Neural dynamics of speech and non-speech motor planning

M. Lancheros*, A-L. Jouen, M. Laganaro

Faculty of Psychology and Educational Science, University of Geneva, Switzerland

ABSTRACT

As the speech apparatus is also involved in producing non-speech movements, understanding whether speech and non-speech planning are controlled by the same brain mechanisms is central to the comprehension of motor speech planning. A crucial issue is whether a specialized motor planning/control system is dedicated to speech or if the motor planning/control system is shared across oromotor behaviors. We investigated the EEG/ERP spatio-temporal dynamics of the motor planning processes preceding articulation by comparing the production of non-speech gestures matched to monosyllabic words and non-words. To isolate motor from pre-motor planning, we used a delayed production task combined with an articulatory suppression task. Results suggest that the planning processes preceding articulation for non-speech, words and non-words recruit the same neural networks but with different involvement for non-speech and speech. The results support the idea of shared motor planning/control systems for speech and non-speech but with *different dynamics* for each oromotor behavior.

1. Introduction

Speech production is a complex process that requires coordinating multiple systems (respiratory, phonatory, resonatory, articulatory and prosodic) in an accurate, efficient and automatic manner. It involves transforming an abstract linguistic message into a motor code. The oromotor structures in charge of producing such motor codes are not only responsible of the production of speech targets, such as phonemes, syllables or utterances, but they also produce non-speech oral gestures. It is, however, unclear whether motor planning involves the same brain mechanisms for speech and non-speech. One challenge in addressing this question relies in separating motor planning from pre-motor processes. Here we take advantage of high temporal-resolution electro-encephalographic/evoked potential (EEG/ERP) approaches, combined with the manipulation of delayed production tasks to investigate speech and non-speech planning mechanisms.

1.1. Speech versus non-speech planning

Since speech and non-speech production involve similar movements and require the use of the same anatomical structures, recent studies on speech motor control have explored whether both oromotor behaviors are governed by the same neural substrates. In the literature, two opposite models debate about the relationship between speech and non-speech motor control systems. On one hand, the task dependent model (Ziegler, 2003a, 2003b; Ziegler & Ackermann, 2013) proposes separate

control subsystems for motor behaviors that overlap at the muscular level, suggesting specialized non-overlapping neural networks for speech as compared to non-speech. On the other hand, the integrative model by Ballard, Robin, and Folkins (2003) proposes a general motor control system in which some principles are shared between speech and other types of motor behaviors, such as non-speech, suggesting overlap at high levels of motor control. In the following paragraphs, we will detail the two models along with initial evidence supporting them, before turning to recent neuroimaging evidence on the relationship between speech and non-speech.

The task dependent model proposes neural substrates responsible for controlling motor functions being organized in a task-specific manner (Ziegler, 2003b). On one hand, innate motor behaviors (e.g. vegetative breathing or emotional expressions) are controlled in a different specialized way than motor behaviors acquired through extensive motor learning, such as speech. Non-speech gestures are, on the other hand, considered as novel oral motor tasks that recruit different basic motor abilities, as compared to speech, in order to achieve an uncommon goal, such as imitating an oral movement. As far as those tasks are constantly trained, a "particularly specialized motor machinery" will be available (Ziegler, 2003b, p. 100). A different neural organization between speech and non-speech tasks is based on neuroimaging studies showing differences in lateralization at the level of the motor cortex and the cerebellum when comparing speaking to tongue movements (e.g. Riecker, Ackermann, Wildgruber, Dogil, & Grodd, 2000; Wildgruber, Ackermann, Klose, Kardatzki, & Grodd,

---

* Corresponding author.
E-mail address: monica.lancherospompeyo@unige.ch (M. Lancheros).

1996). Some clinical findings have also supported the dissociation between the two oromotor tasks: voluntary motor functions such as airflow tracking or diadochokinetic sequences seem to rely more on the integrity of cerebellar circuits than speaking does (Deger, Ziegler, & Wessel, 1999; Ziegler, 2002). Recent behavioral data on healthy and clinical populations have also shown that correlations between non-speech, speech-like and speech tasks (single articulator movements, rapid syllable repetition and oral reading, respectively) differ significantly, suggesting separate latent traits across tasks (Staiger, Schölderle, Brendel, Bötzel, & Ziegler, 2017). Finally, studies on motor learning are taken as well as indirect evidence of a different neural organization for speaking versus voluntary non-speech movements. Since speaking is an automatized motor behavior, due to repetition over time, neural architectures specialized for this task would emerge not only during childhood but even later, when motor cortical activity can still be modulated by motor learning (e.g. Kami et al., 1995).

The opposed view, the integrative model, suggests that speech motor control is integrative or at least partially task independent, instead of being completely task dependent. Thus, it claims that speech shares cortical networks with some non-speech movements. In fact, Ballard et al. (2003) hypothesize that if two tasks share some movement characteristics or demands and are executed by the same groups of muscles, their neural substrates will overlap to some extent. The evidence taken as support comes from studies investigating voluntary oral movements versus mastication in non-human primates. For instance, Luschei and Goldberg (1981) found that the lateral prefrontal cortex played a role in producing both types of oral movements in primates. After bilateral lesions in this area, monkeys were not only impaired in performing previously trained mandibular tasks, but also their jaw movements were affected during mastication. Ballard et al. (2003) claimed that a certain overlapping neural tissue between tasks probably took place given the involvement of the same muscle movements. Consequently, damage in those overlapping areas in the cortex should cause impairment of multiple skills. In the case of humans, this may affect both non-speech and speech abilities. Results issued from research on motor circuits in lower-order animals, suggesting that behaviors are not controlled by separated self-governing systems (Kupfermann & Weiss, 2001), are also taken as arguments in favor of a single motor control system whose organization "is highly flexible, dynamically assembled, and can be rapidly modulated to accommodate a specific initial condition" (Ballard et al., 2003, p. 40).

More recent studies on the neural basis of speech versus non-speech oromotor control have reported results in favor of either positions, namely neural overlap or neural differences (see Maas, 2017 for a review). There are few neuroimaging results supporting the idea of separated motor control systems for speech and non-speech. The two studies we are aware of (Bonilha, Moser, Rorden, Baylis, & Fridriksson, 2006; Horwitz et al., 2003) have found differences in lateralization and different cortical activation when comparing the production of the two oromotor tasks. Thus, both studies reported consistent bilateral cortical activation when producing non-speech stimuli, whereas activation for speech tasks was found to be left-lateralized. Additionally, activation of the Brodmann area (BA) 45 was observed only for speech stimuli in both studies, while non-speech gestures activated other areas such as BA 4, 6 (Bonilha et al., 2006) and 44 (Horwitz et al., 2003).

Shared or overlapping brain areas between speech and non-speech gestures have also been reported in several investigations. Tremblay and Gracco (2009) found that repetitive transcranial magnetic stimulation on the pre-SMA equally affected the volitional selection of both words and non-speech gestures. They suggested an overlap between the production of speech and other motor behaviors, which may be compatible with domain-general neural resources used for speech. In the same vein, the fMRI results by Chang, Kenney, Loucks, Poletto, and Ludlow (2009) showed important activation overlap of left frontal and temporoparietal regions during the production of speech (non-sense syllables) and non-speech gestures, although with larger activations for

non-speech. Likewise, Basilakos, Smith, Fillmore, Fridriksson, and Fedorenko (2017) found that neural regions involved in vowel production were similarly engaged to those of non-speech gestures. Interestingly, fine-grained activation patterns showed differences between those two oromotor behaviors, suggesting that despite the engagement of the same regions during speech and non-speech production, those regions are involved in different ways. The latter results are in agreement with Saarinen, Laaksonen, Parviainen, and Salmelin (2005), reporting MEG activation of the same cortical areas during speech and non-speech production but to a different extent and amplitude according to the stimulus type being produced. In fact, they showed that kinematically similar non-speech movements elicited a higher activation in the motor cortex as compared to speech movements, which is associated with spatially less focal activity (Salmelin & Sams, 2002) within the motor cortex. Finally, Memarian et al. (2012) similarly showed that the production of non-speech oromotor tasks involved the activation of similar sensorimotor control areas as speech -phoneme-production, even though speech required additional activation of brain areas implicated in phoneme selection and articulatory control of speech.

In summary, there are two opposite claims about neuromotor control of speech production: (1) speech is controlled by a specialized, distinct motor system and (2) speech control is integrative and shares some brain processes with other motor behaviors, such as non-speech tasks. Recent studies in the field do not only report findings supporting one or the other views but some of them have also found intermediate results, showing activation of the same cortical areas but with different extent and amplitude of activation between speech and non-speech. In the present study, we aim to investigate whether the brain mechanisms for speech are the same as those of matched non-speech gestures during motor planning through high-density ERPs. Speech and non-speech stimuli will be carefully matched in terms of acoustic and somatosensory targets, in addition to using a procedure enabling the separation of motor planning process from pre-motor processes. These two issues will be further detailed below.

1.2. Matching speech and non-speech gestures

A major challenge in comparing speech and non-speech relies in properly selecting and matching both stimulus types. Given the difficulty to pair both oromotor sequences, some previous studies have selected a very limited number of stimuli, while repeating them many times. For instance, in their MEG study Memarian et al. (2012) only compared the oromotor task of opening/closing the mouth to the basic speech task of saying the phoneme /pa/ aloud. Similarly, in the MEG study by Salmelin and Sams (2002) a few non-speech stimuli (tongue lift and lip protrusion) were compared to different speech stimuli (silent articulation of the Finnish vowel /o/ and noun production). In addition to the limited number of stimuli, it has been argued that non-speech stimuli used in published studies are not comparable to speech stimuli, and thus, any difference in brain activation between oromotor tasks may be due to the selected stimuli involving different articulators (Ballard et al., 2003). This may be the case of soundless non-speech gestures compared to speech items (e.g. Memarian et al., 2012; Staiger et al., 2017).

In some studies, a more careful matching between speech and non-speech gestures was achieved, although often leading to the use of a reduced number of stimuli pairs or to covert production. Tremblay and Gracco (2009) included three stimuli matched on duration, involvement of the same articulators and audibility. Saarinen et al. (2005) included eleven non-speech movements designed to have phoneme-like features (i.e. the non-speech item "smile" was matched to the vowel /i/ or /e/, whereas the non-speech item "lips together" was matched to the consonant /m/) during a silent production task. The use of divergent experimental paradigms (i.e. covert vs. overt production tasks) may explain some contradictory results in terms of brain activation (i.e.

intensity changes of hemisphere dominance on speech movements; Riecker et al. (2000)). In the present study, we will try to overcome the pitfalls acknowledged in the literature by carefully matching 20 speech and non-speech items, which will be overtly produced after a brief delay.

1.3. Separating motor from pre-motor planning

The experimental investigation of the brain mechanisms at the level of motor speech planning has received less attention in the literature on language production than higher-level linguistic encoding processes (Laganaro, 2019). The challenge in investigating motor speech planning consists in its separation from pre-motor processes, i.e. from linguistic encoding. The delayed production task has often been used to those purposes (for instance, Chang et al., 2009; Kawamoto, Liu, Mura, & Sanchez, 2008; Mock, Foundas, & Golob, 2011; Tilsen et al., 2016). However, in a standard delayed production task participants may prepare the utterance up to the motor programs, meaning that only execution of the prepared motor programs is engaged in delayed production. Here, the delayed production task will be combined with an articulatory suppression task (Laganaro & Alario, 2006) in order to look more closely at the motor planning process ("phonetic encoding" in the psycholinguistic literature on speech production) of speech and non-speech stimuli.

In a delayed production task, the vocal response is given after a delay, when a cue is presented. During this delay participants are allowed to retrieve and prepare their response, or part of it. The amount of preparation achieves by speakers can be manipulated under different conditions. In particular, articulatory suppression interferes with the preparation of participants' responses through the repeatedly sub-articulation of a given syllable during the delay. Crucially, the (abstract) phonological process is left relatively intact, while the phonetic encoding is affected. Various lines of evidence support this hypothesis: on one hand, evidence from psycholinguistic and short-term memory studies indicate that phonological processes can survive to articulatory suppression (e.g. Balota & Chumbley, 1985; Goldinger, Azuma, Abramson, & Jain, 1997). On the other hand, combining delayed production with articulatory suppression tasks has allowed to experimentally determining the locus of certain psychological phenomena (i.e. whether the syllable frequency effect arises at the phonological or phonetic encoding stage, see Laganaro and Alario (2006) for a detailed rationale).

In this study, we investigate the neural dynamics of speech and non-speech motor planning through high-density ERPs. We will accurately match speech and non-speech stimuli while using a representative sample of both types of oral gestures. To separate motor planning from pre-motor encoding processes we will use a delayed production task, in which speakers prepare an utterance but produce it overtly only when a cue appears after a short delay (Experiment 1). However, since in a standard delayed production task speakers may not only plan the pre-motor code but also prepare the motor programs within the delay and have them ready to execute at the presentation of the response cue, we will combine the delayed production task with an articulatory suppression task in Experiment 2.

2. Delayed production task – Experiment 1

In Experiment 1, participants produced matched non-speech and speech stimuli presented in video-clips in a standard delayed production task, i.e. after a short variable delay.

2.1. Participants

Participants were all French native speakers, with no reported hearing, language, speech, neurologic or psychiatric disorders. They were all right handed on the Edinburg Handedness Scale (Oldfield, 1971). All subjects gave their informed consent to participate in the study, approved by the local ethics committee, and were paid for their participation. Only participants completing the task with an accuracy >65% were retained. All the 18 subjects that performed the task reached this criterion; however, two were removed because of many contaminated epochs in the ERP data. The 16 remaining participants (5 men, mean age: 23.7, range: 20–30 years old) were included in the analysis.

2.2. Materials

The stimuli were 20 non-speech gestures matched with 20 monosyllabic French words (mean lexical frequency = 57.8 per million words; mean syllabic frequency = 32 per million syllables, SD = 969), and with 20 non-word syllables, obtained by substituting a V or a non-initial C to the words (lexical frequency = 0; mean syllabic frequency = 11.9 per million syllables, SD = 32). Words and non-words had different syllabic structures, including CCV, CCVC, CVC, CVCC and CYV. Non-words were meaningless syllables that devoid of lexically but were existing French syllables.

Since non-speech stimuli could encompass a wide range of motor behaviors (Kent, 2015), we clarify that the non-speech targets in this study corresponded to sounded orofacial movements devoid of linguistic content but meaningful and easily reproduced from video recordings. The non-speech gestures were carefully paired with both speech stimuli (words and non-words) on type and mode of articulation of the gesture onset (i.e. the non-speech stimuli "bisou" -kiss-, was matched with the monosyllabic French word "point" (/pw ~ E/) and with the non-word /pwo/, based on a plosive and bilabial onset). See Appendix A for a detailed description of the non-speech items included in this study.

The material was therefore composed of matched triplets of non-speech gestures, monosyllabic words and non-words (see examples in Table 1 and the complete list in Appendix B). In addition to the 60 target stimuli, 22 fillers were included (see the Section 2.3).

Each stimulus was audio-video recorded, using the same female actor/speaker. Videos (resolution: 512 × 384 pixels; image frequency: 29.97 fps) were centered, cut off to show only the mouth with highlighted lips, producing the stimuli. Videos were also temporally cut, frame-by-frame, so that the mouth was always shown in a neutral position during 300 ms, before producing the target stimulus. Once the articulation was finished, the mouth came back to a neutral position and remained 300 ms on the screen, before the video ended. Background noise was cleaned up. Mean duration (ms) of the videos was similar across stimulus types (conditions hereafter): non-speech = 1269 (SD = 186); words = 1244 (SD = 212); non-words = 1256 (SD = 195), with no statistical difference between them (p > 0.05).

2.3. Procedure

Participants sat in a sound-proof dimly lit room approximately 70 cm in front of a PC screen. The experimental software E-prime (version 2.0, Schneider, Eschman, & Zuccolotto, 2002) was used for stimuli presentation and data collection. Participants were first familiarized with the stimuli by randomly presenting all targets on the screen; they were asked to simply look and listen to the videos. Second,

Table 1
Examples of matched triplets of non-speech, words and non-words.

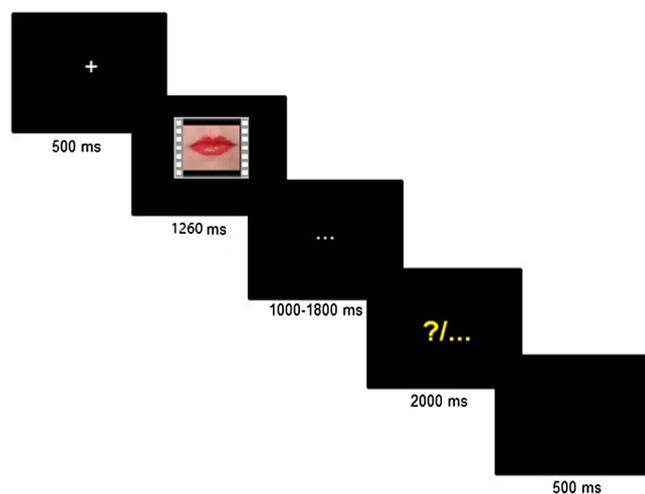| Non-Speech	Words	Non-Words
Bisou (*Kiss*)	Point /pw ~ E /	Pwo /pwo/
Craquement (*Creak*)	Crique /kRik/	Cruke /kRyk/
Moto (*Moto*)	Vrai /vRE/	Vru /vRy/

Fig. 1. Experimental procedure of Experiment 1. Participants responded when the cue "?" appeared on the screen. Reaction times were calculated from the apparition of the "?" to the response onset.

participants underwent a training phase accompanied by the experimenter, who explained the task. Finally, the experimental phase started.

In each experimental trial (see Fig. 1), a "+" sign appeared in the middle of the screen for 500 ms. Then, a video of the target stimulus was presented, followed by "…" in white, which randomly lasted between 1000 and 1800 ms (with steps of 200 ms). Participants were instructed to wait silently until a question mark appeared. A variable delay was used so that participants could not anticipate the response cue presentation (Laganaro & Alario, 2006). See Laganaro and Alario (2006, p. 185) for the rationale behind the shortest and longest delay chosen in this study. The question mark indicated that participants had to repeat the target stimulus shown in the video as fast and accurately as possible. The question mark remained on the screen for 2000 ms. However, if "…" in yellow appeared instead (500 ms), the participants did not have to give any answer and waited until the next trial started. The no-response condition was associated only to filler items (non-analyzed), which corresponded to the shortest and longest delays. The experiment began with five warm-up trials, repeated if necessary.

Each stimulus was presented three times throughout the task, once in each of the three intermediate delays (1200, 1400 or 1600 ms), and each filler item was presented once. Items were pseudo-randomized such that the same stimulus was not presented consecutively and the same delay was not presented in more than three consecutive trials. Three lists of 202 stimuli (60 items × 3 delays and 22 fillers), with a different order, were created and randomly assigned to each participant. The task was divided in three blocks of 68 items to allow participants two brief breaks in between.

2.4. Behavioral analyses

The digitized responses were systematically checked with a speech analysis software (CheckVocal 2.2.6, Protopapas, 2007) to identify correct responses and reaction times (from the question mark to the vocal onset, RT hereafter). No-responses, wrong responses, hesitations and/or auto-corrections were considered as errors. Wrong responses for speech items were any perceived phonemic deviation from the target syllable. For non-speech, only clear deviations from the target sounds were scored as errors. Accuracy was coded by the first author. Additionally, a random subset of 360 items was coded independently by a second judge (the second author of the paper). Inter-rater agreement (94%), was *substantial* (Kappa statistics, Landis & Koch, 1977).

Accuracy and RT data were fitted with mixed models (Baayen,

Davidson, & Bates, 2008) with the R-software (R-project, R-development core team 2005). The model was computed with accuracy or RTs as dependent variable, type of stimuli and order of stimuli presentation as fixed factors, and subjects and triplets as random factors.

2.5. EEG acquisition and pre-analysis

EEG was continuously recorded using the Active-Two Biosemi EEG system (Biosemi V.O.F. Amsterdam, Netherlands) with 128 channels covering the entire scalp. Offline, ERPs were high and low-pass filtered (0.1–30 Hz), notch-filtered (50 Hz) and averaged for each participant. Epochs of 300 ms were extracted, time-locked to the response cue (i.e., stimulus-locked) and to the onset of the vocal response (i.e., response-locked). Epochs contaminated by eye blinking, eye-movements, movements or other noise artefact were excluded from averaging after visual inspection. Only trials with correct responses, valid RTs and both backward and its corresponding forward uncontaminated epochs were retained for further analysis (73.3%, 71.7% and 73.3% of the epochs were retained for words, non-words and non-speech, respectively, with a minimum of 20 epochs per condition). Epoch extraction and averaging was computed for each participant using the Cartool software (Brunet, Murray, & Michel, 2011). Both time-locked ERPs were averaged separately per participant and per condition. Electrodes with signal artefacts were interpolated using 3-D splines interpolation (Perrin, Pernier, Bertnard, Giard, & Echallier, 1987). The average of electrodes interpolated by participants was 13% (max = 16%; or up to 20 of the 128 electrodes).

2.6. EEG data analyses

The aim of the ERP analyses was to determine whether the motor planning of speech and non-speech stimuli is governed by the same neurophysiological processes. To do so, we examined whether non-speech stimuli, words and non-words generated the same waveform amplitudes and the same electric fields. Changes in electric field take place when the underlying generator configuration has changed (Michel, Koenig, Brandeis, Gianotti, & Wackermann, 2009; Michel & Murray, 2012) and, differences in underlying generator suggest activation of different brain networks.

We initially examined the differences in ERP mean amplitudes between non-speech and speech (waveform analysis). We then run topographic consistency tests to ensure that the topographies of each condition showed sufficient consistency (Koenig & Melie-García, 2010). Afterwards we compared the global dissimilarity between conditions to determine objectively whether topographies differed and in which time window(s) (TANOVA, Murray, Brunet, & Michel, 2008). Finally, we performed a topographic pattern analysis (spatio-temporal segmentation) to explore whether differences in global dissimilarity were due to different stable topographic patterns *per-se* or to different time course of the same stable topographic patterns between non-speech and speech. The details of each of these analyses are provided below.

2.6.1. Waveform analysis

ERPs were subjected to a sampling point-wise ERP waveform analysis to determine time-periods presenting local differences in ERP amplitudes across conditions. Waveform amplitude comparisons were run on each electrode and for each time point, using paired ANOVAs with the Sten toolbox, (http://www.unil.ch/fenl/sten), separately on stimulus- and response-locked ERPs. Only differences over at least five adjacent electrodes, extending over at least 20 ms were retained (alpha criterion of 0.01).

2.6.2. Topographic consistency test (TCT)

The TCT (Koenig & Melie-García, 2010) allows to check whether a given scalp field is consistently activated by the event of interest in the time window of analysis. It compares time-point by time-point the

global field power (GFP) of the averaged ERPs to the empirical distribution of the GFP obtained after the random shuffle of the data across electrodes. This test was performed separately on the stimulus- and response-locked ERPs (L2 normalization, 5000 runs and alpha of 0.05) using the Ragu software (Koenig, Kottlow, Stein, & Melie-García, 2011).

2.6.3. Topographic analysis of variance (TANOVA)

The TANOVA (Murray et al., 2008) determines to what extent the topography of the ERPs differs across conditions. This analysis focuses on the global dissimilarity index (GDI, Lehmann & Skrandies, 1980), providing a single measure per time point that reflects the dissimilarity between two electric fields. It is computed by taking the root mean square of the difference between two normalized maps (see Michel & Murray, 2012). We used the procedure implemented in the Ragu software, based on permutation tests (see Koenig, Stein, Grieder, & Kottlow, 2014 for details). For each time point, the GDI was first computed for each condition from the group averaged ERPs. The data were then permuted to build the empirical distribution of the index under the null hypothesis. Thus, topographic maps were randomly assigned to the experimental conditions and the GDI was recomputed from these data. The GDI of the original ERPs was then compared to the empirical distribution resulting from the permutations in order to determine the likelihood of obtaining a higher GDI value than the one actually obtained. In the present study, this analysis was conducted with 5000 iterations and alpha set to 0.05. We considered only significant TANOVAs which lasted more than 20 ms.

2.6.4. Topographic pattern analysis

The aim of topographic pattern analyses (also called spatiotemporal segmentations or microstate analysis) is to determine the stable electrophysiological patterns (or topographic maps) that best explain the data at each time frame (see Michel et al., 2009). This analysis additionally permits to determine whether differences between conditions arise because the topographic maps differ between them, or because of differences in the time course of the same topographic maps. In the present study, the method described by Koenig et al. (2014) and implemented in the RAGU software was used for the spatio-temporal segmentation.

Topographic maps observed in the group-averaged data were then statistically tested by comparing each map templates with the moment-by-moment scalp topography of individual ERPs ("fitting" procedure). Fitting determines how well a topographic template map explains single participant responses for each condition. Each data sampling point was labelled according to the template map with which it best correlated spatially, giving as output variables such as map duration and global explained variance (GEV) in each individual data. Those values were then used to statistically test topographic differences between non-speech and speech: the Friedman non-parametric test was used to test differences in map duration and GEV across conditions. To test for differences between non-speech and non-words and, between non-speech and words, Wilcoxon signed-rank tests were conducted with a Bonferroni correction (significance level set at $p = <0.025$). Templates were fitted in the individual ERPs with the Cartool software (Brunet et al., 2011) to get the fitting statistics for the statistical comparison across conditions.

3. Results

3.1. Behavioural results

Production accuracy for non-speech, words and non-words was 94% (SD = 4), 95% (SD = 5) and 87% (SD = 9) respectively. The generalized mixed model for binomial distribution showed an accuracy significantly higher for non-speech as compared to non-words ($z = -4.60$, $p < 0.0001$, $\beta = -0.78$, SE = 0.17), and no significant difference between non-speech and words ($z = 1.16$, $p = 0.244$, $\beta = 0.23$, SE = 0.20).

Mean response latencies for non-speech, words and non-words were

respectively 533 ms (SD = 128), 527 ms (SD = 123) and 520 ms (SD = 120). The linear mixed model revealed a main effect of order (F (1,2448.3) = 34.08; $p < 0.001$) and condition (F(2,2439.3) = 5.46; $p < 0.01$). Contrasts revealed significant differences between non-speech and non-words (t(2439.8) = -3.31; $p < 0.001$, $\beta = -16.59$, SE = 5.02) but not between non-speech and words (t (2438.8) = -1.67; $p > 0.05$, $\beta = -8.18$, SE = 4.90).

3.2. ERP results

3.2.1. Waveform analysis

Analysis on the stimulus-locked ERPs revealed significant differences in amplitude only between non-speech and non-words from 227 to 275 ms after the presentation of the question mark, on clusters of right posterior and left anterior channels, and no significant differences between non-speech and words. On response-locked ERPs, significant differences in amplitudes were only present between non-speech and non-words from -245 to -228 ms, on a small cluster of 6 right posterior channels (see Fig. 2a).

3.2.2. TCT

The TCT (Koenig & Melie-García, 2010) revealed consistency on the signal over the whole analyzed time window for each condition on both time-locked ERPs (see Appendix C).

3.2.3. TANOVA

Pairwise TANOVAs on the stimulus-locked ERPs revealed significant differences across conditions from 221 to 264 ms, while no significant differences appeared on the response-locked ERPs (see Fig. 2c).

3.2.4. Topographic pattern analysis

The spatio-temporal segmentation revealed 6 different electrophysiological template maps respectively on the stimulus-locked and 4 on the response-locked ERPs. In the stimulus-locked ERPs, a different period of stable electrophysiological pattern seems to be present, at least on the grand-averaged ERPs, for non-words in the same time window revealed by the TANOVA. To confirm differences in global electrophysiological patterns, the topographic map templates present during the time window in which the TANOVA revealed significant differences across conditions (maps C, E and F) were fitted in the individual ERPs from 200 to 300 ms. Map C showed significant differences across conditions on duration ($\chi 2(2) = 6.32$; $p = 0.04$; mean duration (ms): non-speech = 40.28, words = 48.46, non-words = 55.05) but not on GEV ($\chi 2(2) = 3.11$; $p = 0.21$; mean GEV: non-speech = 9%, words = 11%, non-words = 13%). Wilcoxon signed-rank tests revealed no significant differences on duration neither between non-speech and non-words (Z = -2.09, $p = 0.04$), nor between non-speech and words (Z < 1). Similarly, map E showed significant differences across conditions on duration ($\chi 2(2) = 7.51$; $p = 0.02$; mean duration (ms): non-speech = 40.04, words = 38.70, non-words = 25.39,) but not on GEV ($\chi 2(2) = 1.68$; $p = 0.43$; mean GEV: non-speech = 4%, words = 4%, non-words = 3%). Wilcoxon signed-rank tests on duration showed significant differences only between non-speech and non-words (Z = -2.49, $p = 0.01$). Finally, map F showed significant differences across conditions on GEV ($\chi 2(2) = 5.85$; $p = 0.05$; mean GEV: non-speech = 4%, words = 1%, non-words = 4%) but not on duration ($\chi 2(2) = 3.69$; $p = 0.16$, mean duration (ms): non-speech = 17.33, words = 10.15, non-words = 17.21). Wilcoxon signed-rank tests on GEV showed that there were no significant differences neither between non-speech and non-words (Z = -3.38, $p = 0.73$), nor between non-speech and words (Z = -2.02, $p = 0.04$).

In the response-locked ERPs the same four periods of stable global electrophysiological patterns (G, H, I, J in Fig. 2) appeared in all conditions. On the grand averages, the distribution of the four maps seems slightly different for non-speech, but the TANOVA did not reveal significant differences across conditions.
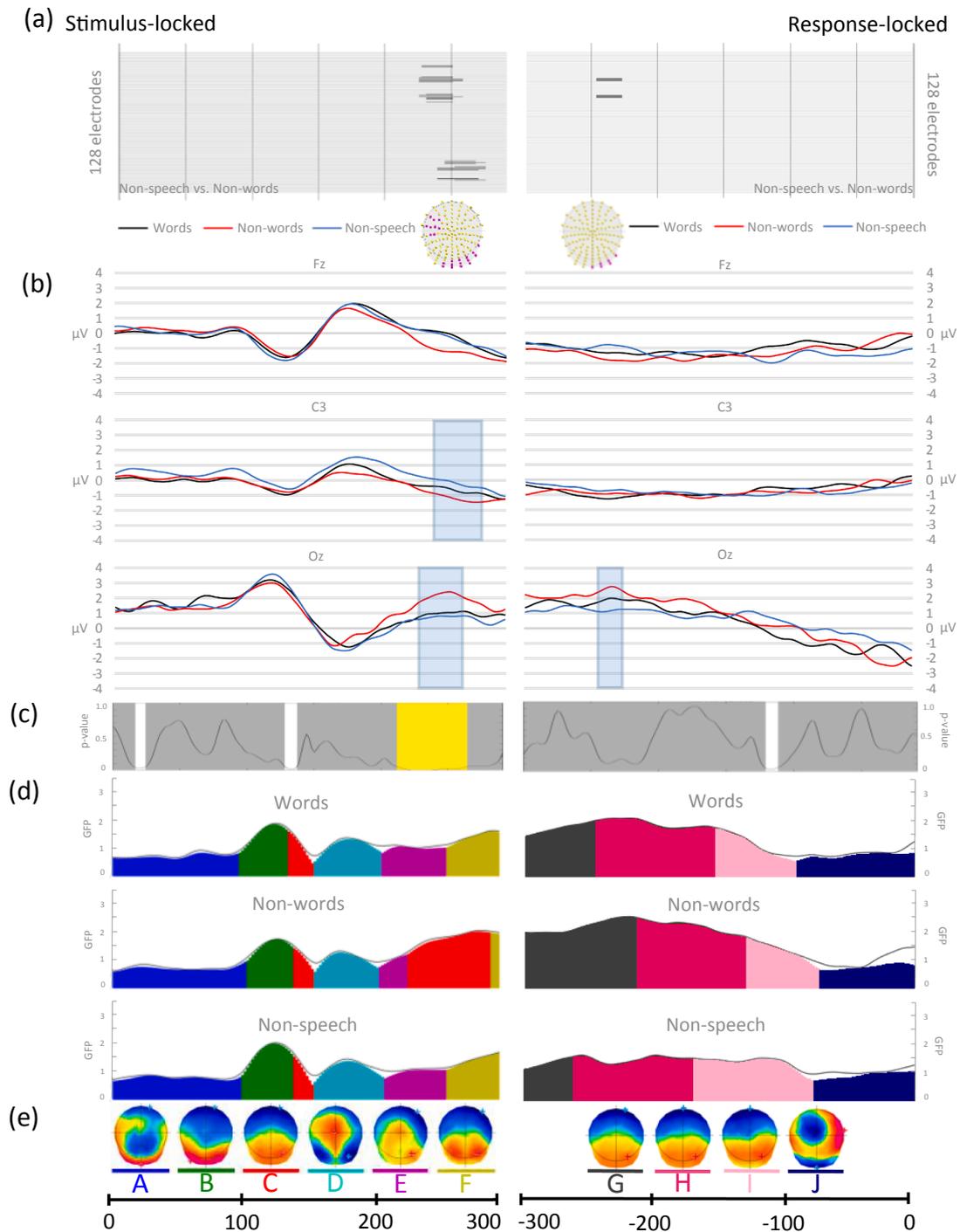
Fig. 2. (a) Results of the waveform analyses across conditions on all time points/electrodes in the stimulus- and response-locked ERPs (left and right panel, respectively) with grey points indicating significant differences (p < 0.01) in amplitudes between non-speech and non-words, (b) examples of waveforms (Fz, FC3, O2) from stimulus- and response-locked ERPs. Highlighted time windows indicate periods of significant differences in mean amplitudes between non-speech and speech. (c) Results of the TANOVA analysis for stimulus- and response-locked ERPs. Yellow areas are periods of significant differences in topographic dissimilarity lasting longer than 20 ms. (d) Temporal distribution of the topographic maps revealed by the spatio-temporal segmentation analysis displayed on the mean GFP from the gran average of each condition for both time-locked ERPs. (e) Map templates (topographic maps) corresponding to each period of stable global electro-physiological pattern at scalp: the colors under the template corresponds to the colors in (d).

4. Discussion

This experiment aimed at comparing motor planning processes for the production of non-speech stimuli and of matched monosyllabic words and non-words. To separate motor planning from pre-motor processes we used a delayed production task, in which speakers prepare a syllable or an orofacial movement (i.e. non-speech stimuli) but

produce it overtly after a brief delay.

Behavioral results, on one hand, suggest similar production accuracy between non-speech stimuli and closely matched monosyllabic words, at least when perceptual criteria are used. On the other hand, they also indicate that in a delayed production task, the production of non-speech sequences does not take longer than producing syllables corresponding to real words, at least when the participants have enough

time to prepare their motor plans. However, we observe lower accuracy, along with slightly (13 ms) but significant faster production latencies for monosyllabic non-words as compared to non-speech sequences. An interpretation of these unexpected results for non-words may be found when integrating also the ERP results.

As the delayed production task allows participants to prepare their response if sufficient time is provided, we expected to find the same electrophysiological patterns (topographies) between non-speech and speech. However, both waveform and global pattern analyses revealed a significant difference only between non-speech and non-words starting around 250 ms after the question mark eliciting the production. Contrarily, similar patterns were found between non-speech and both speech conditions in the response-locked ERPs. The ERP results thus join the behavioral results indicating non-significant neural differences between the production of non-speech sequences and monosyllables corresponding to words, but behavioral and ERP differences with monosyllabic non-words. The ERP differences appearing only in the stimulus-locked ERPs at about 250 ms, fall somewhere in the middle of the time separating the response cue and the vocal onset (as mean RTs are around 520 ms). They fall on a P2-like component, whose amplitudes are larger for non-words (see Fig. 2b).

Waveform modulations of a P2-like component have been previously reported in language tasks. In particular, modulation of the P2 in speech (word) production has been associated with lexical processes. For instance, Strijkers, Holcomb, and Costa (2011) found P2 modulation in an object naming task including items with different lexical frequency, with low frequency words showing more positive amplitudes than high frequency words. In the present study significant differences were found in the stimulus-locked ERPs, at around 250 ms after the presentation of the response cue whereas P2 modulation in the study by Strijkers et al. (2011) started 152 ms after stimulus onset in a different task (picture naming). The P2 modulation we observed is driven by non-word stimuli, which are the only ones devoid of semantic information. Differences between non-speech and non-words on the P2 component may be associated to the retrieval/unpacking of the motor plans prepared during the delay. Actually, when the question mark is presented (beginning of the stimulus-locked ERP recording), participants would access/launch the motor plans that had been prepared during the delay to posteriorly execute them. The differences between non-words and non-speech likely reflect the different ways of keeping meaningless and meaningful sequences in memory. Although it is often supposed that non-speech stimuli do not carry meaning (Kent, 2015), non-speech stimuli in this study can be associated to meaningful representations the same extent as for words.

More crucially for our purpose here is that response-locked ERPs microstates did not differ between non-speech and both speech conditions (words and non-words). This result may suggest that the processes occurring before articulation are common for speech and non-speech, and are governed by the same neural networks. Before any further interpretation however, we will turn to Experiment 2 in which the delayed production is combined with articulatory suppression, disabling the maintenance of a motor plan during the delay.

5. Delayed production task with articulatory suppression – Experiment 2

5.1. Participants

21 French-speaking right-handed adults, with no reported hearing, language, speech, neurologic or psychiatric disorders took part to this experiment. All subjects signed an informed consent form and were paid for their participation. As in Experiment 1, only participants with an accuracy >65% were selected. On the 20 participants reaching this criterion, four were removed because of contaminated epochs in the ERP data. The 16 remaining participants (6 men, mean age: 23, range: 19–30 years old) were included in the analysis. None of them participated in Experiment 1.

5.2. Materials

The materials were those used in Experiment 1.

5.3. Procedure

Participants underwent the same familiarization and training phases as in Experiment 1. The procedure of this experiment was similar to Experiment 1, except for the following: after the video presentation, "bla bla" appeared on the screen, instead of the "…", for a random time of either 1200, 1400 or 1600 ms. Participants were asked to repeatedly produce the syllable "bla", at a rhythm of approximately 2 "bla" per second, until an exclamation mark appeared, indicating to stop their production (see Fig. 3). Then a blank of 100 ms was presented, followed by either a question mark, or "…" in yellow, to respectively repeat or not the target stimulus as in Experiment 1.

5.4. Behavioral analyses

The behavioural analyses and the criteria for coding the vocal responses were the same as those of Experiment 1. In addition, in Experiment 2, responses that included the production of the syllable /bla/ (i.e. when articulatory suppression was not stopped on time) were also excluded. Similar to Experiment 1, accuracy was coded by the first author and a random subset of 360 items was coded independently by the second author. Inter-rater agreement was again high (94%, $\kappa = 0.6$).

5.5. EEG acquisition, pre-analysis and analysis

Acquisition and pre-analysis of the EEG data were the same as in Experiment 1. The retained epochs were 70%, 66.7% and 71.7% for words, non-words and non-speech, respectively, with a minimum of 22 epochs per condition. The average of electrodes interpolated for this Experiment were 14% (max = 18%; or up to 23 of the 128 electrodes).

The same EEG analysis as in Experiment 1 were computed.

6. Results

6.1. Behavioural results

Production accuracy for non-speech, words and non-words was 83% (SD = 11), 85% (SD = 10) and 80% (SD = 11) respectively. The generalized mixed model for binomial distribution showed no significant difference in accuracy either between non-speech and words ($z = 0.98$, $p = 0.32$, $\beta = 0.13$, SE = 0.13) or between non-speech and non-words ($z = -1.82$; $p = 0.07$, $\beta = -0.23$, SE = 0.12).

Mean response latencies for non-speech, words and non-words were respectively of 558 ms (SD = 151), 544 ms (SD = 149) and 559 ms (SD = 152). The linear mixed revealed a main effect of order [F (1,2201.5) = 51.16; $p < 0.001$] and condition [F(2,2194.4) = 3.41; $p < 0.04$]. Contrasts revealed significant differences between non-speech and words (t(2193.9) = -2.31; $p = 0.02$, $\beta = -13.78$, SE = 5.97) but not between non-speech and non-words (t (2194.7) = -0.09; $p = 0.92$, $\beta = -0.58$, SE = 6.08).

6.2. ERP results

6.2.1. Waveform analysis
No significant differences appeared on amplitude in both time-locked ERPs (see Fig. 4a and b).

6.2.2. TCT
The TCT for the stimulus-locked ERPs revealed consistency from the beginning of the recorded data until about 230 ms. Data outside this time window (i.e. from 230 to 300 ms) was discarded for subsequent
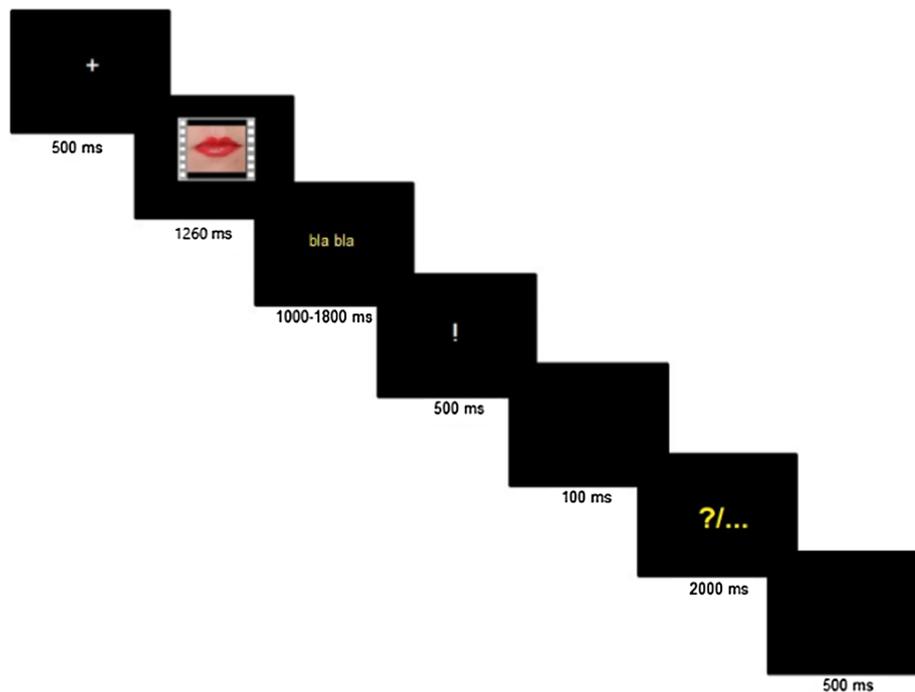
Fig. 3. Experimental procedure of Experiment 2. Participants were asked to respond when the cue "?" appeared on the screen. Reaction times were calculated from the apparition of the "?" to the response onset.

analysis. The response-locked ERPs were wholly consistent over the analyzed time window for the three conditions (see Appendix D).

6.2.3. TANOVA

Pairwise TANOVA revealed significant differences across conditions from 118 to 143 ms on the stimulus-locked ERPs. On the response-locked ERPs, it showed significant differences across conditions between -94 and -126 ms and between -209 and -247 ms (see Fig. 4c).

6.2.4. Topographic pattern analysis

The spatio-temporal segmentation revealed 3 different electrophysiological template maps on the stimulus-locked and 4 on the response-locked ERPs (see Fig. 4d). Similar to Experiment 1, map templates appearing in the time windows in which the TANOVA revealed significant differences between non-speech and speech were fitted in the individual ERPs. Thus, maps A and B for the stimulus-locked ERPs were fitted from 0 to 140 ms. Map A showed no significant differences across conditions (duration: $\chi2(2) = 1.54$; p $= 0.46$; mean duration (ms): non-speech $= 68.60$, words $= 74.45$, non-words $= 73.97$; GEV: $\chi2(2) = 1.58$; p $= 0.45$; mean GEV: non-speech $= 6\%$, words $= 8\%$, non-words $= 8\%$). Similarly, map B did not reveal significant differences across conditions (duration: $\chi2(2) = 1.54$; p $= 0.46$; mean duration (ms): non-speech $= 70.06$, words $= 64.20$, non-words $= 64.69$; GEV: $\chi2(2) = 1.20$; p $= 0.55$; mean GEV: non-speech $= 12\%$, words $= 11\%$, non-words $= 12\%$).

On the response-locked ERPs, given that the TANOVA revealed significant differences across conditions in two different time windows, maps D, E and F were fitted in the individual ERPs from -300 to -100 ms, and maps F and G from -100 to 0 ms. In the first fitting period, map D was observed in all conditions (presence in individual ERPs: non-speech $= 56\%$, words $= 75\%$, non-words $= 63\%$), but its duration was significantly different across conditions (duration: $\chi2(2) = 8.86$; p $= 0.01$; mean duration (ms): non-speech $= 52.61$, words $= 91.80$, non-words $= 87.40$). Wilcoxon signed-rank tests showed significantly shorter duration of map D in non-speech relative to non-words (Z $= -2.35$, p $= 0.02$), and to words (Z $= -2.31$,

p $= 0.02$). There were no significant differences on GEV ($\chi2(2) = 4.12$; p $= 0.13$; mean GEV: non-speech $= 5\%$, words $= 9\%$, non-words $= 8\%$). Even though the Friedman test on GEV revealed no significant difference across conditions, the Wilcoxon signed-rank tests showed significant differences between non-speech and words (Z $= -2.48$, p $= 0.01$). Differences between non-speech and non-words on GEV did not reach significance at the corrected threshold (Z $= -2.11$, p $= 0.03$). Map E, despite not shown on the grand averages for words, was present in some participants in all conditions (presence in individual ERPs: non-speech $= 69\%$, words $= 38\%$, non-words $= 44\%$), but its duration differed significantly across conditions ($\chi2(2) = 6.04$; p $= 0.05$; mean duration (ms): non-speech $= 62.87$, words $= 33.08$, non-words $= 39.92$). Duration of map E appeared to be different between non-speech and words (Z $= -2.39$, p $= 0.02$), but not between non-speech and non-words (Z $= -1.50$, p $= 0.13$). No significant differences were found on GEV ($\chi2(2) = 4.22$; p $= 0.12$; mean GEV: non-speech $= 6\%$, words $= 3\%$, non-words $= 4\%$). Finally, map F was similarly distributed across conditions in terms of duration ($\chi2(2) < 1$, mean duration (ms): non-speech $= 81.79$, words $= 72.39$, non-words $= 69.95$), and GEV ($\chi2(2) = 1.32$; p $= 0.52$; mean GEV: non-speech $= 5\%$, words $= 6\%$, non-word $= 5\%$).

From the second fitting, map F and G showed no significant difference across conditions either on duration (all: $\chi2(2) < 1$)) or on GEV (map f: $\chi2(2) < 1$; map G: $\chi2(2) = 1.26$; p $= 0.53$).

7. Discussion

In this experiment, we aimed to investigate whether motor planning involves the same brain mechanisms for speech and non-speech behaviors requiring similar musculature. To do so, we combined the delayed production task with an articulatory suppression task. As stated in the introduction, articulatory suppression disrupts participants' ability to store/prepare the gestural scores during the delay before the presentation of the response cue.

Behavioural results showed similar accuracy between non-speech and speech, when they are coded on perceptual criteria, and significantly longer RTs for non-speech stimuli relative to words; there was
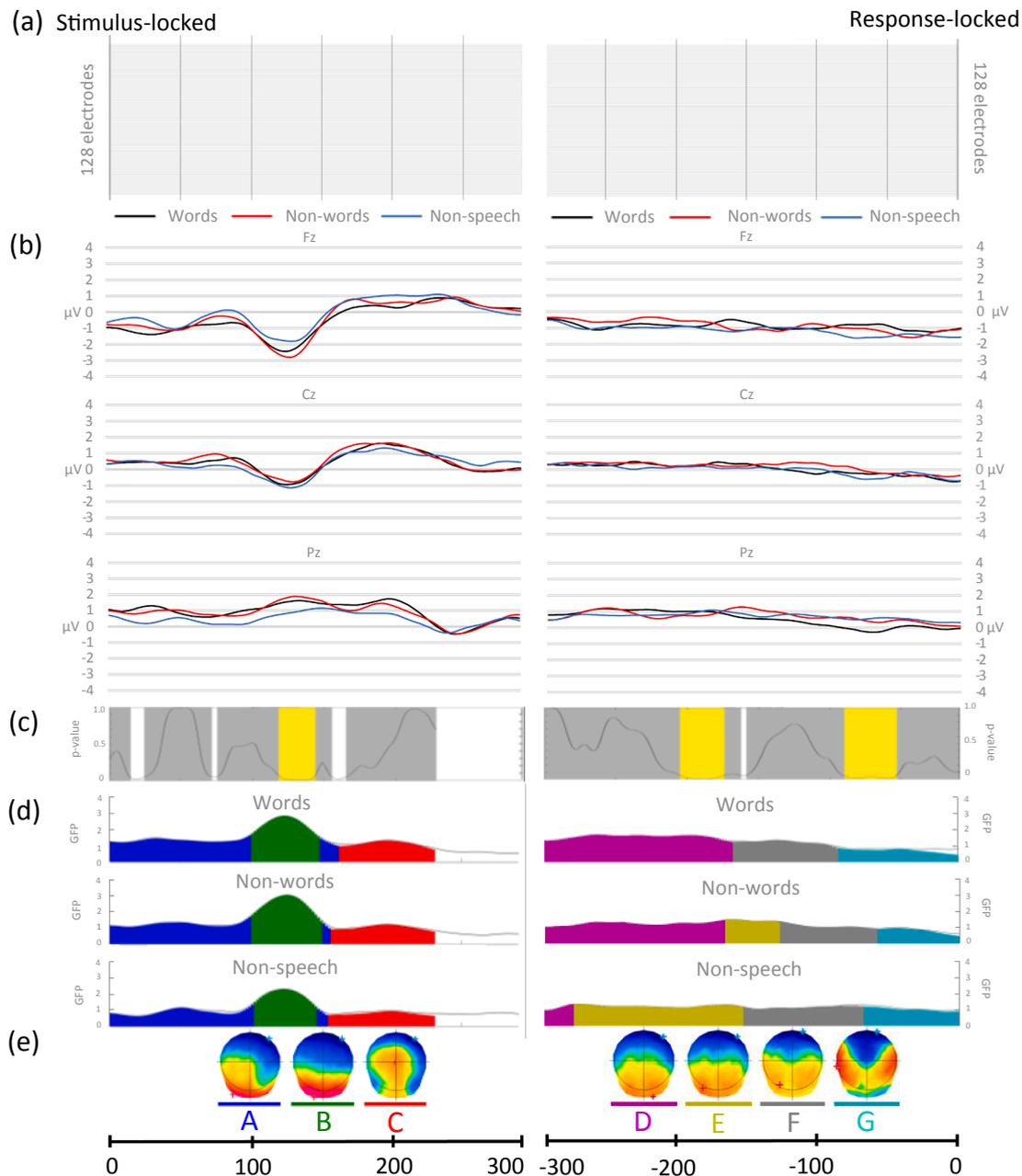
Fig. 4. (a) Significant differences (p < 0.01) in mean amplitudes between non-speech and speech on all time points/ electrodes in the stimulus- and response-locked ERPs (left and right panel, respectively), (b) examples of waveforms at the same electrodes for both time-locked ERPs (Fz, Cz, Pz), (c) results of the TANOVA analysis for both time-locked ERPs. Yellow areas are periods of significant differences in topographic dissimilarity lasting longer than 20 ms. (d) Temporal distribution of the topographic maps per conditions revealed by the spatio-temporal segmentation analysis for stimulus- and response-locked ERPs. (e) Map templates corresponding to each topographic map.

no significant difference between non-speech and non-words. Topographic pattern analysis showed no significant differences in terms of electrophysiological stability between non-speech and speech on the stimulus-locked ERPs. Crucially, the topographic pattern analysis on the 300 ms preceding the vocal onset (response-locked ERPs) revealed significant differences in stable global electrophysiological patterns between non-speech and words and, to a lesser extent (i.e. not on all the analyzed parameters), between non-speech and non-words. As for Experiment 1, the ERP results are in line with the behavioral results, both being different from the results of Experiment 1. As an interpretation of the results is enabled by the comparison of both experiments, we will discuss it only in the general discussion.

8. General discussion

Production of speech and non-speech gestures involves the use of the same anatomical structures. Recent literature on motor speech control has explored whether this means overlapping neural substrates (e.g. Basilakos et al., 2017; Chang et al., 2009). In this study, we compared, behaviorally and on high density EEG/ERPs, the latest processing stages of speech and non-speech planning in order to investigate the dynamics of the transformation of French syllables and of matched non-speech sequences into the corresponding articulation. To separate motor planning from pre-motor processes, in Experiment 1 participants performed a delayed production task, where speakers prepare an utterance or a non-speech gesture, but produce it overtly only after a variable delay. Additionally, in order to avoid

the full preparation or storage of the motor plans (Laganaro & Alario, 2006), in Experiment 2 participants executed an articulatory suppression task during the delay (e.g. repeated articulation of the syllable "bla"). This last task probably required some additional cognitive load as compared to the simple delayed production task in Experiment 1; indeed, even though different participants are involved in the two Experiments, preventing direct comparisons, accuracy is lower for all stimulus types in Experiment 2. However, given that the articulatory suppression was performed in all conditions, the same cognitive charge is involved for speech and non-speech production in Experiment 2. The charge per se should therefore not affect the observed differences across conditions.

Behavioral results of Experiment 1 showed significant differences, both in terms of accuracy and RTs, only between non-speech and non-words, with the former being produced more accurately but slower than the latter ones. ERP results are in line with the behavioral results: they revealed a significant difference at around 250 ms after the question mark, with a microstate lasting longer for non-speech as compared to non-words. In the discussion of Experiment 1 we suggested that this finding could be due to semantic differences between non-speech and non-word stimuli, only the former being meaningful. More importantly for the purpose of this study, global pattern analyses on response-locked ERPs in Experiment 1 did not reveal significant differences between non-speech and speech (whether words or non-words), suggesting that "late" processes preceding articulation are governed by the same neural patterns.

In Experiment 2 behavioral and ERP results appeared mainly between non-speech and words, with different distribution of global electrophysiological patterns at scalp in the response-locked data, i.e. in the time window preceding articulation. Non-speech movements initialising significantly later than words in Experiment 2 suggests that planning non-speech gestures is more costly relative to planning gestures corresponding to monosyllabic words. Similar results have been reported in the past when comparing two types of speech stimuli, namely frequent and infrequent syllables (Cholin, Levelt, & Schiller, 2006; Laganaro & Alario, 2006). Here, producing non-speech is slower than producing matched words, i.e. frequent French syllables, but is not slower than producing matched non-words composed of less frequent syllables. The difference in the distribution of microstates on the response-locked ERPs also points to the same direction, with significant differences between non-speech and words and to a lesser extent between non-speech and non-words. Thus, the first overall interpretation of the results is that when the preparation/storage of motor plans is disabled, non-speech planning activates the neural networks differently as compared to planning monosyllabic words, and seems much more similar to planning less frequent speech gestures. This interpretation will be further refined below.

8.1. Non-speech vs speech

As reviewed in the introduction, the model proposing that speech is different from all other oromotor behaviors and thus, involves separate neuromotor control systems as compared to non-speech (Ziegler, 2003a, 2003b; Ziegler & Ackermann, 2013), is opposed to the model suggesting that both speech and non-speech can be integrated in a general shared system of motor control (Ballard et al., 2003). The present results do not seem to agree with the task dependent model given that the same neural networks were involved during the motor planning of speech and non-speech sequences. On the other hand, the processes are not entirely the same across speech and non-speech in Experiment 2, as they were differently involved depending on the type of stimuli being planned. The results of this study rather support overlapping neural processes as already reported in previous neuroimaging studies (Basilakos et al., 2017; Chang et al., 2009; Memarian et al., 2012; Saarinen et al., 2005; Salmelin & Sams, 2002; Tremblay & Gracco, 2009) but with different dynamics depending on the task. They are also in line with recent results by Mugler et al. (2017), who found activation of the same cortical areas for speech and non-speech stimuli but some of them were involved to a greater extent in the production of one or another type of stimulus. Thus, the results could be consistent

with a weaker version of a task-dependent model, in which the same neural network is involved for speech and non-speech gestures but their control is different. However, this weaker version of the model then becomes virtually indistinguishable from the integrative model, requiring, at the very least, a task-dependent model to be more specific and constrained in its claims (i.e. not requiring non-overlapping neurophysiological networks).

It should be stressed that results of this study were obtained with a careful match of non-speech and speech stimuli in terms of acoustic and somatosensory targets, also ensuring a comparable articulatory complexity between non-speech and speech stimuli. Hence, incongruent results with previous results favoring speech-specific cortical networks (Bonilha et al., 2006; Horwitz et al., 2003) may be due to different accuracy in the stimuli matching: comparisons between speech stimuli and meaningless non-speech movements with very different articulatory properties may indeed result in different brain activation patterns than when comparing the production of meaningful matched speech and non-speech sequences.

8.2. Motor planning vs pre-motor processes

One of the major challenges of this study was to experimentally target motor planning, i.e. to separate pre-motor processes from motor planning. The delayed production task has been used in previous studies to disentangle linguistic from motor encoding (Kawamoto et al., 2008; Kemeny et al., 2005; Mock et al., 2011; Rastle, Croot, Harrington, & Coltheart, 2005; Tilsen et al., 2016), i.e. at the interface between phonological and phonetic encoding according to Levelt's model of word production (Levelt, 1993).

Chang et al. (2009) used a similar experimental paradigm to specifically target speech and non-speech motor planning processes. In their study, participants could retrieve/prepare the motor plans of non-speech and speech targets during the presentation of an arrow, which remained on the screen for a given delay. The participants had to produce the previously planned non-speech or speech response only when the arrow was replaced by a cross. They analyzed the brain activation during the actual production phase and only with a silent delay. The result of overlapping neural networks found in their study is completely in line with the results of Experiment 1, in which participants could prepare and hold the motor programs. However, only the condition with articulatory suppression in Experiment 2 enabled to show that the preparation of motor plans involves the same processes but with *different dynamics* for speech and non-speech.

8.3. Overlapping brain processes with different dynamics for speech and non-speech

The observed differences on the microstate distribution between non-speech and speech on response-locked ERPs in Experiment 2 are likely associated with differences in motor planning because they are observed (1) only when the articulatory suppression is executed and (2) only on response-locked ERPs. As discussed above, the topographic differences on response-locked ERPs do not suggest different underlying neurophysiological networks given that the same topographies are present for non-speech and speech. Instead, those differences could be interpreted as brain processes being differently implicated when planning motor codes for non-speech, non-words and words. One could think of it as neural networks that are more or less activated depending of the type of stimuli that is being planned. The question then is how to interpret them in terms of functional differences.

In speech production models, highly trained speech gestures (syllable-sized in some models) are thought to be stored and retrieved efficiently because they may correspond to highly over-learned motor actions (Guenther, Hampson, & Johnson, 1998; Levelt, 1993), while low-frequency or novel syllables are retrieved with less ease or their motor plans have to be assembled online (see for instance Bürki, Cheneval, & Laganaro, 2015). What happens with the non-speech stimuli is, however, less clear. On one hand, one may wonder whether overlearned non-speech motor gestures are stored in a similar way to

what is supposed for highly trained speech gestures and if the underlying brain structures involved in the storage and retrieval of oromotor gestures are shared for speech and non-speech. On the other hand, if non-speech motor plans are assembled online, one would hypothesize a similar computation as for low frequency/novel speech gestures. The present ERP results suggest that the dynamics of the activated brain networks are distinct between non-speech and words, and that infrequent speech gestures seem to be somehow intermediate between words and non-speech stimuli. To summarize, the present results are not compatible with retrieval versus online assembling of motor gestures, rather with similar functional processes being differently involved for speech and non-speech gestures. Which exact motor planning processes are involved in the brain mechanisms with opposite duration for words and non-speech (maps D and E in Experiment 2) needs further investigation.

9. Conclusion

In this study, the neural dynamics of the motor planning processes for non-speech, words and non-words were investigated by means of a delayed production and an articulatory suppression tasks. Results suggest that the latest planning processes before the actual articulation recruit the same neural networks for the three types of oromotor sequences but they are differently involved. Thus, while previous studies suggested overlapping networks for the production of speech and non-speech, the present study showed that the dynamics of the same mental processes differs between planning speech and non-speech sequences. We could also prove that those results are driven by the encoding process disabled through the articulatory suppression task (i.e. motor planning) and not by late-articulatory processes. Further investigation will have to shed light on a functional understanding of the differential involvement of the same brain processes for planning speech and non-speech gestures.

Acknowledgment

Appendix A. Description of non-speech stimuli.

|		
1	Appétisant (Appetizing)	Closing lips producing a voiced nasal sound with variable prosody.
2	Bisou (Kiss)	Voiceless fricative release of protruded lips.
3	Bouchon (Bottle cap)	Explosive release of inverted lips producing a bilabial voiceless click.
4	Colère (Anger)	Friction of the body tongue against the uvula producing a long vibration of both structures.
5	Craquement (Creak)	Voiceless plosion from the friction between the tongue body against the velum followed by a lasting tongue vibration. The sound finishes with the repetition of the first plosion.
6	Dégoût (Disgust)	Voiced bilabial occlusion followed by a vocalic sound with horizontal lip opening and mid-anterior positioning of the tongue.
7	Dentiste (Dentist)	Voiced bilabial occlusion followed by the release and holding of a voiced fricative sound.
8	Explosion (Explosion)	Voiceless bilabial occlusion followed by a sustained velar-lingual vibration.
9	Frisson (Shiver)	Voiced bilabial occlusion followed by a sustained velar-lingual vibration.
10	Galop (Gallop)	Mouth slightly opened, with the apex of the tongue against the alveolar ridge. The tongue is then released to the floor of the mouth, producing a tongue click.
11	Moto (Moto)	Voiced friction from the upper teeth against the lower lip followed by velar-lingual vibration and a bilabial closure that produces a nasal sound.
12	Pneu (Tire)	Voiceless bilabial occlusion followed by a voiceless alveolar fricative release.
13	Poisson (Fish)	Opening mouth releasing a bilabial click.
14	Rapidité (Rapidity)	Voiceless fricative release stopped by the positioning of the tongue in the alveolar region, producing a voiceless plosive release of the air.
15	Ras le bol (Fed up)	Voiceless bilabial occlusion followed by a voiceless labiodental fricative release with a small friction holding.
16	Sèche cheveux (Hair-dryer)	Friction of the air against the edges of the tongue, positioned in the post-alveolar region of the palate.
17	Silence (Silence)	Sustained production of a voiceless post-alveolar fricative sound.
18	Souffle (Blow)	Protruded lips letting the air come out, producing a soft sound from the friction of the air and the oral structures.
19	Soulagement (Relief)	Voiceless bilabial occlusion followed by a voiceless palatal fricative release and lip protrusion.
20	Voiture (Car)	Voiceless friction from the upper teeth against the lower lip followed by a bilabial closure that produces a voiced nasal sound.

Appendix B. Complete list of stimuli.

		Triplets		
	Non-speech stimuli	Speech stimuli		
		Words	Non-Words	Structure
1	Appétisant (Appetizing)	Même /mom/	Moume /mum/	CVC
2	Bisou (Kiss)	Point /pw5/	Pwo /pwo/	CYV
3	Bouchon (Bottle cap)	Pape /pap/	Pépe /pep/	CVC
4	Silence (Silence)	Tchèque /tSEk/	Tchick /tSik/	CCVC
5	Colère (Anger)	Gros /gRo/	Grou /gRu/	CCV
6	Craquement (Creak)	Crique /kRik/	Cruke /kRyk/	CCVC
7	Dégoût (Disgust)	Barque /baRk/	Borke /boRK/	CVCC
8	Dentiste (Dentist)	Bise /biz/	Boze /bOz/	CVC
9	Explosion (Explosion)	Pro /pRo/	Prœu /pR9/	CCV
10	Frisson (Shiver)	Brais /bRE/	Brœu /bR9/	CCV
11	Galop (Gallop)	Glas /gla/	Glo /glo/	CCV
12	Moto (Moto)	Vrai /vRE/	Vru /vRy/	CCV
13	Pneu (Tire)	Psy /psi/	Psu /psy/	CCV
14	Poisson (Fish)	Pope /pOp/	Pobe /pOb/	CVC
15	Rapidité (Rapidity)	Fûte /fyt/	Fudes /fyd/	CVC
16	Ras le bol (Fed up)	Pouf /puf/	Pof /pof/	CVC
17	Sèche cheveux (Hair-dryer)	Chausse /Sos/	Chouf /Suf/	CVC
18	Souffle (Breath)	Fusse /fys/	Fushe /fyS/	CVC
19	Soulagement (Relief)	Pion /pj§/	Piu /pjy/	CYV
20	Voiture (Car)	Fume /fym/	Fême /fEm/	CVC

Appendix C. Results of the topographic consistency test for ERPs of Experiment 1
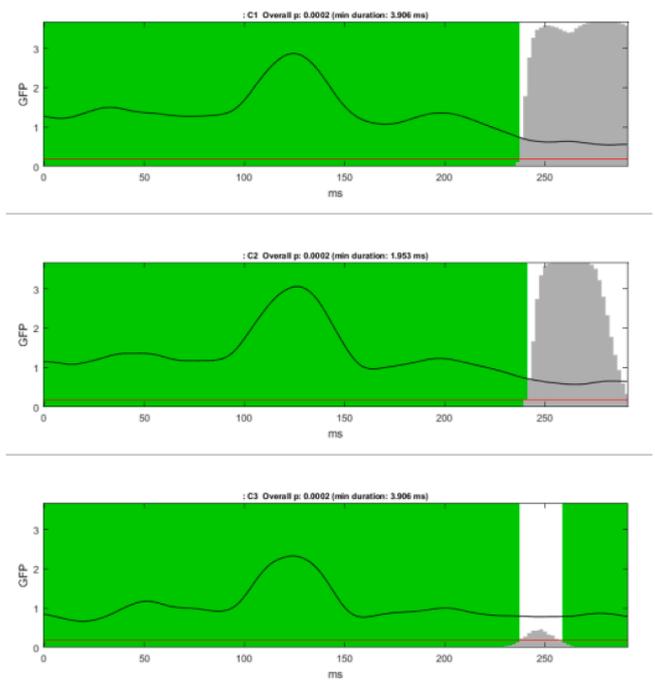
Stimulus-locked ERPs
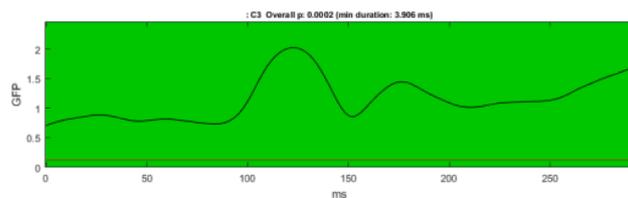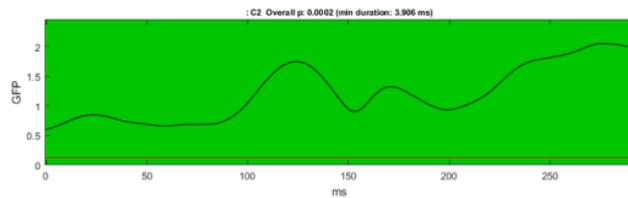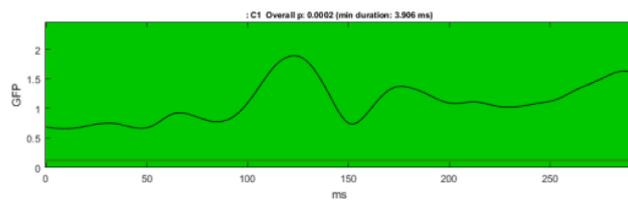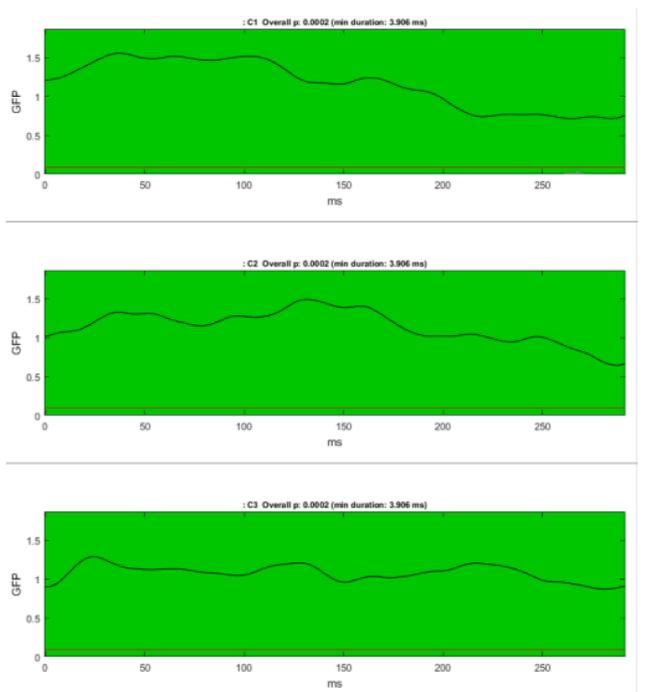

Response-locked ERPs
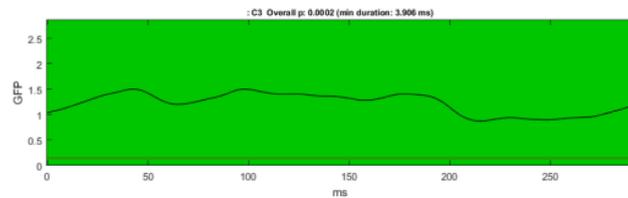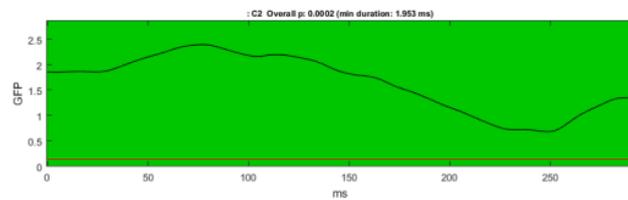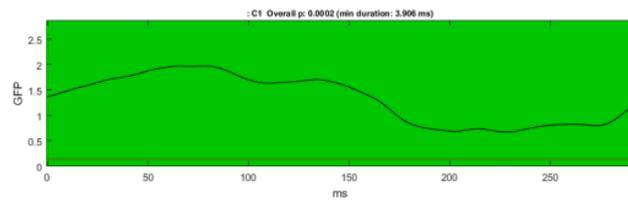
Appendix D. Results of the topographic consistency test for ERPs of Experiment 2

Stimulus-locked ERPs

Response-locked ERPs

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412.

Ballard, K. J., Robin, D. A., & Folkins, J. W. (2003). An integrative model of speech motor control: A response to Ziegler. *Aphasiology, 17*(1), 37–48. https://doi.org/10.1080/02687030244000482.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language, 24*(1), 89–106.

Basilakos, A., Smith, K. G., Fillmore, P., Fridriksson, J., & Fedorenko, E. (2017). Functional characterization of the human speech articulation network. *Cerebral*

Cortex, 28(5), 1816–1830.

Bonilha, L., Moser, D., Rorden, C., Baylis, G. C., & Fridriksson, J. (2006). Speech apraxia without oral apraxia: Can normal brain function explain the physiopathology? *NeuroReport, 17*(10), 1027–1031.

Brunet, D., Murray, M. M., & Michel, C. M. (2011). Spatiotemporal analysis of multi-channel EEG: CARTOOL. *Computational Intelligence and Neuroscience, 2011*, 2.

Bürki, A., Cheneval, P. P., & Laganaro, M. (2015). Do speakers have access to a mental syllabary? ERP comparison of high frequency and novel syllable production. *Brain and Language, 150*, 90–102.

Chang, S.-E., Kenney, M. K., Loucks, T. M., Poletto, C. J., & Ludlow, C. L. (2009). Common neural substrates support speech and non-speech vocal tract gestures. *Neuroimage, 47*(1), 314–325.

Cholin, J., Levelt, W. J., & Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition, 99*(2), 205–235.

Deger, K., Ziegler, W., & Wessel, K. (1999). Airflow tracking in patients with ataxic disorders. *Clinical Linguistics & Phonetics, 13*(6), 433–447.

Goldinger, S. D., Azuma, T., Abramson, M., & Jain, P. (1997). Open wide and say "blah!" Attentional dynamics of delayed naming. *Journal of Memory and Language, 37*(2), 190–216.

Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review, 105*(4), 611.

Horwitz, B., Amunts, K., Bhattacharyya, R., Patkin, D., Jeffries, K., Zilles, K., & Braun, A. R. (2003). Activation of Broca's area during the production of spoken and signed language: A combined cytoarchitectonic mapping and PET analysis. *Neuropsychologia, 41*(14), 1868–1876.

Kami, A., Meyer, G., Jezzard, P., Adams, M. M., Turner, R., & Ungerleider, L. G. (1995). Functional MRI evidence for adult motor cortex plasticity during motor skill learning. *Nature, 377*(6545), 155.

Kawamoto, A. H., Liu, Q., Mura, K., & Sanchez, A. (2008). Articulatory preparation in the delayed naming task. *Journal of Memory and Language, 58*(2), 347–365.

Kemeny, S., Xu, J., Park, G. H., Hosey, L. A., Wettig, C. M., & Braun, A. R. (2005). Temporal dissociation of early lexical access and articulation using a delayed naming task—An FMRI study. *Cerebral Cortex, 16*(4), 587–595.

Kent, R. D. (2015). Nonspeech oral movements and oral motor disorders: A narrative review. *American Journal of Speech-Language Pathology, 24*(4), 763–789.

Koenig, T., Kottlow, M., Stein, M., & Melie-García, L. (2011). Ragu: A free tool for the analysis of EEG and MEG event-related scalp field data using global randomization statistics. *Computational Intelligence and Neuroscience, 2011*, 4.

Koenig, T., & Melie-García, L. (2010). A method to determine the presence of averaged event-related fields using randomization tests. *Brain Topography, 23*(3), 233–242.

Koenig, T., Stein, M., Grieder, M., & Kottlow, M. (2014). A tutorial on data-driven methods for statistically assessing ERP topographies. *Brain Topography, 27*(1), 72–83.

Kupfermann, I., & Weiss, K. R. (2001). Motor program selection in simple model systems. *Current Opinion in Neurobiology, 11*(6), 673–677.

Laganaro, M. (2019). Phonetic encoding in utterance production: a review of open issues from 1989 to 2018. Language, Cognition and Neuroscience, 1–9.

Laganaro, M., & Alario, F.-X. (2006). On the locus of the syllable frequency effect in speech production. *Journal of Memory and Language, 55*(2), 178–196.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 159–174.

Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography and Clinical Neurophysiology, 48*(6), 609–621.

Levelt, W. J. (1993). *Speaking: From intention to articulation, Vol. 1*. MIT press.

Luschei, E. S., & Goldberg, L. J. (1981). Neural mechanisms of mandibular control: Mastication and voluntary biting. *Handbook of Physiology: The Nervous System, 2*, 1237–1274.

Maas, E. (2017). Speech and nonspeech: What are we talking about? *International Journal of Speech-Language Pathology, 19*(4), 345–359.

Memarian, N., Ferrari, P., Macdonald, M. J., Cheyne, D., Luc, F., & Pang, E. W. (2012).

Cortical activity during speech and non-speech oromotor tasks: A magnetoencephalography (MEG) study. *Neuroscience Letters, 527*(1), 34–39.

Michel, C. M., Koenig, T., Brandeis, D., Gianotti, L. R., & Wackermann, J. (2009). *Electrical neuroimaging*. Cambridge University Press.

Michel, C. M., & Murray, M. M. (2012). Towards the utilization of EEG as a brain imaging tool. *Neuroimage, 61*(2), 371–385.

Mock, J. R., Foundas, A. L., & Golob, E. J. (2011). Modulation of sensory and motor cortex activity during speech preparation. *European Journal of Neuroscience, 33*(5), 1001–1011.

Mugler, E. M., Tate, M. C., Livescu, K., Templer, J. W., Goldrick, M. A., & Slutzky, M. W. (2017). Differential representation of articulatory gestures and phonemes in motor, premotor, and inferior frontal cortices. *bioRxiv*, 220723.

Murray, M. M., Brunet, D., & Michel, C. M. (2008). Topographic ERP analyses: A step-by-step tutorial review. *Brain Topography, 20*(4), 249–264.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97–113.

Perrin, F., Pernier, J., Bertnard, O., Giard, M., & Echallier, J. (1987). Mapping of scalp potentials by surface spline interpolation. *Electroencephalography and Clinical Neurophysiology, 66*(1), 75–81.

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.

Rastle, K., Croot, K. P., Harrington, J. M., & Coltheart, M. (2005). Characterizing the motor execution stage of speech production: Consonantal effects on delayed naming latency and onset duration. *Journal of Experimental Psychology: Human Perception and Performance, 31*(5), 1083.

Riecker, A., Ackermann, H., Wildgruber, D., Dogil, G., & Grodd, W. (2000). Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. *NeuroReport, 11*(9), 1997–2000.

Saarinen, T., Laaksonen, H., Parviainen, T., & Salmelin, R. (2005). Motor cortex dynamics in visuomotor production of speech and non-speech mouth movements. *Cerebral Cortex, 16*(2), 212–222.

Salmelin, R., & Sams, M. (2002). Motor cortex involvement during verbal versus non-verbal lip and tongue movements. *Human Brain Mapping, 16*(2), 81–91.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime: User's guide: Psychology Software Incorporated.

Staiger, A., Schölderle, T., Brendel, B., Bötzel, K., & Ziegler, W. (2017). Oral motor abilities are task dependent: A factor analytic approach to performance rate. *Journal of Motor Behavior, 49*(5), 482–493.

Strijkers, K., Holcomb, P. J., & Costa, A. (2011). Conscious intention to speak proactively facilitates lexical access during overt object naming. *Journal of Memory and Language, 65*(4), 345–362.

Tilsen, S., Spincemaille, P., Xu, B., Doerschuk, P., Luh, W.-M., Feldman, E., & Wang, Y. (2016). Anticipatory posturing of the vocal tract reveals dissociation of speech movement plans from linguistic units. *PLoS ONE, 11*(1), e0146813.

Tremblay, P., & Gracco, V. L. (2009). Contribution of the pre-SMA to the production of words and non-speech oral motor gestures, as revealed by repetitive transcranial magnetic stimulation (rTMS). *Brain Research, 1268*, 112–124.

Wildgruber, D., Ackermann, H., Klose, U., Kardatzki, B., & Grodd, W. (1996). Functional lateralization of speech production at primary motor cortex: A fMRI study. *NeuroReport, 7*(15–17), 2791–2795.

Ziegler, W. (2002). Task-related factors in oral motor control: Speech and oral diadochokinesis in dysarthria and apraxia of speech. *Brain and Language, 80*(3), 556–575.

Ziegler, W. (2003a). Speech motor control is task-specific: Evidence from dysarthria and apraxia of speech. *Aphasiology, 17*(1), 3–36.

Ziegler, W. (2003b). To speak or not to speak: Distinctions between speech and nonspeech motor control. *Aphasiology, 17*(2), 99–105.

Ziegler, W., & Ackermann, H. (2013). Neuromotor speech impairment: It's all in the talking. *Folia Phoniatrica et Logopaedica, 65*(2), 55–67.