Characterizing multi-word speech production using event-related potentials

Ries, S. K., Pinet, S., Nozari, N., Knight, R. T.

Abstract

Event-related potentials (ERPs) derived from electroencephalography (EEG) have proven useful for understanding linguistic processes during language perception and production. Words are commonly produced in sequences, yet most ERP studies have used single-word experimental designs. Single-word designs reduce potential ERP overlap in word sequence production. However, word sequence production engages brain mechanisms in different ways than single word production. In particular, speech monitoring and planning mechanisms are more engaged than for single words since several words must be produced in a short period of time. This study evaluates the feasibility of recording ERP components in the context of word sequence production, and whether separate components could be isolated for each word. Scalp EEG data were acquired while participants recited word sequences from memory at a regular pace, using a tongue-twister paradigm. The results revealed a fronto-central error-related negativity, previously associated with speech monitoring, which could be distinguished for each word. Its peak amplitude was sensitive to Cycle and Phonological Similarity. However, an effect of sequential production was also observable on baseline measures, indicating baseline shifts throughout the word sequence due to concurrent sustained medial-frontal EEG activity. We also report a late left anterior negativity (LLAN), associated with verbal response planning and execution, onsetting around 100 ms before the first word in each cycle and sustained throughout the rest of the cycle. This work underlines the importance of considering the contribution of transient and sustained EEG activity on ERPs, and provides evidence that ERPs can be used to study sequential word production.

1. Introduction

The use of electroencephalography (EEG) to study speech production has developed recently (e.g., Ganushchak & Schiller, 2006; Masaki et al., 2001; for a review, see Ganushchak et al., 2011) in comparison to speech and language perception and comprehension (for early reports, see Kutas & Hillyard, 1980; for a review see, Kutas & Federmeier, 2011). One of the main factors impeding this development has been the presence of articulatory-related electromyographic (EMG) artifacts, which heavily contaminate the EEG signal especially around the response (De Vos et al., 2010; Piai et al., 2014). Hence, most studies have focused on single word production to try to limit the impact of EMG artifacts in the EEG signal.

However, speech production outside of the testing room typically involves a sequence of multiple words as we usually speak in sentences. Multi-word production involves all the same processes as single-word production plus additional processes needed to inhibit the previous word and prime the next word in a manner that does not interfere with the production of the present word (e.g., Dell et al., 1997; Nozari et al., 2014). This is why naming the same picture is more difficult in the context of multiple words than in isolation (e.g., Schwartz & Hodgson, 2002). However, even though speakers are more likely to make errors when producing sequences of words, phrases, or sentences, they frequently detect and correct their errors. In fact, neurotypical speakers seem to be able to correct more of their errors when task demands increase the probability of errors (Nozari et al., 2019). Nozari et al. (2019) showed that the increase in the proportion of corrected errors as a function of error probability was not just detectable at the final position of the sentence (where interference from previous items is high), but was also observed on error-prone words earlier in the sentence. This finding implies the involvement of a fast and adaptable monitoring mechanism for each word to optimize sentence production.

A full understanding of how speakers produce connected speech entails understanding the monitoring and control processes that regulate production online, as the acts of production and monitoring are closely intertwined (Nozari, 2018). Incidentally, one of the most established EEG components related to performance monitoring (in language production and in other actions) is the error related negativity (ERN). The ERN, which was first reported outside of language (Falkenstein et al., 1991; Gehring & Knight, 2000), is a frontocentral component starting just before or around the onset of a speech error and peaking shortly after (Ganushchak & Schiller, 2006; Masaki et al., 2001). Importantly, the ERN is also present in correct trials in both linguistic (Riès et al., 2011) and non-linguistic tasks (Vidal et al., 2000), with amplitude reduction in correct trials compared to error trials. Importantly, the ERN in errors and correct trials has been shown to originate from the same source using both source localization and intracranial EEG (Bonini et al., 2014; Roger et al., 2010), and is proposed to index the same underlying action monitoring mechanism. We note that the ERN in correct trials is also sometimes referred to as the Ne-like wave (e.g., Allain et al., 2004; Burle et al., 2008; Riès et al., 2011), but we refer to it as the ERN in the current study. The uncovering of this component in spoken production, signing (Riès et al., 2020), as well as in other modalities of language production such as typing (Pinet & Nozari, 2020), was a critical piece of evidence supporting domain-general monitoring processes in language production (Hanley et al., 2016; Nozari et al., 2011; Riès et al., 2011; see Nozari & Novick, 2017 for a review). Another component which has been less studied, but which has been associated with response preparation and execution in language production is a slow-rising left lateral frontal component (usually seen at electrode FC5 in the 10-20 electrode placement system) (Riès et al., 2013). This left-lateralized component, which we refer to as the Late Left Anterior Negativity (LLAN), starts to rise around 350 ms after stimulus onset and peaks around vocal onset in single word picture naming. Left anterior activity has also been reported using scalp EEG in previous single word processing (at the same recording site in Klimesch et al., 2001) and sentence processing studies (e.g., Wlotko and Federmeier, 2012a, 2012b). In these studies (Wlotko and Federmeier, 2012a, 2012b), this left anterior activity has been linked to working memory or ambiguity resolution processes. While language production is a very different task than sentence processing and the activity observed in these studies did not necessarily have the same time-course or polarity than in language production, working memory and ambiguity resolution processes may also be at play, especially in demanding situations as in a tongue twister paradigm. Magnetoencephalography and intracranial EEG studies of picture naming have also reported late left anterior activities, starting no earlier than 300 ms after stimulus presentation (MEG: Salmelin et al., 1994; Sörös et al., 2003; Vilha et al., 2006; Hultén et al., 2008; ECoG: Edwards et al., 2010, Ries et al., 2017), and which were sometimes also found to be time-locked to vocal-onset (Edwards et al., 2010; Ries et al., 2017; Salmelin et al., 1994). The timing and location of this activity has been often interpreted as consistent with multiple cognitive processes including word retrieval, articulatory programming and preparation. We therefore focused on both of these components, the ERN and the LLAN, in the present study.

Given the higher demands of monitoring in multi- vs. single-word production, the ERN and other indices of monitoring and response preparation such as the LLAN can shed light on processes that might be differentially involved in the production of multi- vs. single-word utterances. However, recovering these indices from individual words during the production of word sequences constitutes a major challenge for several reasons. Indeed, the components that have been described in the literature are known to take place over several hundreds of milliseconds around stimulus onset and around the response (e.g., Riès et al., 2013). With the rate of normal word output being around 3 words per second (Levelt, 1993), the EEG components associated with neighboring words are likely to be overlapping, increasing the difficulty of studying them at the individual word level. Moreover, additional processes are likely involved in producing word sequences, such as planning of the whole sequence or phrases (Martin et al., 2010). EEG components associated with these processes are not well described (although see, Habets et al., 2008). Additional EEG components present during word sequences (such as the frontal negativity or P3b reported in Habets et al., 2008), complicates translating EEG components associated with single word production to a sequential context. Finally, the techniques developed to reduce the impact of articulation-related EMG artifacts, such as Blind Source Separation based on Canonical Correlation Analysis (BSS-CCA), have been developed for single word production studies (e.g., de Vos et al., 2010). Currently, it is unclear whether BSS-CCA would be efficient in the case of word sequences.

Moreover, it is possible that ERP effects are additive with overlapping activity from the prior word or that the generation of sustained brain activity when we produce words in a sequence overlaps with transient ERPs. This cumulative or sustained brain activity may also be sensitive to our experimental manipulations. If that is the case, then the neuronal response may increase or decrease from one word to the next in a sequence and result in shifts in baseline measures. The role of transient versus sustained brain activity (sometimes also equated with evoked versus induced or phase-locked versus non phase-locked brain activity) has been more frequently studied using brain oscillations (e.g., Chen et al., 2012; Laaksonen et al., 2012; Luu et al., 2004; Piai, Roelofs, et al., 2014), although see West et al. (2012) for a study dissociating transient versus sustained ERPs. These transient versus sustained brain responses have been linked to bottom-up versus top-down or reactive versus proactive control mechanisms respectively (e.g., Chen et al., 2012; West et al., 2012). When planning and uttering a sequence of words, both types of cognitive control mechanisms are likely to be at play as participants may be anticipating and trying to prevent interference as well as reacting to interference after it occurs. In the case of sustained responses, baseline shifts may be observed between ERP components time-locked to successive words in the sequence. Taking into account these baseline shifts when analyzing ERPs may provide insight into the concurrent engagement of transient and slow-wave or sustained brain mechanisms in the production of word sequences. Also, since ERP amplitudes are typically calculated relative to the baseline (Luck, 2014), the impact of baseline shifts on the analysis of ERP amplitudes needs to be closely examined.

For all these reasons, studies of word sequence production have generally focused on only part of the sentence, such as the time-interval before the sentence starts being uttered (Habets et al., 2008) or the time-interval preceding the last word of the sentence (e.g., Piai et al., 2015, 2017). A limitation of these paradigms is that they do not allow for studying what happens within the sequence for each word individually. One of the rare studies which attempted to investigate ERPs time-locked to individual words in multi-word utterances is provided by Acheson & Hagoort (2014). However, their study was a nonword reading study and additional processes associated with stimulus perception and reading were engaged during the time-window of interest. In addition, the objective of the study was to compare performance monitoring-related components in a linguistic (i.e., tongue-twister) versus non-linguistic (i.e., arrow-version of the Flanker task) conflict tasks, but the potential impact of producing several words in a sequence on the ERP components was not assessed.

1.1. The current study

We addressed the challenge of using EEG to study word sequence production by characterizing the components associated with words in all positions of the sequence. We used a well-established tongue twister paradigm (Nozari & Dell, 2012; Oppenheim & Dell, 2008) to study previously described speech monitoring and motor preparation related components time-locked to the onset of the verbal response (Riès et al., 2013; Riès et al., 2011). The paradigm we used is particularly well-suited for the study of these monitoring and motor preparation processes as it is designed to elicit conflict at the phonological and articulatory planning levels by manipulating the phonological similarity between words. Participants were asked to recite 4-word tongue-twisters, such as “pub bust bun puff”, twice (i.e., in two cycles) after seeing the word sequence on the screen and practicing the sequence at a specified pace given by a metronome. The metronome pace training was used to ensure that the words produced would be sufficiently far apart to allow for the examination of ERP components time-locked to each word. Importantly, the words were not present on the screen and the metronome stopped before the critical testing period during which the participants recited the words from memory at the practiced pace. This design assured the processes studied were production-related and are not linked to reading or auditory perception of the metronome’s beeps. EEG was recorded during two Cycles of the four words (referred to as words 1-4 with regard to their Word Position in the sequence).

This paradigm has previously not been used in EEG studies, presumably because of the challenges linked to word sequence production and articulatory artifacts noted above. However, this paradigm is suitable for studying monitoring processes because tongue-twisters generate high levels of conflict. This is due to the phonological overlap between the words in the sequence: “pub”, “bust”, “bun”, and “puff” all share at least one segment (e.g., /^/ in this example, but differ in others). In addition, the repeated production of such words (in the different cycles) has been shown to elicit robust interference (e.g., Breining et al., 2016; Nozari et al., 2016). If conflict builds up over more repetitions, one could expect to see effects of both Cycle and Word Position across the sequence. Specifically, we predicted higher levels of conflict in the second compared to the first cycle, and for words in later positions in the sequence (e.g., for “bun” and “puff” compared to “pub” and “bust” in the above example). The words employed also utilize a subtle manipulation of onset similarity: for half the tongue-twisters, the onsets of the first and fourth words differed from those of the second and third words in one phonetic feature (e.g., p/b as in “pub bust bun puff”: differing only in voicing). In the other half, it differed in two phonetic features (e.g., t/b as in “tub bust bun tough”: differing in voicing *and* place of articulation). Behaviorally, there is a small effect of Phonological Similarity on error rates, such that more similar onsets produce more errors (Oppenheim & Dell, 2008). If this is due to greater conflict between phonemes, then we should see this increased conflict in both behavior and ERPs. In sum, we focused on effects of Cycle, Word Position within cycle, and Phonological Similarity.

We focused on the error-related negativity (ERN), a response-related component reflecting action monitoring including speech monitoring, which has been widely used to study conflict with EEG in language (Acheson & Hagoort, 2014; Ganushchak & Schiller, 2006, 2008; Masaki et al., 2001; Pinet & Nozari, 2019; Riès et al., 2011) and in non-linguistic tasks (Allain et al., 2004; Burle et al., 2008; Falkenstein et al., 1991; Gehring & Knight, 2000; Roger et al., 2010). ERN amplitude increases in higher conflict situations (e.g., for errors compared to correct trials), and originates in the medial frontal cortex, including the supplementary motor area (SMA, Bonini et al., 2014) and/or neighboring dorsal anterior cingulate cortex (ACC, e.g., Debener et al., 2005). We also focused on a left-lateralized frontal component, the LLAN, (usually maximal at electrode FC5 in the 10-20 electrode placement system), which has been associated with motor preparation and execution processes during language production, and localized to the left lateral frontal cortex using source localization of scalp EEG data during picture naming (Riès et al., 2013).

We predicted higher ERN amplitudes in correct trials in the second versus first cycle and with increasing position within cycle. Indeed, the amplitude of the ERN in correct trials has been shown to increase in more difficult and error-prone situations (e.g., Acheson & Hagoort, 2014; Allain et al., 2004; Riès et al., 2011). We also predicted greater ERN amplitude in correct trials in the phonologically similar versus dissimilar sequences, as phonologically similar sequences have been shown to elicit greater error rates, and hence greater conflict, than phonologically dissimilar sequences (Oppenheim & Dell, 2008). Finally, if the left-lateralized frontal component is important for motor preparation processes, this component may show a slow ramp-up in amplitude throughout the sequence in line with incremental articulatory planning proposals which posit that articulatory planning can happen simultaneously with speech production (e.g., Ferreira & Swets, 2002). In particular, we predicted that the amplitude of the LLAN would increase with increasing Word Position in each cycle.

Prior to examining ERP amplitudes, we established whether medial frontal and left frontal EEG components associated with each individual word could be distinguished. We also evaluated whether the slope of the rising EEG components was significantly different from zero across words in the sequence and we examined whether the ERP amplitudes of the ERN and left frontal ERP were sensitive to Cycle, Word Position, and Phonological Similarity. Finally, we evaluated whether there was a shift in baseline between different EEG components associated with each word in the sequence and whether any shift in baseline was influenced by Cycle, Word Position, and Phonological Similarity. Importantly, the return to baseline is a pre-requisite for studying the amplitude of EEG components independently for each word, and constitutes a standard in the study of event-related potentials (ERPs) (Luck, 2014). By investigating the change to the components described above and the possible changes to the baseline in between words, we aimed to better characterize the nature of sustained versus transient activities, possibly reflecting top-down or proactive versus bottom-up or reactive processes, respectively. In particular, if sustained EEG activity is also sensitive to a general increase of conflict, then an increase in sustained EEG activity may be observed between cycle 1 and cycle 2, across positions, and in similar versus dissimilar sequences, which would be visible in the baseline comparison measures. Our aim was to assess the impact of producing words in a sequence on EEG measures typically used in single word EEG production studies.

2. Methods

2.1. Participants

Seventeen subjects (11 females) were recruited to participate in the study (mean age: 22 years old, SD = 4.5 years). Of these 17 subjects, 12 were included in the analyses (8 females, mean age: 22 years old, SD = 4.2 years) as the remaining 5 participants either did not follow the instructions (e.g., did not follow the required pace or produced additional utterances) or there were technical difficulties during their behavioral or EEG data acquisition. This number of participants is in line with previous studies of action monitoring targeting similar components using EEG and Laplacian transformation (Pinet et al., 2015; Roger et al., 2010; Vidal et al., 2000, 2003, 2011). All participants were native English speakers, had normal or corrected-to-normal vision and were right-handed. The study was performed in agreement with the Declaration of Helsinki. All subjects gave informed consent approved by the University of California, Berkeley Committee for Protection of Human Subjects. Participants received course credit or financial compensation for their participation.

2.2. Materials

The stimuli were 32 pairs of 4-word tongue-twisters adapted from Oppenheim & Dell, 2008, and similar as in Nozari & Dell, 2012. In all pairs, the word onset consonants were arranged in an ABBA pattern. In both tongue-twisters belonging to the same pair, the second and third words shared an onset different from the onset shared by the first and fourth words (e.g., "pub bust bun puff", “tub bust bun tough”). Onsets were manipulated to create phonologically “similar” and “dissimilar” conditions within each pair. In the “similar” tongue-twisters, onsets differed in only one phonetic feature (e.g., /p/ and /b/ are both bilabial consonants, but /b/ is voiced and /p/ is not). In the “dissimilar” tongue-twisters, onsets differed in two phonetic features (e.g., /t/ is alveolar and unvoiced, whereas /b/ is bilabial and voiced). Vowels were the same in all words within a tongue-twister. Two lists were created, each containing the same number of “similar” and “dissimilar” tongue-twisters, to break each pair into two tongue-twisters. Each list therefore contained only one of the two tongue-twisters in a pair, in order to avoid additional repetitions of the middle words common to the tongue-twisters within a pair. The purpose of the lists was to ensure that half of the subjects got one of the tongue twisters in the pair first (i.e., similar or dissimilar), and the other half the other tongue twisters in the same pair first.

2.3. Procedure

Participants were tested in a sound-attenuated dimly-lit environment. They were seated comfortably 148 cm from a computer screen on which the stimuli were displayed. The experiment was controlled by the Eprime 2.0 Professional software (Psychology Software Tools, Inc., Pittsburgh, PA), which allows on-line recording of the participants' verbal responses.

A trial consisted of the following events: (1) The stimulus (4 words) were presented on the screen. The participant was asked to read the words out loud at least 2 times but more if needed and then recite the words without looking at the screen. When the participant felt ready, they were asked to press the space bar to start the metronome; (2) With words still on the screen, a metronome started at a pace of one beep every 650 ms to allow for comfortable production of the word sequence and the averaging of EEG segments around each word onset. The participant was asked to recite the 4 words twice in a row following the metronome pace. (3) The words disappeared, the metronome stopped but the participant was instructed to continue reciting the 4 words twice in a row at the same pace. This was the critical test phase during which the behavioral and EEG data was collected. Importantly, no stimuli were present on the screen during this critical test phase. (4) A red dot appeared at the center of the screen signaling to the participant to stop speaking. (5) The red dot was replaced by a blank screen and the participant could start the next trial when ready by pressing the space bar (see Figure 1).

Please insert Figure 1 here

The participants were asked not to waste time on correcting their errors and try to stick to the pace. They were also encouraged to remain as relaxed as possible and to avoid making movements that could generate artifacts on the EEG (e.g. eye blinks, frowning) during the trials. The experimental session lasted for about an hour.

The EEG was recorded from 64 Ag/AgCl pre-amplified electrodes (BIOSEMI, Amsterdam) (10-20 system positions). The sampling rate was 1024 Hz (filters: DC to 208 Hz, 3 db/octave). The vertical electro-oculogram (EOG) was recorded by means of surface electrodes above and below the left eye. The horizontal EOG was recorded with two electrodes positioned over the two outer canthi. The passive reference was placed over the right mastoid.

2.4. Data processing and analysis

2.4.1. Behavioral data processing and analysis

All of the sound recordings were processed offline to mark vocal onsets and errors. Two native English speaker research assistants, blind to the hypotheses of the study, transcribed all utterances for errors. Where they disagreed, a third person listened to the recording, and the utterance was marked as an error if two out of the three people agreed that it was an error. Onset times were determined by viewing the acoustic waves in PRAAT 6.0.24 (Boersma & Weenink, 2013). A text grid in PRAAT was used to mark the onset of utterances, and the alignment of the text grid to actual onsets was checked manually for all trials and adjusted if necessary. All vocal onsets were marked from the onset of the blank screen in the test phase. Utterances were coded as errors if the participant produced any kind of verbal error including partial or complete production of incorrect words and verbal disfluencies (stuttering, utterance repairs, etc.).

We analyzed accuracy rates using logistic mixed effect models (Baayen, Davidson, & Bates, 2008; Jaeger, 2008) testing for fixed effects of Phonological Similarity (similar vs. dissimilar), Cycle (first vs. second iteration of the sequence during the test phase), and Word Position within each sequence (1 to 4), possible interactions between these factors, and controlled for random effects of participant.

2.4.2. EEG data processing and analysis

After acquisition, the EEG data were resampled at 256 Hz. We used Independent Component Analysis (ICA) as implemented in EEGLAB (Delorme & Makeig, 2004) to correct for vertical ocular movements, and a Blind Source Separation algorithm based on Canonical Correlation Analysis (BSS-CCA, De Clercq et al., 2006) to correct for muscular artefacts associated with speech production (as developed in De Vos et al., 2010). As in our previous studies, we ran BSS-CCA twice: first on non-overlapping 30 seconds-long time-window to reduce tonic EMG activity due to frowning or muscle fatigue, and second on non-overlapping 2 seconds-long time-windows to target the EMG bursts linked to articulation (Riès et al., 2011, 2013, 2015). Due to the low number of error trials, EEG analyses focused on correct trials, around the onset of each word. Trials with remaining artifacts were rejected by visual inspection (20.3±9.5% trials on average), leaving 185±21 correct trials per subject on average. Laplacian transform was used as a spatial filter to enhance the resolution of components (degree of spline: 3, Legendre polynomial: 15° maximum). We assumed a radius of 10 cm for the sphere representing the head. The resulting unit was µV/cm2.

We focused our analyses on components known to be reliably present in response-locked EEG analysis in speech production, and previously described at medial frontal (Cz and FCz) and left lateral electrodes (FC5) (e.g., Riès et al., 2013; Riès et al., 2011), and performed our analyses per electrode. We first tested whether the slope of the component (as fitted with a linear regression) differed from zero to demonstrate the existence of each component (Vidal et al., 2003). We then tested for a return to baseline by comparing the amplitude of the event-related potentials between a pre-vocal onset time-window (spanning from -300 to -200 ms before response onset for the ERN at FCz and Cz, and -150 to -50 ms for the left fronto-lateral component at FC5) and a post-vocal onset time-window (spanning from 300 to 400 ms after the response onset for the ERN at FCz and Cz, and 500 to 600 ms for the left fronto-lateral component at FC5). The choice of the pre-vocal onset and post-vocal onset time-windows was chosen based on the morphology of the components of interest on the grand averages. We then tested for effects of the experimental variables, Cycle, Word Position, and Phonological Similarity on this return to baseline. In addition, we tested for effects of the experimental variables on the amplitude of the event-related potentials on 100-ms long time-windows around the peak of the components, as commonly performed in ERP studies (e.g., Acheson & Hagoort, 2014). Comparisons were performed using non-parametric statistics (Wilcoxon signed rank test between paired samples). If there were over 15 observations, a z value was estimated and reported. Otherwise the sum of ranks W is reported.

3. Results

3.1. Behavioral results

To test assumptions regarding our experimental manipulations, we analyzed the behavioral results by focusing on error rates since reaction times are not available in this type of paradigm. We collected 268 speech errors (146 in similar, and 122 in dissimilar) over 112 trials across all participants. The average error rate was 8.72 % (SD = 4.27 %). There was an effect of Cycle on accuracy rates (βraw = 1.979, SE= 0.414, Wald Z = 4.78, p<.001): participants made more errors in the second compared to the first iteration of the sequence during the test phase. There was also an effect of Word Position within each sequence on the linear (βraw = 1.944, SE= 0.384, Wald Z = 5.07, p<.001) and quadratic term (βraw = -0.304, SE= 0.068, Wald Z = -4.45, p<.001): error rates tended to increase from the beginning towards the end of the sequence, but this increase was not monotonic, and had its peak on the third word in both cycles. There was also an interaction between Cycle and Word Position (βraw = -0.435, SE= 0.143, Wald Z = -3.05, p=.002): the effect of Word Position on error rates was smaller on the second than the first cycle. Finally, although there were more errors in the similar condition, the difference did not reach significance on accuracy rates (βraw = 0.187, SD= 0.131, Wald Z = 1.43, p=.153) (see Figure 2).

Please insert Figure 2 here

3.2. EEG results

<u>Medial frontal components</u>

The electrophysiological data revealed a negative component over fronto-central electrodes (FCz and Cz), peaking 80ms after the onset of each word, similar to an ERN (Figure 3). The slope of the waveforms were significantly different from zero between -100 ms and vocal onset at FCz (W = 5, p=.0049) and Cz (W = 0, p<.001). Mean amplitudes were not significantly different before and after the onset of each word in the analysis made on averages over all trials per participant, indicating a return to baseline (At FCz: z = -1.41, p = 0.16; At Cz: z= 0.86, p= 0.39).

Please insert Figure 3 here

We also examined whether this return to baseline and the amplitude of the peaks were influenced by the experimental variables under study (Figure 4).

Please insert Figure 4 here

There was an effect of Cycle on the peak amplitude and on the return to baseline analyses at Cz (see Figure 4 top row middle column). The amplitude of the peak was higher for cycle 2 than cycle 1 between 0 and 50 ms after response-onset (z= 2.35, p = 0.019), and the amplitude in the post-response time-window (between 300 and 400 ms) was higher than in the pre-response time-window (-300 to -200 ms) in cycle 2 (z=2.2, p = 0.028). This was not the case at cycle 1 (z= -0.63, p = 0.53). None of the comparisons at FCz reached significance: there was no difference in amplitude between the pre- and post-response time-windows at cycle 1 (z = -0.63, p = 0.53) or cycle 2 (z = -1.18, p = 0.24), and no effect of Cycle on the peak amplitude (z = 0.235, p = 0.814).

There was no clear effect of Word Position within Cycle. Most comparisons were non-significant, except for a marginal effect at position 3 for Cz (z= 1.88, p = 0.06) where the amplitude in the post-response time-window (between 300 and 400 ms) was higher than in the pre-response time-window (-300 to -200 ms) (see Figure 4 middle row middle column). None of the comparisons reached significance on the peak amplitude measurements (all z-values were comprised between 0.3 and -1.02).

There was an effect of Phonological Similarity on the peak amplitude and on the return to baseline analyses at FCz (see Figure 4 bottom row left column). The amplitude of the peak was higher for similar than dissimilar sequences on a 0 to 50 ms time-window following response onset (z = 2.04, p = 0.041). In addition, the amplitude was lower in the post-component time-window compared to the pre-component time-window for dissimilar sequences (z = -2.04, p = 0.041). This was not the case for similar sequences (FCz: z = 0.39, p = 0.69). At Cz, there was no effect of Phonological Similarity on the peak amplitude (z = -0.47, p = 0.64) or on the peak amplitude in dissimilar (z = -0.24, p = 0.81) or similar sequences Cz : z = 1.26, p= 0.81).

Late Left Anterior Negativity (LLAN)

A clear left lateral negativity was observed maximal at FC5, the slope was significantly different from zero between vocal onset and +100 ms (W = 13, p=.0425). This left lateral component showed no significant difference in amplitude between the time-windows before and after the baseline in the analysis made on averages over all trials per participant (-150 :-50ms / 500 :600ms : z = 1.0,p = 0.31) (Figure 5).

Please insert Figure 5 here

As can be seen on Figure 4 (right column), neither Cycle nor Phonological Similarity had an effect on the return to baseline for FC5 (cycle 1: z = 1.18, p = 0.24; cycle 2: z = 1.02, p = 0.31; Dissimilar: z = 1.65, p = 0.10 ; Similar : z = 1.26, p = 0.21), nor on the peak amplitude (Cycle: z = 1.10, p = .27; Phonological Similarity: z = -1.41, p = .16). Word Position, on the other hand, had an effect on the return to baseline and on the peak amplitude. For the first and second position, there was a significant difference in amplitude in the time-windows taken before and after the component (first: z = 2.35, p = 0.019; second: z = 2.43, p = 0.015), but not for the third and fourth position in the sequence (third: z = -1.8, p = 0.07; fourth: z = -0.55,p = 0.58). For the peak amplitude, there was a significant difference between positions 1 and 3 (z = -2.2749, p = .023), and 1 and 4 (z = -2.7456, p = .0060) due to the fact peak amplitude was lower in positions 3 and 4 compared to position 1, but there was no difference between positions 1 and 2 (z = -1.2551, p = .21).

Please insert Figure 6 here

Figure 6 shows the activity at FCz, Cz, and FC5 over the whole sequence produced (2 cycles of 4 words each). It shows that, even if there is no difference between pre- and post-component time-windows on average, the left frontal activity is sustained across the whole sequence, which explains the effect of Word Position in the cycle. More precisely, the left frontal activity seems to follow each cycle, and monotonically increase over the course of one cycle. Even though the EEG activity at FCz and Cz does not show such a large increase across cycle as compared to the EEG activity at FC5, a sustained increase across cycle 2 is visible at Cz.

Simulation results

To assess how slow-wave or sustained activity impacts peak amplitude, we ran simulations with a sequence of four EEG components (see Figure 7) and examined the average of the components. If there is no sustained activity (Figure 7A), the amplitude of the average component is equal to the average of the amplitude of each component. However, if there is a sustained positive activity (Figure 7B, we chose a slope of 1/2000), the amplitude of the average component is higher than the average of each individual component (in our example, the amplitude increased from 1 to 1.14), and the post-component baseline is higher than pre-amplitude baseline. This is what we observed in the manipulation of Cycle at Cz, and demonstrates that increasing slow-wave or sustained activity can contribute to increased peak amplitude. The reversed situation (lower amplitude and decrease in baseline) is observed for a sustained negative activity (Figure 7C). This is what we observed in the manipulation of Phonological Similarity at FCz, and demonstrates that decreasing slow-wave or sustained activity can contribute to decreased peak amplitude.

Please insert Figure 7 here

Discussion

We characterized ERPs associated with vocal response preparation and monitoring during the execution of multi-word utterances in the context of a paced tongue twister paradigm. One might question whether reciting tongue-twisters on a metronome pace is representative of how people produce speech. There is little doubt that everyday production entails conceptual processing, and speakers do not simply recite memorized words after being trained on a metronome beat. That said, the task does tap into some key operations involved in speech production, such as lexical-to-phonological mapping (i.e., phonological encoding) and articulation. The dynamics of the process is also similar to planning everyday speech: analysis of speech errors elicited from tongue-twister have shown both anticipation and perseveration errors (e.g., Nozari et al., 2012). These errors, which are also observed in everyday speech, indicate proactive planning of future words, and occasional failures of suppressing already-produced words, respectively. Finally, tongue-twisters exaggerate phonological competition effects that are present in other tasks such as picture naming (e.g., Nozari et al., 2016), and more generally in the lexicon (Sadat et al., 2014). As such, tongue-twisters are well-suited for capture the processes involved in mapping lexical items to their sounds and ultimately to articulatory-phonetic features.

Our results show that both the ERN (observed at FCz and Cz), and the LLAN peaking after vocal onset (observed at FC5) can be reliably detected in multi-word utterances in this paradigm. There was no difference in amplitude between pre- and post-component time-windows in the analyses performed on the averages over all trials per participant, but we found effects of Cycle, Word Position within the sequence, and Phonological Similarity on this return to baseline. These results indicate that, in the context of multi-word utterances, the ERP signatures of utterance planning and monitoring are combinations of transient EEG activity and sustained EEG activity that develop over time in the word sequence. The implications of these findings are discussed below.

The ERN in multi-word utterances

The ERN is a fronto-central component maximal at electrodes Cz and FCz, which is sensitive to increased task difficulty. This component has been studied in numerous studies of speech monitoring (e.g., Acheson & Hagoort, 2014; Ganushchak & Schiller, 2006, 2008; Masaki et al., 2001; Pinet & Nozari, 2020; Riès et al., 2011), and action monitoring in general (e.g., Falkenstein et al., 1991; Gehring & Knight, 2000; Roger et al., 2010; Vidal et al., 2000, 2003). As outlined in the introduction, multi-word production requires efficient, fast, and adaptable monitoring processes as speakers are more likely to make errors when producing word sequences compared to an isolated word. The involvement of these highly efficient monitoring processes has been evidenced through the increased proportion of corrected errors as a function of error probability on error-prone words throughout the sentence (Nozari et al., 2019). The ERN provides a useful electrophysiological index to study how monitoring processes are engaged throughout sentence production. However, in order to make best use of this index, the impact of producing multiple words in a sequence must be closely examined. To this end, we asked participants to produce words at a specified pace after training them with a metronome so that each word would be 650 ms apart from the other words in the sequence. This avoided having another word produced during the time-window of analysis and allowed for the examination of ERP components time-locked to each word. While this does not exactly mimic the varying rhythm of speech, it is the first step towards analyzing multi-word utterances.

In our study, we were able to identify clear negative components at FCz and Cz, where the ERN is typically observed. The slopes of this component were significantly different from zero in the 100 ms preceding vocal onset and the component peaked within 100 ms post-vocal onset, in agreement with previous reports (Pinet & Nozari, 2020; Riès et al., 2011, 2013). The fact that we could observe the ERN clearly in the context of multi-word utterances is in itself noteworthy as the words were produced within 650 ms from one another, which is relatively fast-paced even if slower than normal speech production (during which we produce 2 to 3 words per second in connected speech, Levelt et al., 1999). This indicates that EMG artifacts could be successfully addressed with the blind source separation technique we used, BSSCCA, even in the context of multi-word utterances. We also examined whether or not there was a clear return to baseline after each ERN. We found that there was no statistical difference in amplitude when we compared the pre- and post-component time-windows over all trials, suggesting that the ERN locked to each word can indeed be studied as a separable component. Finally, we examined whether or not the effect of experimental manipulations likely to affect the amplitude of the ERN could be studied in the context of multi-word utterances. We found effects of Cycle and Phonological Similarity on the return to baseline as well as on the amplitude of the ERN.

An effect of Cycle was observed at Cz: In the second cycle, the amplitude in the post-component time-window was higher than in the pre-component time-window. In addition, the amplitude around the peak of the component was higher in the second cycle than in the first cycle. This parallels our behavioral findings, where we saw increased error rates in the second cycle compared to the first cycle.

In particular, in more difficult and error-prone situations, while the amplitude of the ERN in errors has been shown to decrease, its amplitude in correct trials (as analyzed here) has been shown to increase (e.g., Acheson & Hagoort, 2014; Allain et al., 2004; Riès et al., 2011). Observing the ERN in the context of multi-word sequences provides a tool to ask questions related to speech monitoring beyond single word production.

ERN amplitude increase in the pre versus post-component time-windows in cycle 2 indicates that the brain activity underlying this component is not completely transient and is superimposed on a baseline shift. A dissociation between sustained versus transient cognitive control mechanisms has been linked to bottom-up versus top-down or reactive versus proactive control mechanisms respectively (e.g., Chen et al., 2012; West et al., 2012; for parallel investigations using fMRI, see Braver, 2012; Reynolds et al., 2009). The ERN has usually been interpreted as reflecting a performance monitoring mechanism assumed to operate on a trial-by-trial basis (e.g., Riès et al., 2011; Vidal et al., 2000, 2003). However, investigations of response-locked medial frontal theta activity, which has been related to the ERN, have suggested that the neuronal activity underlying the ERN may be associated with a long-lasting increase in theta power rather than with a transient burst of neural activity (Yeung et al., 2007, for similar simulation-based results see Holroyd & Coles, 2002). Our results suggest that, at least in the case of producing multiple words relatively close together in a sequence, sustained neuronal activity may contribute to the ERN in addition to transient activity time-locked to each word. This could be interpreted as a general increase of proactive control or top-down activity when the task gets harder (in cycle 2 compared to cycle 1), in addition to an increase in reactive control time-locked to each word. We note that there was no significant effect of Word Position within the Cycle on the amplitude of the ERN or on its return to baseline, suggesting the general increase in activity in cycle 2 versus cycle 1 is relatively slow. This differs from what we observed for the left lateral activity as we describe below.

One question is how effects of experimental manipulations on the baseline measures may affect component amplitudes. Peak amplitude measurements for a component of interest are generally measured relative to a pre-defined baseline (Luck, 2014). In our analysis of the effect of Cycle on ERN amplitude, we employed two different baselines: one on the average of the trials in cycle 1 for the measurement of the amplitude of the ERN in cycle 1, and one on the average of the trials in cycle 2 for the measurement of the amplitude of the ERN in cycle 2. Therefore, the fact that the net amplitude of the EEG signal may be larger in cycle 2 than in cycle 1 cannot explain the difference in ERN amplitude in cycle 2 compared to cycle 1.

However, it is unclear how much of an influence the slow increase of activity in cycle 2 may have on the shape of the ERN for each word. It is possible that the upward slope of the sustained EEG activity throughout cycle 2 (as seen in Figure 6) could have increased ERN amplitude. Since this sustained upward slope was not visible in cycle 1, the difference in amplitude between the ERN in cycle 1 versus cycle 2 may be explained by both an increase in transient and/or sustained EEG activity, as suggested by our simulation results.

An effect of Phonological Similarity was also found on the baseline and peak amplitude of the ERN at FCz with higher peak amplitude in similar than dissimilar sequences. We note that we did not observe a significant effect of Phonological Similarity on accuracy rates, which is likely due to the lower number of participants in our study compared to the study by Oppenheim & Dell (2008), for example. Indeed, this effect is expected as performance in phonologically similar sequences is typically lower than in dissimilar sequences (Nooteboom, 2005; Oppenheim & Dell, 2008). However, although there was no difference in amplitude between the time-windows taken before and after the peak in phonologically similar sequences, this was not the case for phonologically dissimilar sequences. In dissimilar sequences, the amplitude was lower in the post-component compared to the pre-component time-window. This effect may have been due to the fact that the positive component following the ERN was larger in dissimilar than in similar sequences. In errors, a positive component following the ERN, the error positivity (Pe), has been linked to conscious error detection in previous linguistic (e.g., Pinet & Nozari, 2019) and non-linguistic studies (e.g., Nieuwenhuis et al., 2001; Overbeek et al., 2005). However, here we focused on correct trials and no Pe is usually observed in correct trials. Therefore, it is unclear at this point which cognitive process the positivity following the ERN may reflect in correct trials. Importantly, the difference in baseline amplitude that we observed may be linked to the fact that the words were produced in a sequence and only separated from one another by around 650 ms (i.e., the pace of the metronome beeps that participants were trained on). There was perhaps not enough time for the EEG signal to return to baseline after this positivity occurred (i.e., between 300 and 400 ms following vocal onset) in dissimilar sequences. We note that this potential problem would only be increased if words were produced at their normal pace of about 2 to 3 words per second in connected speech (Levelt et al., 1999). This baseline shift may also indicate a general decrease in sustained activity in phonologically dissimilar versus similar sequences. As for the effect of Cycle, this may be impacting the effect of Phonological Similarity on the amplitude of the ERN in our paradigm if the shape of the ERN in dissimilar sequences is affected by the general downward slope of the EEG activity throughout the sequence, as suggested by our simulation results.

Our results illustrate the importance of taking into account effects on the return to baseline for studying the ERN in the context of multi-word utterances. In particular, they suggest that a general increase or decrease in sustained EEG activity as a function of task difficulty may influence the shape of the ERN on a word-by-word basis. How much of the ERN's amplitude is explained by underlying transient versus sustained EEG activity has been the focus of previous research outside of the language domain (Luu et al., 2004).

Left lateral frontal component

We observed a left lateral frontal component, onsetting shortly before vocal onset and peaking around 300 ms post-vocal onset, which we called the LLAN. The slope of this component was significantly different from zero on the 100 ms-time-window following vocal onset and there was no difference in amplitude between the pre- and post-component time-windows over all trials, suggesting a return to baseline. This return to baseline and the peak amplitude was not affected by Cycle or Phonological Similarity. However, the return to baseline was affected by Word Position in the sequence, as was the peak amplitude. In positions 1 and 2, the amplitude of the post-component time-window was higher than that of the pre-component time-window. This was not the case for positions 3 and 4 and the amplitude of the peak was lower in positions 3 and 4 compared to 1 and 2. As can be seen on Figure 6, this was due to the fact the activity started to rise at the beginning of each cycle in the sequence over positions 1 and 2, and then stayed sustained and seemed to reach a plateau at positions 3 and 4, before decreasing at the end of the cycle. A left lateral frontal component has been previously described at FC5 in picture naming using Laplacian transformation (Riès et al., 2013), although its time-course was different. In this previous study, the LLAN started to rise at about 350 ms post stimulus presentation and peaked right around vocal onset. Because of this time-course and its source localization (obtained through both dipole and surface minimum norm modelling), we associated it with phonological encoding and response programming and execution. In the present study, this activity was more sustained as it started to rise around 100 ms before the onset of the first word in the sequence, then increased throughout the first half of each cycle and then plateaued until the end of the sequence. This sustained activity could be compatible with a response planning interpretation given that speakers tend to plan the words they are going to say early on in a word sequence, as has been argued in the case of sentence production (e.g., Ferreira & Swets, 2002; Ford & Holmes, 1978; Konopka, 2012). How much speakers plan ahead of time in sentence production is under strategic control (Ferreira & Swets, 2002), and stimulation of the left frontal cortex through anodal transcranial direct current stimulation (tDCS) decreases errors and shortens planning time during sentence production (Arnold & Nozari, 2017; Nozari et al., 2014). Although the utterance produced in our study were not sentences, such planning mechanisms may be at play in the case of multi-word utterances and the localization of this EEG component to the left frontal cortex through Laplacian transformation is compatible with this interpretation. In addition, planning is one of the functions which is commonly affected in patients with prefrontal lesions (for a review, see Szczepanski & Knight, 2014). We note however that we did not test utterance planning in our study, and that further studies would be needed to confirm this interpretation.

In conclusion, our results show that extraction of ERP components time-locked to vocal onset is possible in word sequence production. We isolated medial frontal activity related to speech monitoring time-locked to each word, as well as a slower left frontal component potentially associated with word planning. One caveat is that baseline amplitudes may shift across the sequence and the impact of this additional sustained EEG activity must be considered to understand the effects of experimental manipulations on the peak amplitudes of the ERPs of interest in the context of word sequence production.

Acknowledgements

This research was supported by an NIDCD grant 1R21DC016985 to S.K.R., and an NINDS grant 2R37NS21135 to R.T.K. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We are very thankful to Elizabeth Johnson and research assistants for their help with EEG recordings and data processing, and to the participants who took part in this study.

References

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

Allain, S., Carbonnell, L., Falkenstein, M., Burle, B., & Vidal, F. (2004). The modulation of the Ne-like wave on correct responses foreshadows errors. *Neuroscience Letters*, *372*(1–2), 161–166. https://doi.org/10.1016/j.neulet.2004.09.036

Arnold, J. E., & Nozari, N. (2017). The effects of utterance timing and stimulation of left prefrontal cortex on the production of referential expressions. *Cognition*, *160*, 127–144. https://doi.org/10.1016/j.cognition.2016.12.008

Bonini, F., Burle, B., Liégeois-Chauvel, C., Régis, J., Chauvel, P., & Vidal, F. (2014). Action monitoring and medial frontal cortex: leading role of supplementary motor area. *Science (New York, N.Y.)*, *343*(6173), 888–891. https://doi.org/10.1126/science.1247412

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113. https://doi.org/10.1016/j.tics.2011.12.010

Breining, B., Nozari, N., & Rapp, B. (2016). Does segmental overlap help or hurt? Evidence from blocked cyclic naming in spoken and written production. *Psychonomic Bulletin & Review*, *23*(2), 500–506. https://doi.org/10.3758/s13423-015-0900-x

Burle, B., Roger, C., Allain, S., Vidal, F., & Hasbroucq, T. (2008). Error negativity does not reflect conflict: a reappraisal of conflict monitoring and anterior cingulate cortex activity. *Journal of Cognitive Neuroscience*, *20*(9), 1637–1655. https://doi.org/10.1162/jocn.2008.20110

Chen, C.-C., Kiebel, S. J., Kilner, J. M., Ward, N. S., Stephan, K. E., Wang, W.-J., & Friston, K. J. (2012). A dynamic causal model for evoked and induced responses. *NeuroImage*, *59*(1), 340–348. https://doi.org/10.1016/j.neuroimage.2011.07.066

De Clercq, W., Vergult, A., Vanrumste, B., Van Paesschen, W., & Van Huffel, S. (2006). Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram. *IEEE Transactions on Bio-Medical Engineering*, *53*(12 Pt 1), 2583–2587. https://doi.org/10.1109/TBME.2006.879459

De Vos, M., Vos, D. M., Riès, S., Vanderperren, K., Vanrumste, B., Alario, F.-X., Van Huffel, S., Huffel, V. S., & Burle, B. (2010). Removal of muscle artifacts from EEG recordings of spoken language production. *Neuroinformatics*, *8*(2), 135–150. https://doi.org/10.1007/s12021-010-9071-0

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

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, *104*(1), 123–147. https://doi.org/10.1037/0033-295X.104.1.123

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior*, *20*(6), 611–629. https://doi.org/10.1016/S0022-5371(81)90202-4

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

Edwards, E., Nagarajan, S. S., Dalal, S. S., Canolty, R. T., Kirsch, H. E., et al. (2010) Spatiotemporal imaging of cortical activation during verb generation and picture naming. *NeuroImage, 50*, 291–301.

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

Ferreira, F., & Swets, B. (2002). How Incremental Is Language Production? Evidence from the Production of Utterances Requiring the Computation of Arithmetic Sums. *Journal of Memory and Language*, *46*(1), 57–84. https://doi.org/10.1006/jmla.2001.2797

Ford, M., & Holmes, V. M. (1978). Planning units and syntax in sentence production. *Cognition*, *6*(1), 35–53. https://doi.org/10.1016/0010-0277(78)90008-2

Ganushchak, L. Y., Christoffels, I. K., & Schiller, N. O. (2011). The use of electroencephalography in language production research: a review. *Frontiers in Psychology*, *2*, 208. https://doi.org/10.3389/fpsyg.2011.00208

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

Ganushchak, L. Y., & Schiller, N. O. (2008). 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

Gehring, W. J., & Knight, R. T. (2000). Prefrontal-cingulate interactions in action monitoring. *Nature Neuroscience*, *3*(5), 516–520. https://doi.org/10.1038/74899

Habets, B., Jansma, B. M., & Münte, T. F. (2008). Neurophysiological correlates of linearization in language production. *BMC Neuroscience*, *9*, 77. https://doi.org/10.1186/1471-2202-9-77

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

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679–709. https://doi.org/10.1037/0033-295X.109.4.679

Hultén, A., Vihla, M., Laine, M., Salmelin, R. (2009) Accessing Newly Learned Names and Meanings in the Native Language. *Human Brain Mapping, 30*, 976–989.

Klimesch, W., Doppelmayr, M., Wimmer, H., Gruber, W., Röhm, D., Schwaiger, J., & Hutzler, F. (2001) Alpha and beta band power changes in normal and dyslexic children. *Clinical Neurophysiology, 112*, 1186–1195. https://doi.org/10.1016/S1388-2457(01)00543-0

Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language*, *66*(1), 143–162. https://doi.org/10.1016/j.jml.2011.08.003

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Kutas, Marta, & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Laaksonen, H., Kujala, J., Hultén, A., Liljeström, M., & Salmelin, R. (2012). MEG evoked responses and rhythmic activity provide spatiotemporally complementary measures of neural activity in language production. *NeuroImage*, *60*(1), 29–36. https://doi.org/10.1016/j.neuroimage.2011.11.087

Levelt, W. J. M. (1993). *Speaking: From Intention to Articulation*. MIT Press.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *The Behavioral and Brain Sciences*, *22*(1), 1-38; discussion 38-75.

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique, Second Edition* (2nd edition). MIT Press. https://mitpress.mit.edu/books/introduction-event-related-potential-technique-second-edition

Luu, P., Tucker, D. M., & Makeig, S. (2004). Frontal midline theta and the error-related negativity: neurophysiological mechanisms of action regulation. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *115*(8), 1821–1835. https://doi.org/10.1016/j.clinph.2004.03.031

Martin, R. C., Crowther, J. E., Knight, M., Tamborello, F. P., & Yang, C.-L. (2010). Planning in sentence production: Evidence for the phrase as a default planning scope. *Cognition*, *116*(2), 177–192. https://doi.org/10.1016/j.cognition.2010.04.010

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

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P. H., & 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. https://doi.org/10.1111/1469-8986.3850752

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

Nozari, N. (2018). Chapter Seven - How Special Is Language Production? Perspectives From Monitoring and Control. In K. D. Federmeier & D. G. Watson (Eds.), *Psychology of Learning and Motivation* (Vol. 68, pp. 179–213). Academic Press. https://doi.org/10.1016/bs.plm.2018.08.006

Nozari, N., Arnold, J. E., & Thompson-Schill, S. L. (2014a). The Effects of Anodal Stimulation of the Left Prefrontal Cortex on Sentence Production. *Brain Stimulation*, *7*(6), 784–792. https://doi.org/10.1016/j.brs.2014.07.035

Nozari, N., Arnold, J. E., & Thompson-Schill, S. L. (2014b). The Effects of Anodal Stimulation of the Left Prefrontal Cortex on Sentence Production. *Brain Stimulation*, *7*(6), 784–792. https://doi.org/10.1016/j.brs.2014.07.035

Nozari, N., & Dell, G. S. (2012). Feature migration in time: reflection of selective attention on speech errors. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *38*(4), 1084–1090. https://doi.org/10.1037/a0026933

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

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

Nozari, N., Martin, C. D., & McCloskey, N. (2019). Is repairing speech errors an automatic or a controlled process? Insights from the relationship between error and repair probabilities in English and Spanish. *Language, Cognition and Neuroscience*, *34*(9), 1230–1245. https://doi.org/10.1080/23273798.2019.1637007

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

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

Overbeek, T. J. M., Nieuwenhuis, S., & Ridderinkhof, K. R. (2005). Dissociable Components of Error Processing. *Journal of Psychophysiology*, *19*(4), 319–329. https://doi.org/10.1027/0269-8803.19.4.319

Piai, V., Riès, S. K., & Knight, R. T. (2014). The electrophysiology of language production: what could be improved. *Frontiers in Psychology*, *5*, 1560. https://doi.org/10.3389/fpsyg.2014.01560

Piai, V., Roelofs, A., Jensen, O., Schoffelen, J.-M., & Bonnefond, M. (2014). Distinct patterns of brain activity characterise lexical activation and competition in spoken word production. *PloS One*, *9*(2), e88674. https://doi.org/10.1371/journal.pone.0088674

Piai, V., Roelofs, A., Rommers, J., & Maris, E. (2015). Beta oscillations reflect memory and motor aspects of spoken word production. *Human Brain Mapping*, *36*(7), 2767–2780. https://doi.org/10.1002/hbm.22806

Piai, V., Rommers, J., & Knight, R. T. (2017). Lesion evidence for a critical role of left posterior but not frontal areas in alpha-beta power decreases during context-driven word production. *The European Journal of Neuroscience*. https://doi.org/10.1111/ejn.13695

Pinet, S., Hamamé, C. M., Longcamp, M., Vidal, F., and Alario, F.-X. (2015) Response planning in word typing: Evidence for inhibition. *Psychophysiology, 52*, 524-531. https://doi.org/10.1111/psyp.12373

Pinet, S., & Nozari, N. (2020). Electrophysiological Correlates of Monitoring in Typing with and without Visual Feedback. Journal of Cognitive Neuroscience, 32(4), 603–620. http://doi.org/10.1162/jocn_a_01500

Reynolds, J. R., West, R., & Braver, T. (2009). Distinct Neural Circuits Support Transient and Sustained Processes in Prospective Memory and Working Memory. *Cerebral Cortex*, *19*(5), 1208–1221. https://doi.org/10.1093/cercor/bhn164

Riès, S. K., Dhillon, R. K., Clarke, A., King Stephens, D., Laxer, K. D., Weber, P. B., et al. (2017). Spatiotemporal dynamics of word retrieval in speech production revealed by cortical high-frequency band activity. *Proceedings of the National Academy of Sciences of the United States of America, 114(23)*, E4530–E4538. https://doi.org/10.1073/pnas.1620669114.

Riès, S., Janssen, N., Burle, B., & Alario, F.-X. (2013). Response-locked brain dynamics of word production. *PloS One*, *8*(3), e58197. https://doi.org/10.1371/journal.pone.0058197

Riès, S., 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

Riès, S. K., Fraser, D., McMahon, K. L., & de Zubicaray, G. I. (2015). Early and Late Electrophysiological Effects of Distractor Frequency in Picture Naming: Reconciling Input and Output Accounts. *Journal of Cognitive Neuroscience*, *27*(10), 1936–1947. https://doi.org/10.1162/jocn_a_00831

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

Riès, S. K., Nadalet, L., Mickelsen, S., Mott, M., Midgley, K. J., Holcomb, P. J., & Emmorey, K. (2020). Preoutput Language Monitoring in Sign Production. *Journal of Cognitive Neuroscience*, 1–14. https://doi.org/10.1162/jocn_a_01542

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

Roger, C., Bénar, C. G., Vidal, F., Hasbroucq, T., & Burle, B. (2010). Rostral Cingulate Zone and correct response monitoring: ICA and source localization evidences for the unicity of correct- and error-negativities. *NeuroImage*, *51*(1), 391–403. https://doi.org/10.1016/j.neuroimage.2010.02.005

Sadat, J., Martin, C. D., Costa, A., & Alario, F. X. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. Cognitive psychology, 68, 33-58. doi: 10.1016/j.cogpsych.2013.10.001.

Salmelin, R., Hari, R., Lounasmaa, O.V., Sams, M. (1994) Dynamics of brain activation during picture naming. *Nature, 368*, 463–465.

Schwartz, M. F., & Hodgson, C. (2002). A new multiword naming deficit: Evidence and interpretation. *Cognitive Neuropsychology*, *19*(3), 263–288. https://doi.org/10.1080/02643290143000187

Sörös, P., Cornelissen, K., Laine, M., Salmelin, R. (2003) Naming actions and objects: cortical dynamics in healthy adults and in an anomic patient with a dissociation in action/object naming. *NeuroImage, 19*, 1787–1801.

Szczepanski, S. M., & Knight, R. T. (2014). Insights into Human Behavior from Lesions to the Prefrontal Cortex. *Neuron*, *83*(5), 1002–1018. https://doi.org/10.1016/j.neuron.2014.08.011

Vidal, F., Burle, B., Bonnet, M., Grapperon, J., & Hasbroucq, T. (2003). Error negativity on correct trials: a reexamination of available data. *Biological Psychology*, *64*(3), 265–282. https://doi.org/10.1016/s0301-0511(03)00097-8

Vidal, F., Burle, B., Grapperon, J., Hasbroucq, T. (2011). An ERP study of cognitive architecture and the insertion of mental processes: Donders revisited. *Psychophysiology 48(9),* 1242–1251.

Vidal, F., Hasbroucq, T., Grapperon, J., & Bonnet, M. (2000). Is the “error negativity” specific to errors? *Biological Psychology*, *51*(2–3), 109–128. https://doi.org/10.1016/S0301-0511(99)00032-0

Vihla, M., Laine, M., Salmelin, R. (2006). Cortical dynamics of visual/semantic vs. phonological analysis in picture confrontation. *NeuroImage, 33*, 732–738.

West, R., Bailey, K., Tiernan, B. N., Boonsuk, W., & Gilbert, S. (2012). The temporal dynamics of medial and lateral frontal neural activity related to proactive cognitive control. *Neuropsychologia*, *50*(14), 3450–3460. https://doi.org/10.1016/j.neuropsychologia.2012.10.011

Wlotko, E. W., & Federmeier, K. D. (2012a). Age-related changes in the impact of contextual strength on multiple aspects of sentence comprehension. *Psychophysiology, 49(6),* 770–785. doi:10.1111/j.1469-8986.2012.01366.x.

Wlotko, E. W., & Federmeier, K. D. (2012b). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *Neuroimage, 62(1)*, 356-366. doi:10.1016/j.neuroimage.2012.04.054.

Yeung, N., Bogacz, R., Holroyd, C. B., Nieuwenhuis, S., & Cohen, J. D. (2007). Theta phase resetting and the error-related negativity. *Psychophysiology*, *44*(1), 39–49. https://doi.org/10.1111/j.1469-8986.2006.00482.x

Figure captions:

Figure 1: Tongue twister paradigm. During the practice phase (on the left), the participants see the word sequence on the screen and practice iterating the sequence at the pace indicated by the metronome (once every 650 ms). During the test phase (on the right), the words disappear from the screen and the metronome stops. The participants are asked to keep on iterating the words at the same pace as when they were practicing for two cycles.

Figure 2: Error rate by Word Position in cycle 1 (i.e., positions 1 to 4) and 2 (positions 5 to 8) for similar and dissimilar word sequences with standard error bars.

Figure 3: Laplacian-transformed event-related potentials time-locked to vocal onset at fronto-medial sites FCz (left) and Cz (middle), and associated topography at the peak of the ERN (between 0-100 ms post-vocal onset) (right). Grayed time-windows indicate where the baseline measures were taken for the return-to-baseline analyses. On the grand averages (shown here), there was no difference in amplitude between the pre- and post-component time-windows.

Figure 4: Laplacian-transformed event-related potentials per condition time-locked to vocal onset at fronto-medial sites FCz (left column) and Cz (middle column), and left frontal site FC5 (right column). ERPs are plotted on the top row by Cycle (blue: cycle 1, red: cycle 2), in the middle row by Word Position within Cycle (blue: position 1, red: position 2, yellow: position 3, and purple: position 4), and in the bottom row by Phonological Similarity (red: similar, blue: dissimilar). Grayed time-windows indicate where the baseline measures were taken for the return-to-baseline analyses. Our analyses showed an effect of Cycle on the peak amplitude and on the return to baseline at Cz (top row middle column), an effect of Word Position on the peak amplitude and on the return to baseline at FC5 (middle row, right column), and an effect of Phonological Similarity on the peak amplitude and on the return to baseline at FCz (bottom row, left column).

Figure 5: Laplacian-transformed event-related potentials time-locked to vocal onset at left frontal site FC5 (blue) and right frontal site FC6 (red) (left), and associated topography at the peak of the component at FC5 (between 300-400 ms post-vocal onset) (right). Grayed time-windows indicate where the baseline measures were taken for the return-to-baseline analyses. On the grand averages (shown here), there was no difference in amplitude between the pre- and post-component time-windows.

Figure 6: Laplacian-transformed event-related potentials time-locked to vocal onset of the first word of the whole sequence at left frontal site FC5 (blue), and midline sites (FCz in red, Cz in yellow). A low-pass filter of 150 Hz was used for display. Vertical lines mark the approximate timing of each word, based on the timing of the metronome presented during training (every 650ms). Lighter vertical lines signal the beginning of each cycle.

Figure 7: Impact of sustained activity on simulated transient recordings. Top panels represent a sequence of four components (amplitude fixed at 1 A.U.), bottom panels represent the average of all four components. (A) No sustained activity. (B) Sustained positive activity. (C) Sustained negative activity.