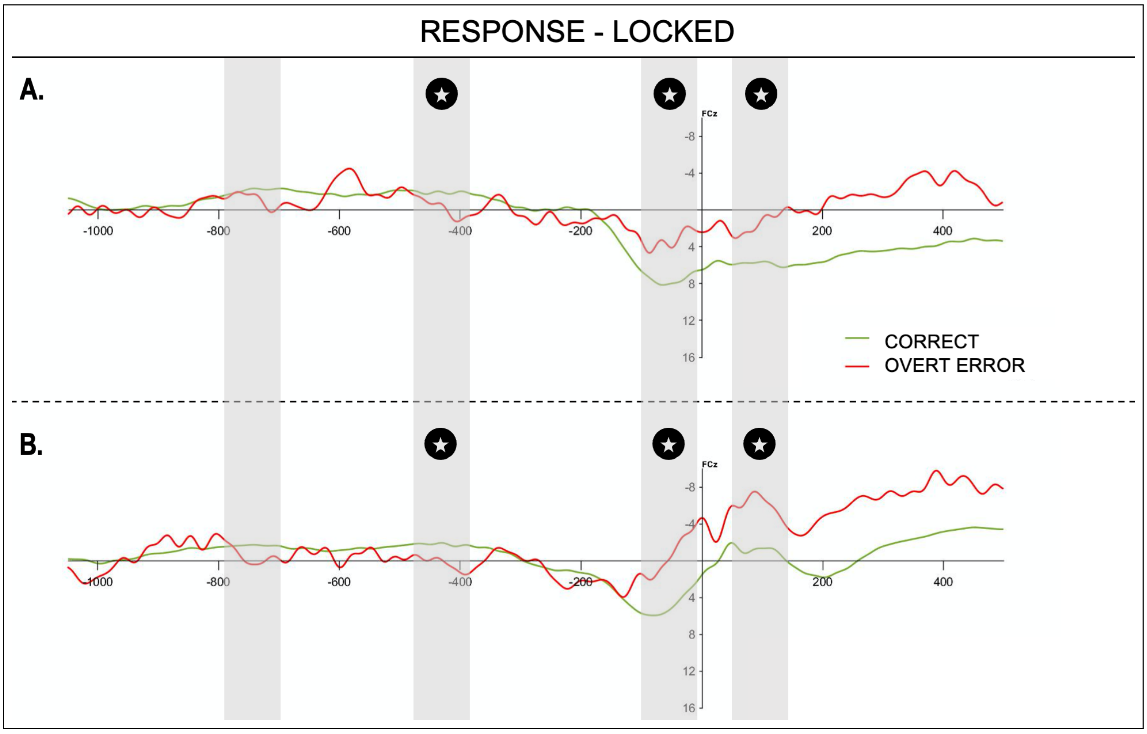

FIGURE 6 – Grand average wave of FCz electrode for correct (green) and overt error (red) trials in Experiment I (panel A.) and Experiment II (panel B.) in response-locked epochs. Stars indicate significant effect in GLM on 9 fronto-central electrodes (FC1 FCz FC2 C1 Cz C2 CP1 CPz CP2).

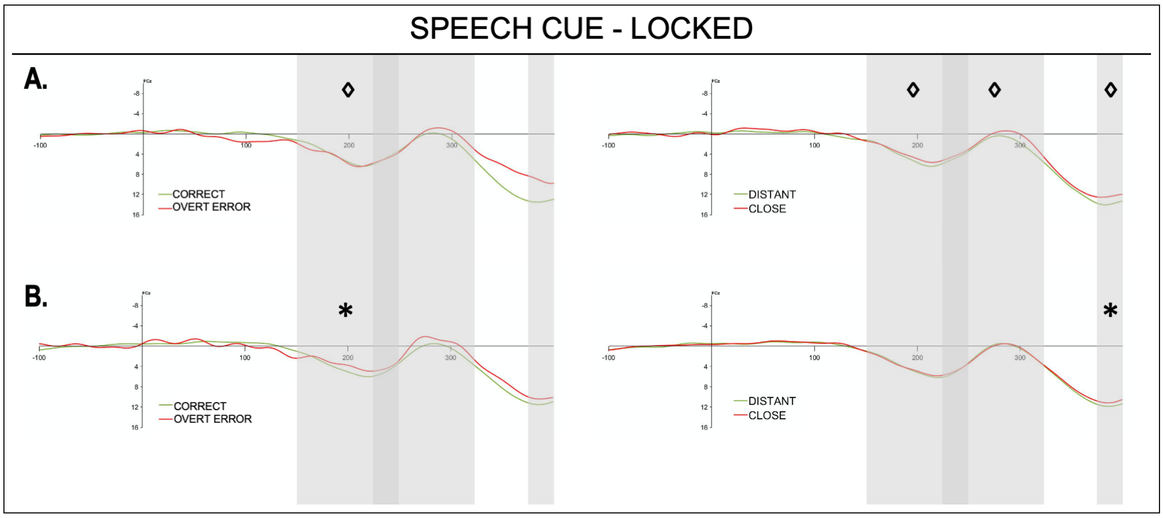

FIGURE 7 – Left : Grand average wave of FCz electrode for correct (green) and overt error (red) trials in Experiment I (panel A.) and Experiment II (panel B.) in speech cue-locked epochs. Right : Grand average wave of FCz electrode for AP close (red) and AP distant (green) correct trials in Experiment I (panel A.) and Experiment II (panel B.) in speech cue-locked epochs. Asterisks and diamonds indicate significant effect in one or the other GLM (9 fronto-central electrodes or all electrode) of the ERP.

Discussion

The principal aim of this study was to thoroughly investigate the temporal dynamics of the monitoring process, encompassing speech planning, speech-motor preparation, and articulation. Our specific focus was to explore potential variations in the temporal dynamics of monitoring of (a) correctly produced utterances with high error probability inflicted by either lexical or articulatory-phonetic related variables; and (b) incorrect utterances. To accomplish this, we conducted two EEG experiments. To have a comprehensive temporal map of the entire speech production process, the three contrasts of interest (high vs low lexically driven error probability, high vs low articulatory phonetically driven error probability, and errors vs correct trials) were examined on three distinct epochs that allowed us to asses speech planning, speech motor preparation and articulation. Finally, we employed two types of analyses (ERP and MVPA). It is worth noting that our emphasis was on robust effects that consistently emerged across both experiments, ensuring the reliability and validation of the results. In what follows we will discuss the results we obtained for initial stages of speech planning, speech motor planning and articulation respectively.

Initial stages of speech planning

The initial stages of speech planning were examined through the stimulus-locked epochs and through the early part of the response-locked epochs (see Figure **??**) . There was an effect of lexicality on the early pre-response part of the response-locked time window [-793 -693]. That is, based on the electrophysiological reponse across all electrodes, the MVPA distinguished above chance those correct trials that were more error prone due to the lexical response competition from those that were less error prone due to an absence of lexical competition. Interestingly, this effect falls almost in the same time-window where previous studies had observed ERP effects of semantic response conflict (GANUSHCHAK et al. 2008a) and of response conflict (MOLLER et al. 2006).Taken together, these findings seem to indicate that the kind of response competition affecting early stages of speech planning is related to the meaning or appropriateness of a potential response (i.e., competing responses that are semantically related as opposed to unrelated, or that are real words as opposed to non-words are all more appropriate).Additionally, this time window resulted in an overt errors vs. correct effect in Experiment I, similar to MOLLER et al. 2006, but the absence of this effect in Experiment II, despite a larger number of observations, indicates its lesser robustness. One possibility is that the effect was more robust in their study because all their critical trials were primed to result in lexical errors and were thus always response appropriate. To gain further insights on the nature of the effect of lexical error probability that we observed, we will briefly consider the results of the fMRI study by RUNNQVIST, CHANOINE et al. 2021, using the exact same task as here and the same stimuli as in our Experiment 1. In that study, lexically driven error probability engaged the Crus I of the cerebellum, which was linked to internal modeling of upcoming speech as a means of error monitoring. Interestingly, and consistent with this interpretation of predictive internal modeling, another study using EEG found that the readiness-potential (RP), was modulated by predictability (GRISONI, MOHR et al. 2019). Although we did not observe a distinguishable RP in our data, the timing of our MVPA lexicality effect preceding the verbal response is consistent with this component that is ususally observed preceding a motor response,. Thus, linking our findings with a previously found cerebellar origin of the effect and with modulations related to predictability occuring in similar time windows in previous studies, a plausible interpretation is that monitoring during the initial stages of speech planning is carried out through predictive internal modeling (e.g., PICKERING et al. 2013, RUNNQVIST, BONNARD et al. 2016, RUNNQVIST, CHANOINE et al. 2021; RUNNQVIST 2023).

Speech motor preparation

The speech motor preparation period was investigated through the speech-cue locked epochs and the late pre-response part of response-locked epochs. Leveraging multivariate pattern analysis (MVPA), we observed significant decoding rates for both overt errors as compared to correct trials and for high as compared to low articulatory-phonetic error probability on correct trials within a [164 264 ms] time-window after the speech cue. This time window is similar to the one where previously more negative event-related potential (ERP) for errors were reported(MOLLER et al. 2006). Our MVPA findings seem to mirror the EEG findings of MOLLER et al. 2006. in what concerns overt errors, but also extend their findings by showing that articulatory phonetic conflict also impacts the same time window. In that study, the spatial source of the EEG effect was localized to the medial frontal cortex (SMA), with a potential involvement of the anterior cingulate region. Additionally, the SMA activation during speech planning (post - speech cue) was previously linked by MOLLER et al. 2006 to conflicts in articulatory gestures planning, which is in line with recent findings by TODOROVIĆ et al. 2023 and the appearance of the pre-SMA region activation in situations involving increased articulatory-motor complexity (e.g., ALARIO et al. 2006). In summary, our findings are consistent with the SMAs known implication in phonetic encoding and articulation complexity, and provides support for the hypothesis that also response conflict involving articulatory phonetic representations may lead to greater SMA activation during speech motor planning. Connecting both sets of findings (fMRI and EEG), the SMA might work in concert with frontal and parietal structures and the superior cerebellum in a forward modeling loop preparing for motor execution (e.g., RIECKER et al. 2005, TODOROVIĆ et al. 2023). Regarding the late pre-response window of the response locked epochs, we observed a significant difference between errors ans correct trials in the ERPs [-115 to -15 ms]. We propose two plausible interpretations for this effect : The first interpretation is consistent with the findings discussed in the speech-planning section regarding the readiness potential (RP) from GRISONI, MILLER et al. 2017. In fact, this window corresponds to the greatest negativity of the RP, preceding the positive drop. The effect on this window supports the idea of prediction mechanisms as a monitoring component and implies that errors are inherently more unpredictable than correct trials. Alternatively, the effect may also be attributed to proprioceptive error detection involving somatosensory speech targets and stemming from the preparation of articulatory muscles before the onset of speaking (RIÈS, NADALET et al. 2020; GUENTHER et al. 2016).

Articulation

The articulation-related effects were examined by analyzing the post-response part of the response-locked epochs. The contrast of overt errors vs. correct trials yielded a significant effect in the time window [62-162 ms] that is consistent with the error-related negativity (ERN) (GANUSHCHAK et al. 2008a). We thus interpret the effect in the [62-162 ms] time window as the ERN, which is typically associated with conflict monitoring in the Anterior Cingulate Cortex (ACC) (DEHAENE et al. 1994, FALKENSTEIN et al. 1991,NOZARI et al. 2011). However, the ACC was found to be active only for overt errors (and not for internal monitoring) in previous studies (RUNNQVIST, CHANOINE et al. 2021; TODOROVIĆ et al. 2023), supporting the idea of the ACC having a feedback-related function especially for the time-window when the error is articulated. Surprisingly, multivariate pattern analysis (MVPA) did not show significant decoding rates for overt errors vs. correct trials, while the event-related potential (ERP) effects were robustly significant even after applying Bonferroni correction. This discrepancy raises questions about the underlying factors reflected by these two types of analyses and will be discussed further.

To summarize the global picture from both experiments : Internal monitoring of correct trials reveals early differences, discernible only during speech planning, suggesting that conflict on correct trials related to response appropriateness or meaning is resolved or substantially diminished by the time motor planning and articulation take place. On the other hand, articulatory-phonetic conflict on correct trials triggers differences only during speech motor preparation, but this conflict is also resolved or greatly diminished before the actual response. In contrast, overt fluent errors trigger differences during both early and late speech motor preparation, as well as during articulation. Thus, when examining correct trials, there seems to be a temporal coincidence between the moment of conflict emergence and the moment when the object of this conflict becomes task-relevant (i.e., response appropriateness or meaning related conflict arises when participants read and process meaning, and articulatory-phonetic conflict arises during speech motor planning). Note that while these results indeed suggest that processes operating on the already activated linguistic representations to be produced unfold in a sequential manner according to their task relevance, they may be compatible with both sequential and parallel processing in language production (e.g., FAIRS et al. 2021). For instance, FAIRS et al. 2021 proposed that while all dimensions of words are subject to a first pass of parallel activation (ignition) due to their holistic nature, selection and checking processes are likely to proceed sequentially during later reverberation processes. Concerning the mechanism underlying these internal monitoring effects, the combined evidence of this and previous studies suggest that this monitoring is carried out through predictive internal modeling. If correct, such internal modeling seems to generate error signals in a temporally distributed and task-relevant fashion (as opposed to only, for instance, upon phonological encoding). Conversely, overt errors show both pre-response and response differences in processing, suggesting that when the error signal of the predictive internal modeling is not enough to stop an error, additional, presumably more feedback dependent, processes are triggered during articulation. That is, the persistence of the effect related to overt errors suggests that errors may be detected multiple times, possibly through different processes. Concretely, error detection may occur through internal modeling before the response, followed by proprioceptive feedback, and finally, feedback-related mismatch. This interpretation aligns with fMRI studies that have identified the cerebellum, housing the internal models, as playing a significant role for monitoring of both error probability and overt errors (RUNNQVIST, CHANOINE et al. 2021 ; TODOROVIĆ et al. 2023), while other structures seemed to be exclusively triggered by overt errors (e.g., the ACC).

In a more comprehensive discussion of the present study, two intriguing questions remain unanswered. The first question pertains to determining the most pertinent event during speech production for precisely accounting for the underlying monitoring processes. Specifically, we consider two types of segmentation : one based on external events, such as stimulus presentation and speech cue appearance, and the other based on the participant's initiated response. While the response-locked epochs provided dynamic windows across participants, the stimulus-locked epochs remained stable. To effectively analyze groups of individuals, alignment in processes is essential to observe and statistically quantify the effects. Remarkably, the response-locked segmentation yielded a greater number of significant effects in both experiments and revealed the lexical effect during the speech planning stage, which coincided with the window of the stimulus-locked segmentation where this effect was not observed. This finding suggests that the initiation of the response may serve as the departure point that allows for alignment among individuals, as external events may introduce perceptual differences at multiple levels (e.g., reading speed, encoding period, lexical retrieval speed), potentially disaligning the group from a process-oriented perspective.

The second question emerges from the diverging results obtained from the two types of analyses conducted on the data. For example, while both the ERP analysis and MVPA yielded consistent cross-validated results in the early pre-response window [-483 to -383 ms], the strong error-related negativity (ERN) effect observed in the post-response window with the ERP analysis was not observed with MVPA. This discrepancy raises the question of what each type of analysis reflects and how to interpret the differences between them.

Finally, one potential limitation of our study should be pointed out. While our approach of cross validation at multiple levels (time window selection, consideration of significnt effects) ensures that the effects observed are robust and generalizable, it is possible that this rather stringent approach made certain true but more subtle effects go undetected. As we focused on the discussion of the effects that we did observe here it does not compromise our conclusions, and can be addressed in future studies by, for instance, including the time windows reported as significant in one of our two experiments in more focused an hypothesis driven analyses.

Conclusion

Cross-validated results from two experiments revealed robust electrophysiological effects of high versus low lexically and articulatory phonetically driven error probability on correct trials, and of overt speech errors versus correct trials. Temporal dissociations were observed across these contrasts with lexicality affecting the early stages of speech planning, articulatory phonetic proximity the early stages of speech motor preparation, and overt errors affecting both early and late stages of speech motor preparation as well as articulation. These results suggest the presence of temporally distributed predictive internal modeling in charge of monitoring before articulation, and of an additional mechanism relying on somatosensory and auditory feedback recruited successively in the case of the occurrence of a speech error.

Funding

This work, carried out within the Institute of Convergence ILCB (ANR-16-CONV-0002), has benefited from support from the French government (France 2030), managed by the French National Agency for Research (ANR) and the Excellence Initiative of Aix-Marseille University (A*MIDEX). E.R. has benefited from support from the French government, managed by the French National Agency for Research (ANR) through a research grant (ANR-18-CE28-0013).

References

ALARIO, F., CHAINAY, H., LEHÉRICY, S., & COHEN, L. (2006). The role of the supplementary motor area (SMA) in word production [Publisher : Elsevier]. Brain Research, 1076(1), 129-143. https://doi.org/10.1016/j.brainres.2005.11.104 (cf. p. 22)

BATES, D., MAECHLER, M., BOLKER, B., & WALKER, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. https://doi.org/10.18637/jss.v067.i01 (cf. p. 9)

BATES, D., MAECHLER, M., BOLKER, B., & WALKER, S. (2021). Matrix : Sparse and Dense Matrix Classes and Methods [R package version 1.3-4]. https://CRAN.R-project.org/package=Matrix. (Cf. p. 9)

BAUS, C., SANTESTEBAN, M., RUNNQVIST, E., STRIJKERS, K., & COSTA, A. (2020). Characterizing lexicalization and self-monitoring processes in bilingual speech production. Journal of Neurolinguistics, 56, 100934. https://doi.org/10.1016/j.jneuroling.2020.100934 (cf. p. 3)

CHAMBERS, J., HASTIE, T., & PREGIBON, D. (1990). Statistical Models in S. In K. MOMIROVIĆ & V. MILDNER (Éd.), Compstat (p. 317-321). Physica-Verlag HD. https://doi.org/10.1007/978-3-642-50096-1_48. (Cf. p. 9)

DEHAENE, S., POSNER, M. I., & TUCKER, D. M. (1994). Localization of a Neural System for Error Detection and Compensation [_eprint : https ://doi.org/10.1111/j.1467-9280.1994.tb00630.x]. Psychological Science, 5(5), 303-305. https://doi.org/10.1111/j.1467-9280.1994.tb00630.x (cf. p. 22)

DELORME, A., & MAKEIG, S. (2004). EEGLAB : an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods, 134(1), 9-21. https://doi.org/10.1016/j.jneumeth.2003.10.009 (cf. p. 11)

FAIRS, A., MICHELAS, A., DUFOUR, S., & STRIJKERS, K. (2021). The Same Ultra-Rapid Parallel Brain Dynamics Underpin the Production and Perception of Speech. Cerebral Cortex Communications, 2(3), tgab040. https://doi.org/10.1093/texcom/tgab040 (cf. p. 23)

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. https://doi.org/10.1016/0013-4694(91)90062-9 (cf. p. 22)

GANUSHCHAK, L. Y., & SCHILLER, N. O. (2006). Effects of time pressure on verbal self-monitoring : An ERP study. Brain Research, 1125(1), 104-115. https://doi.org/10.1016/j.brainres.2006.09.096 (cf. p. 2)

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. https://doi.org/10.1162/jocn.2008.20514 (cf. p. 2, 3, 21, 22)

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. https://doi.org/10.1016/j.neuroimage.2007.09.001 (cf. p. 3)

GAUVIN, H. S., DE BAENE, W., BRASS, M., & HARTSUIKER, R. J. (2016). Conflict monitoring in speech processing : An fMRI study of error detection in speech production and perception. NeuroImage, 126, 96-105. https://doi.org/10.1016/j.neuroimage.2015.11.037 (cf. p. 2)

GRISONI, L., MILLER, T. M., & PULVERMÜLLER, F. (2017). Neural Correlates of Semantic Prediction and Resolution in Sentence Processing. The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 37(18), 4848-4858. https://doi.org/10.1523/JNEUROSCI.2800-16.2017 (cf. p. 22)

GRISONI, L., MOHR, B., & PULVERMÜLLER, F. (2019). Prediction mechanisms in motor and auditory areas and their role in sound perception and language understanding. NeuroImage, 199, 206-216. https://doi.org/10.1016/j.neuroimage.2019.05.071 (cf. p. 12, 15, 21)

GUENTHER, F. H., & HICKOK, G. (2016). Neural Models of Motor Speech Control. In Neurobiology of Language (p. 725-740). Elsevier. https://doi.org/10.1016/B978-0-12-407794-2.00058-4. (Cf. p. 22)

HANSEN, S. J., MCMAHON, K. L., & de ZUBICARAY, G. I. (2019a). Neural Mechanisms for Monitoring and Halting of Spoken Word Production. Journal of Cognitive Neuroscience, 31(12), 1946-1957. https://doi.org/10.1162/jocn_a_01462 (cf. p. 4)

HANSEN, S. J., MCMAHON, K. L., & de ZUBICARAY, G. I. (2019b). The neurobiology of taboo language processing : fMRI evidence during spoken word production. Social Cognitive and Affective Neuroscience, 14(3), 271-279. https://doi.org/10.1093/scan/nsz009 (cf. p. 4)

HARTSUIKER, R. J., CORLEY, M., & MARTENSEN, H. (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly : Related beply to Baars et al. (1975). Journal of Memory and Language, 52(1), 58-70. https://doi.org/10.1016/j.jml.2004.07.006 (cf. p. 2)

HARTSUIKER, R. J., & KOLK, H. H. J. (2001). Error Monitoring in Speech Production : A Computational Test of the Perceptual Loop Theory. Cognitive Psychology, 42(2), 113-157. https://doi.org/10.1006/cogp.2000.0744 (cf. p. 2)

HOTHORN, T., BRETZ, F., & WESTFALL, P. (2021). multcomp : Simultaneous Inference in General Parametric Models [R package version 1.4-25]. https://CRAN.R-project.org/package=multcomp. (Cf. p. 10)

INC., T. M. (2020). MATLAB version : 9.8.0 (R2020a). Natick, Massachusetts, United States. https://www.mathworks.com. (Cf. p. 11)

KING, J.-R., & DEHAENE, S. (2014). Characterizing the dynamics of mental representations : the temporal generalization method. Trends in Cognitive Sciences, 18(4), 203-210. https://doi.org/10.1016/j.tics.2014.01.002 (cf. p. 4)

KUZNETSOVA, A., BROCKHOFF, P. B., & CHRISTENSEN, R. H. B. (2021). lmerTest : Tests in Linear Mixed Effects Models [R package version 3.1-4]. https://CRAN.R-project.org/package=lmerTest. (Cf. p. 10)

LEVELT, W. (1983). Monitoring and self-repair in speech. Cognition, 14(1), 41-104. https://doi.org/10.1016/0010-0277(83)90026-4 (cf. p. 2)

LOPEZ-CALDERON, J., & LUCK, S. J. (2014). ERPLAB : an open-source toolbox for the analysis of event-related potentials. Frontiers in Human Neuroscience, 8. Récupérée 8 mai 2023, à partir de https://www.frontiersin.org/articles/10.3389/fnhum.2014.00213 (cf. p. 11)

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 (cf. p. 3)

MOLLER, J., JANSMA, B. M., RODRIGUEZ-FORNELLS, A., & MUNTE, T. F. (2006). What the Brain Does before the Tongue Slips. Cerebral Cortex, 17(5), 1173-1178. https://doi.org/10.1093/cercor/bhl028 (cf. p. 2, 3, 21, 22)

NIZIOLEK, C. A., & GUENTHER, F. H. (2013). Vowel Category Boundaries Enhance Cortical and Behavioral Responses to Speech Feedback Alterations. The Journal of Neuroscience, 33(29), 12090-12098. https://doi.org/10.1523/JNEUROSCI.1008-13.2013 (cf. p. 2)

NOOTEBOOM, S. G., & QUENÉ, H. (2008). Self-monitoring and feedback : A new attempt to find the main cause of lexical bias in phonological speech errors [Accepted : 2017-06-09T19 :35 :25Z]. Journal of Memory and Language, 58, 837-861. Récupérée 8 septembre 2021, à partir de http://localhost/handle/1874/350522 (cf. p. 6)

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. https://doi.org/10.1016/j.jml.2017.01.007 (cf. p. 2)

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. https://doi.org/10.1016/j.cogpsych.2011.05.001 (cf. p. 22)

OKADA, K., MATCHIN, W., & HICKOK, G. (2018). Neural evidence for predictive coding in auditory cortex during speech production. Psychonomic Bulletin & Review, 25(1), 423-430. https://doi.org/10.3758/s13423-017-1284-x (cf. p. 4)

OPPENHEIM, G. M., & DELL, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. Cognition, 106(1), 528-537. https://doi.org/10.1016/j.cognition.2007.02.006 (cf. p. 6)

PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V., VANDERPLAS, J., PASSOS, A., COURNAPEAU, D., BRUCHER, M., PERROT, M., & DUCHESNAY, E. (2011). Scikit-learn : Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830 (cf. p. 13).

PICKERING, M. J., & GARROD, S. (2013). An integrated theory of language production and comprehension. Behavioral and Brain Sciences, 36(4), 329-347. https://doi.org/10.1017/S0140525X12001495 (cf. p. 21)

POSTMA, A., & NOORDANUS, C. (1996). Production and Detection of Speech Errors in Silent, Mouthed, Noise-Masked, and Normal Auditory Feedback Speech [Publisher : SAGE Publications Ltd]. Language and Speech, 39(4), 375-392. https://doi.org/10.1177/002383099603900403 (cf. p. 2)

PROTOPAPAS, A. (2007). Check Vocal : A program to facilitate checking the accuracy and response time of vocal responses from DMDX. Behavior Research Methods, 39(4), 859-862. https://doi.org/10.3758/BF03192979 (cf. p. 8)

R CORE TEAM. (2022). R : A Language and Environment for Statistical Computing [Version 4.2.2, released on 2022-10-31]. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/. (Cf. p. 9, 10)

RIECKER, A., MATHIAK, K., WILDGRUBER, D., ERB, M., HERTRICH, I., GRODD, W., & ACKERMANN, H. (2005). fMRI reveals two distinct cerebral networks subserving speech motor control. Neurology, 64(4), 700-706. https://doi.org/10.1212/01.WNL.0000152156.90779.89 (cf. p. 22)

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), 1419-1436. https://doi.org/10.1162/jocn.2010.21467 (cf. p. 3)

RIÈS, S. K., NADALET, L., MICKELSEN, S., MOTT, M., MIDGLEY, K. J., HOLCOMB, P. J., & EMMOREY, K. (2020). Pre-output Language Monitoring in Sign Production. Journal of Cognitive Neuroscience, 32(6), 1079-1091. https://doi.org/10.1162/jocn_a_01542 (cf. p. 22)

RUNNQVIST, E. (2023). Self-Monitoring : The Neurocognitive Basis of Error Monitoring in Language Production [Num Pages : 23]. In Language Production. Routledge. (Cf. p. 4, 21).

RUNNQVIST, E., BONNARD, M., GAUVIN, H. S., ATTARIAN, S., TRÉBUCHON, A., HARTSUIKER, R. J., & ALARIO, F.-X. (2016). Internal modeling of upcoming speech : A causal role of the right posterior cerebellum in non-motor aspects of language production. Cortex, 81, 203-214. https://doi.org/10.1016/j.cortex.2016.05.008 (cf. p. 5, 21)

RUNNQVIST, E., CHANOINE, V., STRIJKERS, K., PATTAMADILOK, C., BONNARD, M., NAZARIAN, B., SEIN, J., ANTON, J.-L., DOROKHOVA, L., BELIN, P., & ALARIO, F. X. (2021). Cerebellar and Cortical Correlates of Internal and External Speech Error Monitoring. Cerebral Cortex Communications, 2(1), tgab038. https://doi.org/10.1093/texcom/tgab038 (cf. p. 3, 21-23)

SAVARIAUX, C., PERRIER, P., & ORLIAGUET, J. P. (1995). Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube : A study of the control space in speech production. The Journal of the Acoustical Society of America, 98(5), 2428-2442. https://doi.org/10.1121/1.413277 (cf. p. 2)

SCHNEIDER, W., ESCHMAN, A., & ZUCCOLOTTO, A. (2002). E-prime User's Guide. (Cf. p. 7).

SEVERENS, E., JANSSENS, I., KÜHN, S., BRASS, M., & HARTSUIKER, R. J. (2011). When the brain tames the tongue : Covert editing of inappropriate language [_eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-8986.2011.01190.x]. Psychophysiology, 48(9), 1252-1257. https://doi.org/10.1111/j.1469-8986.2011.01190.x (cf. p. 2, 3)

SZIRTES, J., & VAUGHAN, H. (1977). Characteristics of cranial and facial potentials associated with speech production. Electroencephalography and Clinical Neurophysiology, 43(3), 386-396. https://doi.org/10.1016/0013-4694(77)90261-9 (cf. p. 5)

TEGHIPCO, A., OKADA, K., MURPHY, E., & HICKOK, G. (2023). Predictive Coding and Internal Error Correction in Speech Production. Neurobiology of Language (Cambridge, Mass.) 4(1), 81-119. https://doi.org/10.1162/nol_a_00088 (cf. p. 4)

TODOROVIĆ, S., ANTON, J.-L., SEIN, J., NAZARIAN, B., CHANOINE, V., RAUCHBAUER, B., KOTZ, S., & RUNNQVIST, E. (2023). Cortico-cerebellar monitoring of speech sequence production. Neurobiology of Language, 1-47. https://doi.org/10.1162/nol_a_00113 (cf. p. 3, 22, 23)

TREMBLAY, S., SHILLER, D. M., & OSTRY, D. J. (2003). Somatosensory basis of speech production. Nature, 423(6942), 866-869. https://doi.org/10.1038/nature01710 (cf. p. 2)

VOLFART, A., MCMAHON, K. L., HOWARD, D., & de ZUBICARAY, G. I. (2022). Neural Correlates of Naturally Occurring Speech Errors during Picture Naming in Healthy Participants. Journal of Cognitive Neuroscience, 35(1), 111-127. https://doi.org/10.1162/jocn_a_01927 (cf. p. 4)

WICKHAM, H. (2016). ggplot2 : Elegant Graphics for Data Analysis [R package version 3.4.3]. https://CRAN.R-project.org/package=ggplot2. (Cf. p. 10)

WICKHAM, H., FRANÇOIS, R., HENRY, L., & MÜLLER, K. (2021). dplyr : A Grammar of Data Manipulation [R package version 1.0.7]. https://CRAN.R-project.org/package=dplyr (cf. p. 10)

Supplementary materials

		Time-window	Overt error		Lexicality		AP proximity	
			p-values					
				(uncorrected)		(uncorrected)		(uncorrected)
Exp.1	Stimulus	62 162	0.132	0.066	0.904	0.452	0.460	0.230
		142 242	0.212	0.106	0.643	0.322	0.505	0.253
	Speech cue	164 264	2.141	0.714	0.353	0.118	0.141	**0.047**
		226 326	1.989	0.663	0.754	0.251	0.120	**0.040**
		372 400	**0.010**	**0.003**	2.545	0.848	0.098	**0.033**
	Response	-793 -693	0.806	0.269	0.483	0.161	1.739	0.580
		-483 -383	**0.001**	**<0.001**	1.083	0.361	**0.044**	**0.015**
		-115 - 15	**<0.001**	**<0.001**	0.783	0.261	0.349	0.116
		62 162	**0.003**	**0.001**	2.099	0.700	**0.036**	**0.012**
Exp.2	Stimulus	62 162	0.490	0.245	0.890	0.445	0.634	0.317
		142 242	1.560	0.780	0.578	0.289	1.745	0.872
	Speech cue	164 264	1.884	0.628	**0.048**	**0.016**	1.919	0.640
		226 326	1.475	0.492	0.132	**0.044**	2.850	0.950
		372 400	1.566	0.522	0.139	**0.046**	0.764	0.255
	Response	-793 -693	0.669	0.223	1.560	0.520	0.324	0.108
		-483 -383	**0.007**	**0.002**	0.215	0.072	1.856	0.619
		-115 - 15	**<0.001**	**<0.001**	0.855	0.285	**0.015**	**0.005**
		62 162	**0.002**	**0.001**	1.646	0.549	0.574	0.191

TABLE S.2 – Summary of p-values of 9 fronto-central electrode GLM. Significant values in one experiment are in bold, significant in both experiments are on grey background

		Time-window	Overt error		Lexicality		AP proximity	
			p-values					
				(uncorrected)		(uncorrected)		(uncorrected)
Exp.1	Stimulus	62 162	1.100	0.367	1.716	0.572	0.853	0.284
		142 242	0.243	0.081	2.480	0.827	2.855	0.952
	Speech cue	164 264	0.097	**0.032**	0.905	0.302	0.729	0.243
		226 326	0.161	0.054	2.834	0.945	1.670	0.557
		372 400	0.129	0.064	1.842	0.921	1.944	0.972
	Response	-793 -693	0.124	0.062	1.876	0.938	1.689	0.844
		-483 -383	2.734	0.911	1.035	0.345	1.637	0.546
		-115 - 15	2.351	0.784	1.351	0.450	1.537	0.512
		62 162	**0.014**	**0.005**	2.175	0.725	1.272	0.424
Exp.2	Stimulus	62 162	0.500	0.167	1.414	0.471	0.853	0.284
		142 242	0.283	0.094	0.780	0.260	2.855	0.952
	Speech cue	164 264	**0.028**	**0.009**	1.950	0.650	0.729	0.243
		226 326	0.070	**0.023**	1.391	0.464	1.670	0.557
		372 400	0.315	0.157	1.941	0.970	1.944	0.972
	Response	-793 -693	1.085	0.542	1.939	0.969	1.689	0.844
		-483 -383	2.265	0.755	0.148	**0.049**	1.637	0.546
		-115 - 15	1.962	0.654	0.217	0.072	1.537	0.512
		62 162	1.400	0.467	1.203	0.401	1.272	0.424

TABLE S.3 – Summary of p-values of all electrode GLM. Significant values are in bold.

		Time-window	Overt error	Lexicality	AP proximity
			p-values		
Exp.1	Stimulus	62 162	0,350	0,431	0,178
		142 242	**0,027**	0,271	0,809
	Speech cue	164 264	**0,002**	0,095	**<0.001**
		226 326	0,750	0,374	0,359
		372 400	0,099	0,073	**0,005**
	Response	-793 -693	**0,001**	**<0.001**	0,072
		-483 -383	**<0.001**	0,345	0,992
		-115 - 15	0,101	0,130	0,089
		62 162	0,081	**0,003**	**0,007**
Exp.2	Stimulus	62 162	0,085	0,692	0,111
		142 242	0,644	**0,028**	0,494
	Speech cue	164 264	**0,011**	0,060	**0,002**
		226 326	0,697	0,278	0,520
		372 400	0,578	0,214	0,629
	Response	-793 -693	0,097	**0,004**	0,055
		-483 -383	**0,001**	0,863	0,590
		-115 - 15	**0,012**	0,569	0,420
		62 162	0,583	0,151	0,201

TABLE S.4 – Summary of p-values of standard parametric one-tail paired t-test of the distribution of MVPA ROC_AUC values compared to chance level. Significant values are in bold, significant in both experiments are on grey background

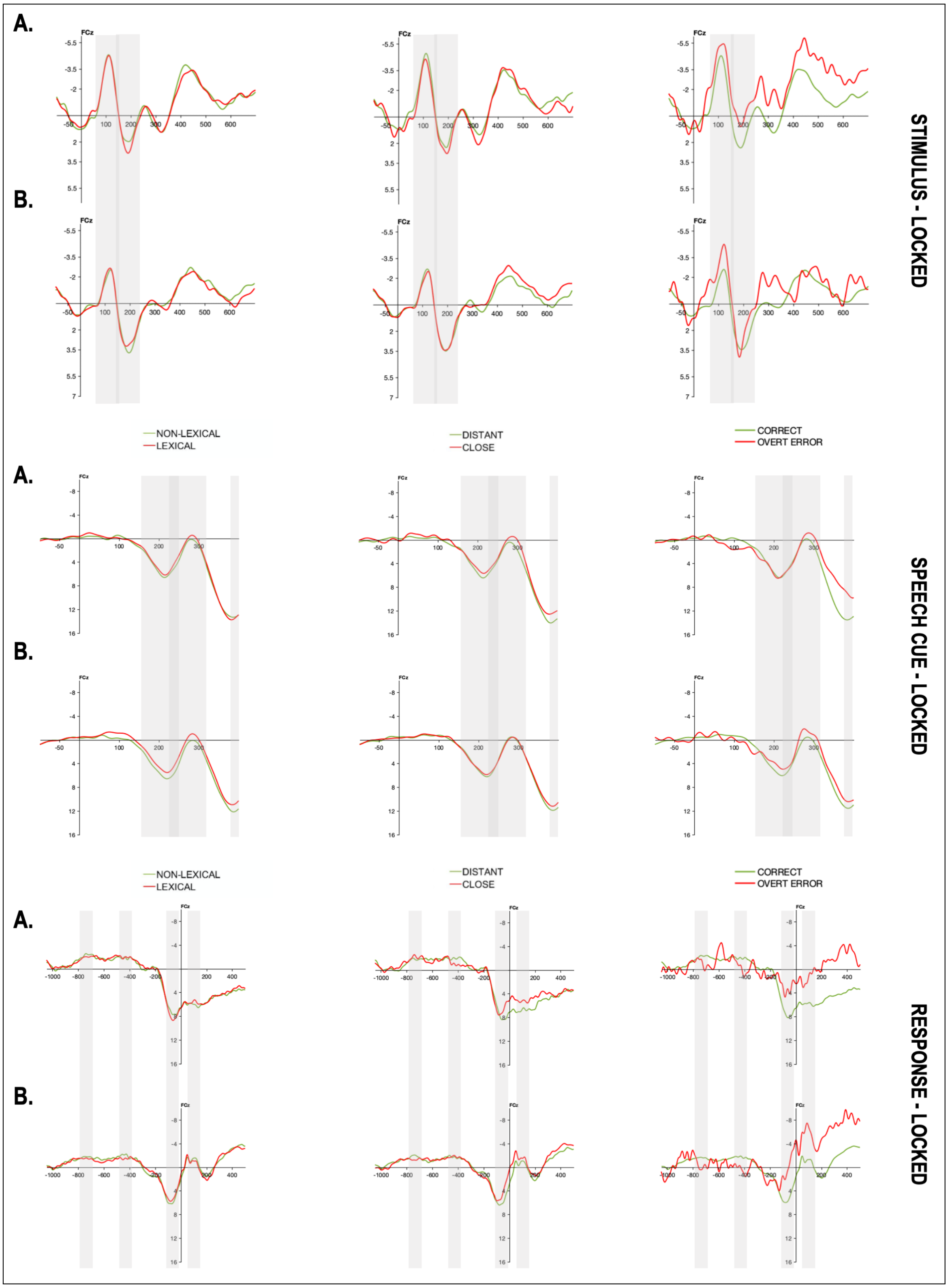

FIGURE S.8 – Grand average wave of FCz electrode for non-lexical (green) and lexical (red) trials (left column) ; for AP close (red) and AP distant (green) correct trials (middle column) ; for correct (green) and overt error (red) trials (right column) in Experiment I (panel A) and Experiment II (panel B) in stimulus - locked (top panel), speech cue-locked (central panel) and response-locked (down panel) epochs.

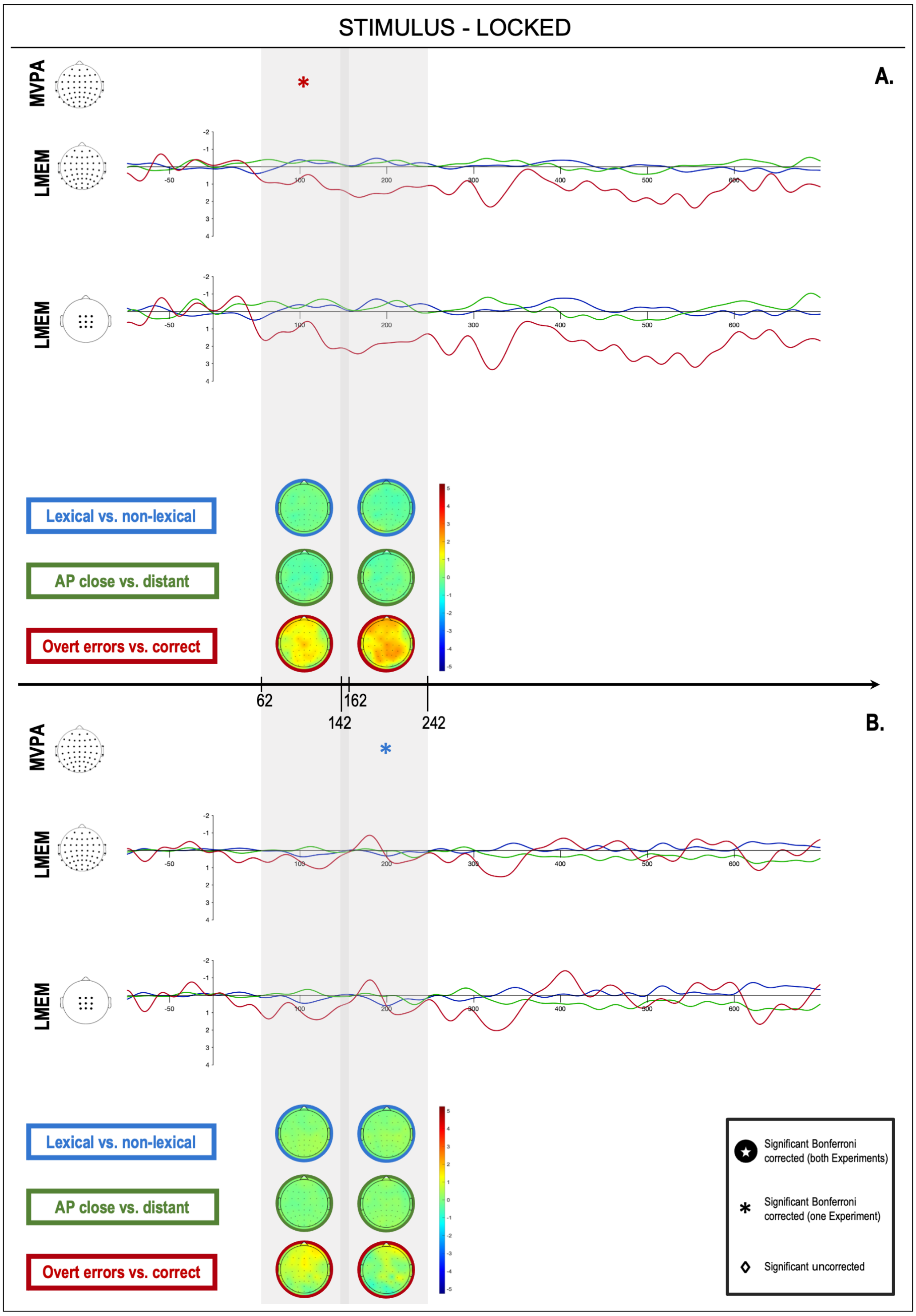

FIGURE S.9 – Summary of results of all performed analyses : MVPA, mean difference waves of all and 9 fronto-central electrodes and their topographic maps across Experiment I (panel A) and Experiment II (panel B.) in stimulus-locked epochs. Color code is used to differentiate the conditions : blue for lexicality, green for articulatory – phonetic proximity and red for overt error vs. correct contrast. Asterisks mark significant Bonferroni corrected p-values if observed in one Experiment, empty diamonds mark significant p-values without correction while bold stars mark significant p-values if observed in both Experiments.

Experimental lists of Experiment I

A				B			
lexical		nonlexical		lexical		nonlexical	
0SF	cadeau rocher	0SF	belote ciment	0SF	cage rap	0SF	carreau roulette
0SF	douche salle	0SF	bordée fournée	0SF	fagot rumeur	0SF	caverne mouton
0SF	lierre poupe	0SF	cÅŞur robe	0SF	farine mission	0SF	ceinture région
0SF	malade sinus	0SF	cote lueur	0SF	patte nièce	0SF	doc geste
0SF	marine fission	0SF	dague four	0SF	pierre loupe	0SF	fête lobe
0SF	ministre seringue	0SF	disque suite	0SF	pote nuits	0SF	glace fuite
0SF	natte pièce	0SF	façon gardon	0SF	radeau cocher	0SF	lanière fêtard
0SF	note puits	0SF	filleul monteur	0SF	salade minus	0SF	lapin fusée
0SF	rage cap	0SF	fosse masque	0SF	sinistre meringue	0SF	lecteur joker
0SF	ragot fumeur	0SF	jointure boulette	0SF	souche dalle	0SF	maçon journée
0SF	tenue voiture	0SF	jonction loto	0SF	venue toiture	0SF	pelle risque
1SF	titre voile	0SF	lamelle têtard	0SF	vitre toile	0SF	rampe soeur
1SF	butte lave	0SF	manière cuisson	1SF	berger vison	0SF	verbe tour
1SF	clé bol	0SF	panne roc	1SF	blé col	1SF	boisson caresse
1SF	crochet briquet	0SF	pion vase	1SF	bonus toucan	1SF	case sueur
1SF	dentier répit	0SF	recteur tracas	1SF	brochet criquet	1SF	casque fraise
1SF	dune lieu	0SF	rouleau tonton	1SF	butin local	1SF	cordée frimeur
1SF	durée pédale	0SF	tête liège	1SF	clic foin	1SF	coupe frange
1SF	flic coin	1SF	barreau légion	1SF	gosier râteau	1SF	fonction troupier
1SF	foire prime	1SF	brique vieux	1SF	loir sac	1SF	gamelle ponton
1SF	gag troupe	1SF	contre braise	1SF	lune dieu	1SF	linge bourse
1SF	garage palette	1SF	coussin bouton	1SF	lutte bave	1SF	lion vote
1SF	gaule tare	1SF	fable place	1SF	maison raquette	1SF	molosse lardon
1SF	lutin bocal	1SF	filon croupier	1SF	parage galette	1SF	pause mouche
1SF	matin passage	1SF	forge course	1SF	patin massage	1SF	pelote fracas
1SF	nature ration	1SF	fraction paresse	1SF	poire frime	1SF	poussin savoir
1SF	pillage sommier	1SF	gerbe plaque	1SF	purée dédale	1SF	sable poudre
1SF	raison maquette	1SF	lampe boule	1SF	rame dose	1SF	tilleul buisson
1SF	rosier gâteau	1SF	moisson lavoir	1SF	rature nation	1SF	toison boulet
1SF	serre valve	1SF	peinture musée	1SF	rentier dépit	2SF	biche dieux
1SF	soir lac	1SF	pomme mâche	1SF	sillage pommier	2SF	bosse montre
1SF	tonus boucan	1SF	sapin poker	1SF	tag groupe	2SF	bouleau moto
1SF	verger bison	1SF	singe veste	1SF	taule gare	2SF	crique pousse
2SF	banque marque	1SF	touche grange	1SF	verre salve	2SF	flaque somme
2SF	casse tube	2SF	colosse primeur	2SF	bec mise	2SF	gorge canne
2SF	ciel fil	2SF	fiche selle	2SF	fiel cil	2SF	pilon bouture
2SF	coteau poupon	2SF	foudre soupe	2SF	manque barque	2SF	pointure conteur
2SF	dame rose	2SF	gousse cause	2SF	poteau coupon	2SF	tâche piège
2SF	faveur semelle	2SF	poison couture	2SF	saveur femelle	2SF	traction piment
2SF	mec bise	2SF	taverne poulet	2SF	tasse cube	2SF	vague foule

C				D			
lexical		nonlexical		lexical		nonlexical	
0SF	boule fosse	0SF	cadeau voiture	0SF	biche fraise	0SF	cil nièce
0SF	braise fiche	0SF	casse voile	0SF	bosse foule	0SF	farine galette
0SF	colosse monteur	0SF	faveur maquette	0SF	gorge frange	0SF	foin dalle
0SF	façon musée	0SF	lutin fission	0SF	lion piège	0SF	gare sac
0SF	grange forge	0SF	marine semelle	0SF	maçon fusée	0SF	loir toile
0SF	liège pion	0SF	natte coin	0SF	molosse conteur	0SF	lune cage
0SF	mâche touche	0SF	ragot palette	0SF	montre casque	0SF	lutte cube
0SF	masque contre	0SF	raison pédale	0SF	mouche tâche	0SF	maison cocher
0SF	vase cote	0SF	rosier sommier	0SF	verbe geste	0SF	poteau vison
1SF	veste gerbe	0SF	serre rose	0SF	vote case	0SF	purée rumeur
1SF	barreau cuisson	0SF	verger poupon	1SF	bouleau roulette	0SF	rentier pommier
1SF	bordée couture	1SF	butte cap	1SF	carreau buisson	0SF	salade dépit
1SF	course brique	1SF	ciel tare	1SF	ceinture piment	0SF	saveur massage
1SF	filleul tracas	1SF	coteau bocal	1SF	cordée bouture	0SF	sillage râteau
1SF	filon primeur	1SF	crochet fumeur	1SF	crique bourse	0SF	sinistre nation
1SF	four tête	1SF	dentier passage	1SF	fête tour	0SF	tag loupe
1SF	fraction têtard	1SF	douche puits	1SF	fonction journée	1SF	barque clic
1SF	jointure poker	1SF	dune prime	1SF	gamelle lardon	1SF	blé taule
1SF	jonction fournée	1SF	foire troupe	1SF	lanière moto	1SF	butin coupon
1SF	lamelle gardon	1SF	gag marque	1SF	lapin savoir	1SF	fagot toucan
1SF	lueur singe	1SF	garage ration	1SF	lecteur région	1SF	gosier toiture
1SF	manière loto	1SF	gaule lac	1SF	linge sueur	1SF	patin femelle
1SF	peinture ciment	1SF	lierre valve	1SF	pilon frimeur	1SF	pierre manque
1SF	place gousse	1SF	malade répit	1SF	pointure joker	1SF	poire groupe
1SF	plaque foudre	1SF	matin rocher	1SF	poudre flaque	1SF	pote fiel
1SF	recteur légion	1SF	mec lave	1SF	pousse glace	1SF	radeau mission
1SF	robe lampe	1SF	nature bison	1SF	rampe lobe	1SF	rame bec
1SF	roc disque	1SF	note lieu	1SF	risque doc	1SF	rap dieu
1SF	rouleau boulette	1SF	rage bise	1SF	soeur coupe	1SF	rature local
1SF	sapin lavoir	1SF	soir tube	1SF	somme pelle	1SF	tasse salve
1SF	selle pomme	1SF	tenue sinus	1SF	tilleul fracas	1SF	venue raquette
1SF	soupe coeur	1SF	titre fil	1SF	traction fêtard	1SF	verre bave
1SF	vieux dague	1SF	tonus seringue	1SF	vague dieux	1SF	vitre dose
2SF	belote poulet	2SF	banque poupe	2SF	boisson mouton	2SF	berger minus
2SF	cause panne	2SF	clé pièce	2SF	canne pause	2SF	bonus meringue
2SF	coussin paresse	2SF	dame bol	2SF	caverne troupier	2SF	brochet dédale
2SF	moisson bouton	2SF	durée gâteau	2SF	pelote boulet	2SF	mise nuits
2SF	poison tonton	2SF	flic salle	2SF	poussin caresse	2SF	parage criquet
2SF	suite fable	2SF	ministre boucan	2SF	sable fuite	2SF	patte col
2SF	taverne croupier	2SF	pillage briquet	2SF	toison ponton	2SF	souche frime

E				F			
lexical		nonlexical		lexical		nonlexical	
0SF	cap rage	0SF	boulette jointure	0SF	cocher radeau	0SF	cité biche
0SF	fission marine	0SF	ciment belote	0SF	dalle souche	0SF	fêtard lanière
0SF	fumeur ragot	0SF	cuisson manière	0SF	loupe pierre	0SF	fuite glace
0SF	pièce natte	0SF	four dague	0SF	meringue sinistre	0SF	fusée lapin
0SF	poupe lierre	0SF	fournée bordée	0SF	minus salade	0SF	geste doc
0SF	puits note	0SF	gardon façon	0SF	mission farine	0SF	journée maçon
0SF	rocher cadeau	0SF	liège tête	0SF	nièce patte	0SF	lobe fête
0SF	salle douche	0SF	loto jonction	0SF	nuits pote	0SF	mouton caverne
0SF	seringue ministre	0SF	lueur cote	0SF	rap cage	0SF	région ceinture
0SF	sinus malade	0SF	masque fosse	0SF	rumeur fagot	0SF	risque pelle
0SF	voile titre	0SF	monteur filleul	0SF	toile vitre	0SF	roulette carreau
1SF	voiture tenue	0SF	robe coeur	0SF	toiture venue	0SF	soeur rampe
1SF	bison verger	0SF	roc panne	1SF	bave lutte	0SF	tour verbe
1SF	bocal lutin	0SF	suite disque	1SF	col blé	1SF	boulet toison
1SF	bol clé	0SF	têtard lamelle	1SF	criquet brochet	1SF	bourse linge
1SF	boucan tonus	0SF	tonton rouleau	1SF	dédale purée	1SF	buisson tilleul
1SF	briquet crochet	0SF	tracas recteur	1SF	dépit rentier	1SF	caresse boisson
1SF	coin flic	0SF	vase pion	1SF	dieu lune	1SF	fracas pelote
1SF	gâteau rosier	1SF	boule lampe	1SF	dose rame	1SF	fraise casque
1SF	lac soir	1SF	bouton coussin	1SF	foin clic	1SF	frange coupe
1SF	lave butte	1SF	braise contre	1SF	frime poire	1SF	frimeur cordée
1SF	lieu dune	1SF	course forge	1SF	galette parage	1SF	joker lecteur
1SF	maquette raison	1SF	croupier filon	1SF	gare taule	1SF	lardon molosse
1SF	palette garage	1SF	grange touche	1SF	groupe tag	1SF	mouche pause
1SF	passage matin	1SF	lavoir moisson	1SF	local butin	1SF	ponton gamelle
1SF	pédale durée	1SF	légion barreau	1SF	massage patin	1SF	poudre sable
1SF	prime foire	1SF	mâche pomme	1SF	nation rature	1SF	savoir poussin
1SF	ration nature	1SF	musée peinture	1SF	pommier sillage	1SF	sueur case
1SF	répit dentier	1SF	paresse fraction	1SF	raquette maison	1SF	troupier fonction
1SF	rose dame	1SF	place fable	1SF	râteau gosier	1SF	vote lion
1SF	sommier pillage	1SF	plaque gerbe	1SF	sac loir	2SF	bouture pilon
1SF	tare gaule	1SF	poker sapin	1SF	salve verre	2SF	canne gorge
1SF	troupe gag	1SF	veste singe	1SF	toucan bonus	2SF	conteur pointure
1SF	valve serre	1SF	vieux brique	1SF	vison berger	2SF	foule vague
2SF	bise mec	2SF	cause gousse	2SF	barque manque	2SF	montre bosse
2SF	fil ciel	2SF	couture poison	2SF	cil fiel	2SF	moto bouleau
2SF	marque banque	2SF	poulet taverne	2SF	coupon poteau	2SF	piège tâche
2SF	poupon coteau	2SF	primeur colosse	2SF	cube tasse	2SF	piment traction
2SF	semelle faveur	2SF	selle fiche	2SF	femelle saveur	2SF	pousse crique
2SF	tube casse	2SF	soupe foudre	2SF	mise bec	2SF	somme flaque

G				H			
lexical		nonlexical		lexical		nonlexical	
0SF	contre masque	0SF	coin natte	0SF	case vote	0SF	cage lune
0SF	cote vase	0SF	fission lutin	0SF	casque montre	0SF	cocher maison
0SF	fiche braise	0SF	maquette faveur	0SF	conteur molosse	0SF	cube lutte
0SF	forge grange	0SF	palette ragot	0SF	foule bosse	0SF	dalle foin
0SF	fosse boule	0SF	pédale raison	0SF	fraise biche	0SF	dépit salade
0SF	gerbe veste	0SF	poupon verger	0SF	frange gorge	0SF	galette farine
0SF	monteur colosse	0SF	rose serre	0SF	fusée maçon	0SF	loupe tag
0SF	musée façon	0SF	semelle marine	0SF	geste verbe	0SF	massage saveur
0SF	pion liège	0SF	sommier rosier	0SF	piège lion	0SF	nation sinistre
0SF	touche mâche	0SF	voile casse	0SF	tâche mouche	0SF	nièce cil
1SF	boulette rouleau	0SF	voiture cadeau	1SF	bourse crique	0SF	pommier rentier
1SF	brique course	1SF	bise rage	1SF	bouture cordée	0SF	râteau sillage
1SF	ciment peinture	1SF	bison nature	1SF	buisson carreau	0SF	rumeur purée
1SF	coeur soupe	1SF	bocal coteau	1SF	coupe soeur	0SF	sac gare
1SF	couture bordée	1SF	cap butte	1SF	dieux vague	0SF	toile loir
1SF	cuisson barreau	1SF	fil titre	1SF	doc risque	0SF	vison poteau
1SF	dague vieux	1SF	fumeur crochet	1SF	fêtard traction	1SF	bave verre
1SF	disque roc	1SF	lac gaule	1SF	flaque poudre	1SF	bec rame
1SF	foudre plaque	1SF	lave mec	1SF	fracas tilleul	1SF	clic barque
1SF	fournée jonction	1SF	lieu note	1SF	frimeur pilon	1SF	coupon butin
1SF	gardon lamelle	1SF	marque gag	1SF	glace pousse	1SF	dieu rap
1SF	gousse place	1SF	passage dentier	1SF	joker pointure	1SF	dose vitre
1SF	lampe robe	1SF	prime dune	1SF	journée fonction	1SF	femelle patin
1SF	lavoir sapin	1SF	puits douche	1SF	lardon gamelle	1SF	fiel pote
1SF	légion recteur	1SF	ration garage	1SF	lobe rampe	1SF	groupe poire
1SF	loto manière	1SF	répit malade	1SF	moto lanière	1SF	local rature
1SF	poker jointure	1SF	rocher matin	1SF	pelle somme	1SF	manque pierre
1SF	pomme selle	1SF	seringue tonus	1SF	piment ceinture	1SF	mission radeau
1SF	primeur filon	1SF	sinus tenue	1SF	région lecteur	1SF	raquette venue
1SF	singe lueur	1SF	tare ciel	1SF	roulette bouleau	1SF	salve tasse
1SF	têtard fraction	1SF	troupe foire	1SF	savoir lapin	1SF	taule blé
1SF	tête four	1SF	tube soir	1SF	sueur linge	1SF	toiture gosier
1SF	tracas filleul	1SF	valve lierre	1SF	tour fête	1SF	toucan fagot
2SF	bouton moisson	2SF	bol dame	2SF	boulet pelote	2SF	col patte
2SF	croupier taverne	2SF	boucan ministre	2SF	caresse poussin	2SF	criquet parage
2SF	fable suite	2SF	briquet pillage	2SF	fuite sable	2SF	dédale brochet
2SF	panne cause	2SF	gâteau durée	2SF	mouton boisson	2SF	frime souche
2SF	paresse coussin	2SF	pièce clé	2SF	pause canne	2SF	meringue bonus
2SF	poulet belote	2SF	poupe banque	2SF	ponton toison	2SF	minus berger
2SF	tonton poison	2SF	salle flic	2SF	troupier caverne	2SF	nuits mise

Experimental list of Experiment II

lexical		nonlexical	
0SF	2SF	0SF	2SF
balade sonnet	bague digue	ballet sapeur	bande membre
boule fosse	banque marque	bourse foin	baroque patin
braise fiche	bêche peau	bride farce	beige poste
cadeau rocher	bêcheur piquet	courant record	beurrier passion
durée fiction	belote poulet	dépit fourmis	bible daube
festin dada	carte torse	façade dépense	cerne forces
fête bac	casse tube	ferme brute	coin tact
gaule sang	cause panne	glace faune	compte taux
grange forge	ciel fil	gué signe	confit puma
lavoir sapin	coteau poupon	lamelle sofa	coup puce
liqueur pétale	croupier taverne	légume pédale	croyante tampon
lueur secteur	daron bélier	levier souris	dopage barbu
mâche touche	derme ton	mari talent	double tombe
masque contre	faveur semelle	mer cape	fiston sondage
matin taxi	gage coût	miracle sérum	galion bascule
ministre seringue	gallon bourde	modèle famille	gamin bonheur
musée façon	garage bateau	muette top	guide corne
natte pièce	malais barquette	nombre pont	maillet baleine
note pain	mec bise	nymphe peuple	mot bave
poker jointure	mouture boisson	potion jumelle	muret bacille
rage pÃťle	nature marine	rabat ficus	niveau moquette
ragot fondeur	paresse coussin	roque pelle	palier cuiller
râpe paille	poing coupe	rythme poivre	plumet tournée
soupe liège	primeur tension	sucre loutre	pouce corde
tacle rÃťt	prune baume	tapis visage	proue berge
tenue voiture	suite fable	tireuse vanille	sÅŞur feutre
terrine vernis	têtard ponton	toge rêne	tabou pavot
vase cote	torche prÃťne	veau cure	tofu cobra
verre toile	tracas cantine	vieux taupe	tonne ponce
village panneau	veste gerbe	visite pétrin	vers gifle