Cortex

Research Report

Predictive and perceptual phonemic processing in articulatory motor areas: A prediction potential & mismatch negativity study

Luigi Grisoni [a,b,*] *and Friedemann Pulvermüller* [a,b,c,d]

[a] *Freie Universität Berlin, Brain Language Laboratory, Department of Philosophy and Humanities, Berlin, Germany*
[b] *Cluster of Excellence 'Matters of Activity. Image Space Material', Humboldt Universität zu Berlin, Berlin, Germany*
[c] *Berlin School of Mind and Brain, Humboldt Universität zu Berlin, Berlin, Germany*
[d] *Einstein Center for Neurosciences, Berlin, Germany*

ABSTRACT

The recent finding of predictive brain signals preceding anticipated perceptual and linguistic stimuli opens new questions for experimental research. Here, we address the possible brain basis of phonological predictions regarding the features of specific speech sounds and their relationship to phonological priming. To this end, we recorded EEG correlates of both pre- and post-stimulus brain responses in a phonological priming study. Redundant spoken sounds induced stimulus expectations, which manifested as a slow-wave anticipatory activity (the Prediction Potential, PP), whereas articulatory-congruent (e.g.,/bӘ/in the context of expected/pӘ/) pairs elicited weaker post-stimulus MMN-like responses as compared to the articulatory-incongruent (e.g.,/bӘ/in the context of expected/ dӘ/) pairs, a pattern reminiscent of perceptual priming mediated by articulatory-motor areas. Source analysis reveal clusters of activation in lateral prefrontal, temporal and ventral motor areas, thus providing the proof of the relevance of multimodal representation units subserving predictive and perceptual phonemic processing.

1. Introduction

The neurophysiological basis of perceptual predictions has been a target of much recent EEG research, which showed that the expectation of upcoming acoustic stimuli induces cortical pre-activation observable as negative-going potential shifts starting hundreds of milliseconds prior to predictable stimuli (for recent reviews, see Pickering & Gambi, 2018; Pulvermüller & Grisoni, 2020). As it is specific to predictable items, but absent or reduced for less predictable ones (Grisoni, Dreyer, & Pulvermüller, 2016; Grisoni, Mohr, & Pulvermuller, 2019; Grisoni, Moseley, et al., 2019; Kilner, Vargas, Duval, Blakemore, & Sirigu, 2004; Leon-Cabrera, Flores, Rodriguez-Fornells, & Moris, 2019; León-Cabrera, Rodríguez-Fornells, & Morís, 2017), this neurophysiological activity is sometimes

* *Corresponding author.* Brain and Language Laboratory Department of Philosophy and Humanities, WE4 Freie Universität Berlin Habelschwerdter Allee 45 14195 Berlin, Germany
E-mail address: grisoniluigi@zedat.fu-berlin.de (L. Grisoni).

called the Prediction Potential, or PP (Pulvermüller & Grisoni, 2020). The PP can be elicited by a wide range of expected items, including visual (Kilner et al., 2004), auditory (Grisoni, Mohr, et al., 2019), as well as spoken (Grisoni et al., 2016; Grisoni, Miller, & Pulvermüller, 2017; Grisoni, Tomasello, & Pulvermüller, 2020) and written language stimuli (Leon-Cabrera et al., 2019). It emerges in tasks forcing subjects to attend to these stimuli, but it is likewise present when subjects are instructed to ignore the stimuli and focus their attention elsewhere (Pulvermüller & Grisoni, 2020). Some recent studies even revealed brain correlates of quite specific semantic predictions (Grisoni et al., 2017, 2020). In language understanding, the first words of a sentence sometimes make it likely that a particular word will follow after a given initial string fragment. In this case, a semantic PP can be observed and, intriguingly, the specific topography of this potential may index rather fine-grained aspects of the meaning of the expected and most likely upcoming word, for example whether it is a noun referring to an animal or to a tool (Grisoni et al., 2020), or whether it is a verb addressing an action typically performed with the hand or mouth (Grisoni et al., 2017). Similarly, in oddball paradigms, where subjects expect the frequent repetition of a so-called "standard" stimulus so that, within a continuous train of such repeated standards, a PP occurs before each of them and also before the occasionally interspersed rare "deviant" stimuli (Grisoni et al., 2016; Grisoni, Mohr, et al., 2019).

Having a brain index of prediction available may be useful for assessing theories of predictive and perceptual processing and for testing whether cortical pre-activation facilitates stimulus processing, as it is assumed for perceptual priming. The priming effect is a well-known phenomenon according to which the exposition to one (prime) stimulus leads to a quicker and more accurate recognition of a subsequent (target) stimulus, for example when the prime and target stimuli share perceptual and/or semantic features as compared to matched nonprime—target pairs with less or no similarity. The neurophysiological correlate of priming is the relative reduction of the cortical response to the primed (target) stimulus (Bentin, McCarthy, & Wood, 1985; Grisoni et al., 2016; Holcomb & Neville, 1990). That accurate predictions might induce priming is justified by a conceptual consideration: the activations elicited before the target stimulus must somehow (and somewhere) improve stimulus recognition when the target stimulus overlaps with the expectations.

Having available a brain index of prediction might also be of help in revealing the brain areas involved in predictive and perceptual phonemic processing. For example, some theories claim a main role of prefrontal and motor systems in prediction generation (Pickering & Gambi, 2018), and other models hold that perception per se primarily involves posterior perceptual areas (Hickok & Poeppel, 2007). An integrative perspective suggests that circuits distributed across frontocentral (including motor) and posterior (including sensory) areas and interlinking stimulus-related perceptual and action-related information may carry both prediction and perception processes (Pulvermüller & Fadiga, 2010;

Pulvermüller, Tomasello, Henningsen-Schomers, & Wennekers, 2021).

Previous work on phonological processing reported evidence for mappings of the nature of speech sounds or phonemes on cortical activation patterns. For example, a range of studies looked at brain correlates of different speech sounds, mapping their phonological features on the activation of specific cortical areas. Results showed phonological mapping in temporal areas (Arsenault & Buchsbaum, 2015; Mesgarani, Cheung, Johnson, & Chang, 2014; Obleser & Eisner, 2009), and further studies revealed the additional involvement of frontoparietal areas as well, including sensorimotor cortex, both in speech production and perception (Chartier, Anumanchipalli, Johnson, & Chang, 2018; Khoshkhoo, Leonard, Mesgarani, & Chang, 2018; Pulvermüller et al., 2006). In particular, some research pointed to the differential involvement of frontoparietal as well as temporal areas in the processing of specific phonemic types and a mapping of distinctive features of phonemes — for example the place-of-articulation features [labial] versus [coronal] (Evans & Davis, 2015; Pulvermüller et al., 2006; Skipper, Devlin, & Lametti, 2017; Strijkers, Costa, & Pulvermüller, 2017). Although cortical stimulation studies reported some evidence for a functional role of sensorimotor cortex in phoneme perception and language understanding (D'Ausilio et al., 2009; Schomers, Kirilina, Weigand, Bajbouj, & Pulvermüller, 2015; Schomers & Pulvermüller, 2016), some theorists still claim that frontocentral and motor circuits are not of particular relevance for speech perception per se (Stokes, Venezia, & Hickok, 2019).

The current study applied an oddball-like roving paradigm to elicit predictive and mismatch negativity (MMN) brain responses reflecting phonological prediction, perception and priming (Cowan, Winkler, Teder, & Näätänen, 1993). In a roving paradigm, the same stimulus type is presented repeatedly. A train of speech items of the same type was followed by a second train of a different stimulus and so on. As the first stimulus of a train is always unexpected, this first item is called the "deviant" stimulus. The second stimulus and its followers are called "standards". We expect that before standard stimuli, *predictive brain activity* will emerge and that after perception of a novel "deviant stimulus", *perceptual activation* emerges. Similarity between the novel deviant and the previous standard stimulus, which constitutes the context, leads to *priming*.

In the paradigm applied, we used speech sounds characterized by different Places-of-Articulation (PoA), coronal and labial phonemes; of each, both voiced and unvoiced versions were presented. All critical phonemes were stop consonants/bƏ/,/pƏ/,/dƏ/and/tƏ/. A stimulus train included from 2 to 8 identical phonemes, then followed by a different train. Any two successive stimulus trains could thus be either congruent with regard to the place-of-articulation of the phoneme (if both stimulus trains included coronal phonemes, or if both included labials) or incongruent (if there was a change from labial to coronal or vice versa). In this setup, there is priming of phonemic features if a PoA congruent deviant stimulus follows a standard train. If standard train and subsequent deviant are incongruent with regard to their articulation locus, we assume absence of priming related to the PoA feature.

Specifically, we asked here whether:

(i) predictions about the nature of expected speech sounds can be reflected in the predictive potentials emerging before the critical stimuli. Although pre-existing research shows that PPs emerge for a wide range of expected items (see above), there is thus far no indication whether phoneme expectation induce a PP response.

(ii) phonemic predictive processing originates from a representation unit distributed across frontocentral (including articulatory-motor) and posterior areas (see above).

(iii) phonemes with a different place of articulation (PoA) modulate the topographical distribution of the PP and whether this difference also emerges in source space. Specifically, we tested whether the PoA feature brings about different activation patterns in prefrontal and temporal areas (see above).

(iv) whether repetition increases the precision and specificity of the prediction reflected in the PP, so that, after relatively longer trains of standard stimuli more pronounced and more topographically specific PPs could be observed. To address this issue, we considered the position of the stimulus in a train and tested the hypothesis that topographical modulations reflecting the PoA emerge after relatively longer trains of repeated identical phoneme sounds, but not, or less so, in short trains. Specifically, we compared the PP before Early (i.e., 3rd sound in a train) and Late (i.e., 6th, 7th and 8th) standard sounds.

(v) post-stimulus MMN responses are modulated by phonological similarity and show phonological priming. This was tested by considering phoneme pairs with the same or different PoA. We expected smaller MMN-like responses for articulatory-congruent pairs which share the same PoA (e.g.,/bƏ/after a train of/pƏ/sounds) as compared with incongruent pairs (e.g.,/bƏ/in the context of/dƏ/).

(vi) The same pattern of results in signal and source space supposed for the PP (see hypothesis iii) also emerge at post-stimulus latencies (i.e., MMN).

To test hypothesis iv (precision) in conjunction with (v) (priming), we tested whether the potential PoA congruency effects occurred in the same positions (i.e., Early vs Late) where the PP showed PoA driven topographical modulations.

In the context of these questions and predictions, the roving paradigm has one main advantage: it allows to investigate how the activation of the memory trace for the standard sound influences both the subsequent expectations (reflected by the PP prior to the subsequent stimulus) and response to the violation of these expectations (MMN to a deviant stimulus). Therefore, the paradigm was designed so as to always have the same number of deviants (i.e., 16) at each position (i.e., after 2, 3, 4, 5, 6, 7, or 8 standard repetitions). This design enabled us to obtain and separate the brain correlates of phonological prediction and phonological priming and also to investigate potential modulations of both the PP and the MMN responses depending on the phoneme's position (i.e., Early vs Late hypotheses iii and iv; Fig. 1c).

2. Materials and methods

2.1. Participants

The sample size was determined on the basis of a previous MMN and PP study (Grisoni, Mohr, et al., 2019). Twenty-five healthy adults (mean age 25.5 years, range 20–36 years) were tested. Subjects were German native speakers with normal hearing, normal motor control, normal or corrected-to-normal sight and no record of neurological or psychiatric disease. Datasets from 6 participants were excluded, due to technical problems during data acquisition or because of too many (i.e., >30%) rejected trials. Therefore, data from 19 participants (mean age 25.2 years, range 20–31 years; 11 females), all of them right-handed, as determined by the Edinburgh Handedness Inventory (Oldfield, 1971) (mean laterality quotient 77.8 ± 18.9 SD), were included in the final analysis. All participants provided written informed consent and were paid for their participation (12EUR/h). Procedures were approved by the Ethics Committee of Charité Universitätsmedizin, Campus Benjamin Franklin, Berlin, Germany.

2.2. Stimuli, apparatus and experimental design

As stimuli, we chose naturally spoken CV syllables starting with one of four consonants,/bƏ/,/pƏ/,/dƏ/and/tƏ/, followed by a neutral vowel, the 'schwa' sound/Ə/thus yielding/bƏ/,/pƏ/,/dƏ/and/tƏ/; Note that these are all meaningless syllables with stop consonants differing in 1 or 2 phonetic distinctive features, place-of-articulation and voicing. We recorded multiple repetitions of these CV syllables uttered by a male native speaker of German, who had no marked dialectical inflections, and selected a quadruplet that was matched for length (~210 msec), F0 frequency (~340Hz) and
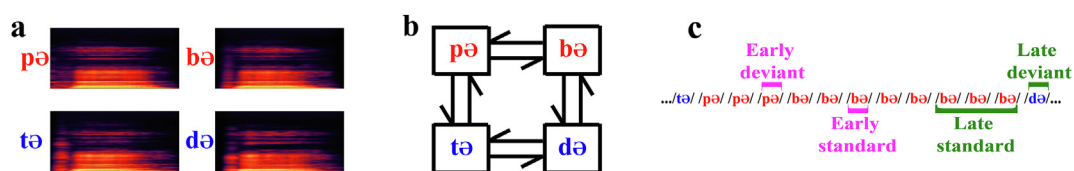

Fig. 1 − a) Spectrograms of the labial (red) and coronal (blue) spoken phonemes stimuli used in this study (from top left clockwise/pƏ//bƏ//tƏ//dƏ/). b) Experimental conditions tested in this study along with (c) schematic representation of the oddball roving paradigm underlying Early (magenta) and Late (green) deviant and standard.

loudness/sound energy (RMS). All four stimuli appeared in a distraction-oddball roving paradigm (Cowan et al., 1993) and were used both as frequently repeated 'standard' stimuli and new 'deviant' stimuli. The EEG experiment lasted for 45 min length plus a short break in the middle.

The four CV syllables including the critical stop consonants constituted a structured sequence in which train of sounds succeeded each other. Within each stimulus train one of the four CV syllables (/bƏ/,/pƏ/,/dƏ/or/tƏ/) was presented binaurally to the participants and it was followed by a train of a different syllable. By convention, the first phoneme of a new train is called the unexpected 'deviant stimulus', which is known to elicit a Mismatch negativity or MMN. The subsequent repetitions of the same stimulus are called 'standard stimuli' and elicit a much smaller brain response compared with the MMN-eliciting first presentation. The length of the trains varied, randomly between 2 and 8 and the probability with which a deviant appeared after 2, 3, 4, 5, 6, 7, or 8 standards was kept constant. Overall, each deviant occurred 112 times, that is, 16 times after each of the standard train lengths (i.e., 2, 3, 4, 5, 6, 7 or 8). The stimulus onset asynchrony (SOA) between any two sounds was 1200 msec. During EEG recording all the acoustic stimuli were presented binaurally to participants through high-quality headphones (Ultrasone HFI-450 S-LOGIC™, Wielenbach, Germany) and the hearing level was determined, per each participant, before the beginning of the roving paradigm. During the EEG recording participants were instructed to watch a silent movie free of human interactions and they were specifically advised not to pay attention to the sounds presented through the headphones. Furthermore, participants were asked to avoid any unnecessary movements during the EEG recording. The experiment took place in the electrically and acoustically shielded chamber of the Brain Language Laboratory at the Freie Universität Berlin. Participants were monitored during the entire EEG session through a camera to ensure that they did not move during the EEG recording. Stimulus presentation and timing was managed by E-prime 2.0.8.90 software (Psychology Software Tools, Inc., Pittsburg, PA, USA) running on a personal computer (PC) located outside the chamber, whereas inside the chamber, a different PC showed the silent movie to participants, who were seated at a distance of approximately 1m from the monitor.

2.3. Electrophysiological recordings and pre-processing

The EEG was recorded with 129 active electrodes (actiCAP system, BrainProducts, Gilching, Germany). The reference electrode was placed at the nose tip, whereas the EOG channels were placed below the left eye and with two electrodes placed on the side of each eye and embedded in the fabric cap (i.e., F9 and F10). During the EEG recording, the EOG channels had the same reference as all the other EEG electrodes. All the impedances were kept lower than 10 kΩ. Electrophysiological data were amplified and recorded using BrainVision Recorder software (version: 1.20.0003; Brain Products GmbH) with a passband of .1–250 Hz and a sampling rate of 1000 Hz. Data were stored on a disk. Offline pre-processing followed standard procedure for ERPs analysis (Luck, 2014). It started with data down-sampling to 512 Hz. Bipolar EOG channels were created as follows, VEOG channel was created by subtracting the Fp1 from the lower eye electrode, whereas the new HEOG channel was created by subtracting the right (i.e., F10) from the left (i.e., F9) eye electrode signal. Therefore, raw data have been epoched in segments that were 4000 msec long (i.e., from −600 msec before sound onset to 3400 msec after sound onset). After data segmentation, independent component analysis (ICA) was carried with EEGLAB 13 (Delorme & Makeig, 2004) on all the 129 electrodes using the default infomax algorithm "runica" (Bell & Sejnowski, 1995). A component was classified artifactual when the peak activity appeared over the horizontal or vertical eye electrodes and when the power spectrum was smoothly decreasing (Delorme & Makeig, 2004). On average, 2.4 (range: 1–4) components out of 129 were removed from each participant's EEG dataset. After artefact correction, off-line data analyses were carried out with BrainVision Analyzer (Brain Products GmbH, Munich, Germany). The electrophysiological signal was offline filtered using a Butterworth zero phase filter with a digital .1 Hz high-pass, a 20 Hz low-pass and a notch filter at 50 Hz (24 dB/oct); these are typical filter settings for both the MMN and the slow brain (i.e., PP) potentials (Kappenman & Luck, 2012). Afterwards, epochs were shortened to different lengths depending on the kind of the trials (i.e., standard or deviant). Standard trials have been segmented from 600 msec before to 1600 msec after phoneme presentation, whereas deviant trials from 100 msec before to 1600 msec after sound onset. In both cases, the first 100 msec of the epoch were used as baseline (i.e., Standard sounds from −600 to −500 msec; Deviant sounds the last 100 msec before stimulus onset) which represent a standard procedure for the two components (Kappenman & Luck, 2012). Epochs with voltage fluctuation of >100 μV at any location and/or >75 μV at EOG and VEOG channels, and those contaminated with amplifier clipping, electromyographic activity, or excessive alpha power were excluded from further analysis. Participants with more than 30% of rejected trials were excluded, overall, 11% of trials were rejected in the dataset finally evaluated.

3. Data analysis

3.1. Standard sounds: pre-stimulus anticipatory activity (PP)

The neurophysiological signals anticipating standard sound onset has been tested as the brain signature of phoneme expectation. The Prediction Potentials (PPs) emerging before standard sounds were analysed as the mean amplitudes (in microvolts) extracted from the last 100 msec before standard sound onset at 9 fronto-central electrodes (average of FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, CP2), where the PP has been previously reported to be largest and, therefore, at these channel locations the best signal-to-noise ratio was expected (Deecke, Scheid, & Kornhuber, 1969; Grisoni, Mohr, et al., 2019). The anticipatory activity before onset of the four syllables was analysed separately for each phoneme and, in a further 'partitioned' analysis, specifically for the Early (i.e., 3rd sound presentation) and Late (i.e., 6th, 7th and 8th sound presentation) positions (to address issue (iv), see

introduction). The Early and Late standard presentations so defined had a similar number of trials. Therefore, the statistical evaluations were carried out for both the whole and the partitioned datasets. First, we determined whether the anticipatory signals were reliable by means of t-tests against zero. Second, potential differences in PP amplitudes from the whole dataset were tested by means of a 2×2 Repeated measures ANOVA with the factors Place-of-articulation (i.e., labial vs coronal) and Phoneme (i.e.,/bƏ/,/pƏ/,/dƏ/and/tƏ/) on the mean amplitudes extracted during the last 100 msec before syllable onset, and by means of a 2×2 x 2 repeated measures ANOVA with the additional factor Position (Early vs Late) on the mean amplitudes extracted from the partitioned dataset.

Furthermore, we performed a second set of analysis to address any possible differences in topographical distributions between conditions, with a focus on putative differences between labial and coronal phonemes (see Introduction, prediction (iii and iv)). To this end, two repeated-measures ANOVAs were performed on the mean amplitudes extracted from a large array of fronto-parietal electrodes (FT7, FC3, FCz, FC4, FT8; T7, C3, Cz, C4, T8; TP7, CP3, CPz, CP4, TP8; P7, P3, Pz, P4, P8; PO7, PO3, POz, PO4, PO8). First, we performed a 2×5 x 5 repeated measures ANOVA with the factors Place-of-articulation (labial vs coronal), Gradient (five levels, frontal-parietal) and Laterality (five levels, left-central-right) on PP mean amplitudes extracted from the whole dataset, then any potential effect of standard sounds' position on PP scalp distribution was tested by means of a 2×2 x 5 $\times 5$ repeated measure ANOVA with the additional factor Position (2 levels, Early vs Late). Since previous observations (Grisoni et al., 2017; Grisoni, Mohr, et al., 2019) reported topographical modulations of the PP activity at the very end of the anticipatory wave we performed this last analysis for both the last 100 msec and 50 msec before syllable onset.

3.2. Source localization. (PP)

Because our main hypothesis was to observe specific activations in articulatory motor areas, it was crucial to estimate the sources underlying both the pre- and post-stimulus presentation. Therefore, distributed cortical sources were estimated using the standard methods implemented in SPM12 (Litvak et al., 2011). Although, this method, as any other method, cannot overcome the non-uniqueness of the inverse problem (Helmholtz, 1853), it uses established priors which have been extensively used in cognitive experiments and it has been shown to produce plausible source solutions. We used the structural MRI included in SPM to create a cortical mesh of 8196 vertices. The electrode cap space was, then, co-registered using three fiducials: Cz, TP9 and TP10. The volume conductors were constructed with an EEG (three-shell) boundary element model. The PP responses (i.e., subject averages) were inverted using the multiple sparse prior technique (i.e., the "Greedy Search" algorithm) at the group level (Friston et al., 2008; Litvak & Friston, 2008). Since our expectation was to observe articulatory-motor activations before predictable, standard, phoneme, we extracted the activation maps for each syllable including the crucial phonemes (i.e.,/bƏ/,/pƏ/,/dƏ/,/tƏ/) for both the Early and Late trials thus resulting in

eight images per participant. Images were smoothed using a Gaussian kernel of full-width at half-maximum (FWHM) of 16 mm. Given the well-known limited spatial resolution of the EEG technique, we did not focus the analysis on the comparison between labial and coronal phonemes whose motor representations are too close. Therefore, we focused on the sources underlying the PP elicitation by comparing the activations maps extracted from the last 50 msec before phoneme presentation with the activation maps extracted from the baseline. Indeed, this comparison is sufficient to confirm the involvement of articulatory (face) motor areas promoting phonemes' expectation. To this end, first we collapsed the four phonemes (i.e.,/bƏ/,/pƏ/,/dƏ/,/tƏ/) and the two positions (i.e., Early vs Late) and we computed the differences between the sources extracted from the last 50 msec before phoneme presentation and the sources extracted from the baseline, then the same strategy was also applied to the Early and Late standard trials and, finally, activity underlying labial and coronal phonemes expectations were also tested. Therefore, five voxel-by-voxel t-tests against zero were performed to reveal the sources underlying the PP emerging before all the standard trials, before Early and Late stimuli and before labial and coronal phonemes. These statistical evaluations were performed on both the whole brain and on regions of interest (ROIs) located on the left hemisphere. The ROIs were created with WFU_PickAtlas (Maldjian, Laurienti, Kraft, & Burdette, 2003) and included traditional language areas, as well as somatosensory, auditory and motor areas (Brodmann areas: 1, 4, 6, 21, 22, 39, 40, 41, 42, 44, 45, and 46). All these regions were merged in one mask image.

For all the t-tests against zero on whole brain and ROIs p values were thresholded at $p = .05$ and corrected for multiple comparisons with the Family-wise error (FWE) procedure, only those clusters >30 voxels were considered.

3.3. Post-deviant MMN-like potential

Phoneme-related negative-going potential responses following deviant stimuli were analysed as mean amplitudes extracted from 9 fronto-central electrodes (i.e., average of FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, CP2) which are the typical MMN electrodes. As in previous reports (Grisoni et al., 2016; Grisoni, Mohr, et al., 2019), the deviant-related responses were tested from the unsubtracted data. As the MMN is normally obtained by subtracting the standard response from that of the deviant, this deviant-elicited potential includes the MMN; therefore, we speak of a 'MMN-like' response. More specifically, the MMN-like response was defined as the mean amplitude in the 50 msec time window centered at 120 msec after syllable onset (i.e., time range 95—145 msec). This latency represents the local maximum latency of the grand average obtained by collapsing all the deviants.

First, we tested for effects of consonant voicing and devoicing on the MMN-like responses. Consonant voicing was tested collapsing the MMN-like responses elicited by the phonemes/bƏ/presented in the context of/pƏ/and/dƏ/presented in the context of standard/tƏ/sounds, whereas consonant devoicing consisted of the MMN-like responses elicited by /pƏ/ in the context of /bƏ/ and /tƏ/ in the context of /dƏ/ sounds. Potential effects of consonant voicing and devoicing

so defined were tested by means of a *t*-test comparing these two MMN-like responses from the whole dataset and by means of a 2 × 2 repeated measures ANOVA with the factors Sound change (voicing *vs* devoicing) and Position (Early *vs* Late) for the partitioned dataset.

Then, potential effects of phoneme and context were assessed with a 2 × 2 repeated-measures ANOVA with the factors Place-of-Articulation (*lips vs tongue*) and Context (*lips vs tongue*) for the whole dataset and with a 2 × 2 x 2 repeated-measures ANOVA with the additional factor Position (Early *vs* Late) for the partitioned dataset. Early responses were defined as the deviants occurring after 2, 3 and 4 standard presentations, whereas the Late MMN-like responses were those occurring after 6, 7 and 8 standard presentations. The deviant presented after 5 standard presentations was excluded from the analysis as it was (n)either Early (n)or Late. Early and Late responses had the same number of trials. Potential topographical differences between conditions, were further tested at the larger array of fronto-parietal electrodes (see above). Therefore, two additional five-levels factors (i.e., Gradient: frontal-parietal; Laterality: left-central-right) were integrated in the previous repeated measures ANOVA thus resulting in a new 2 × 2 x 2 × 5 x 5 design. Finally, to be consistent with the MMN litterature, we have also performed the statistical evaluations with the canonical, that is subtracted, MMN responses. To this end we calculated the MMN by subtracting the ERP elicited by the immediately preceding standard sounds from the subsequent deviant related responses.

3.4. Source localization (MMN-like)

To test whether the enlargement of the MMN-like signal in response to articulatory incongruent conditions was due to activity in articulatory motor areas we computed the sources also at MMN-like latency. To this end, we apply the same methods used for the PP responses. However, since our main interest here was to test where the difference between congruent and incongruent conditions emerged, we first subtracted the ERP response elicited by the congruent conditions from the incongruent ones and then, we calculated the sources underlying this new signal. Furthermore, since the MMN-like expected modulations appeared at Late but not at Early trials, we focused exclusively on Late trials. The sources obtained from the ERPs obtained as a difference between the Incongruent and Congruent conditions were, then averaged across phoneme type (i.e., labial and coronal) to ensure a better signal to noise ratio. This procedure resulted in one image per participant which was then submitted to a *t*-test against zero. *T*-test against zero was performed both at whole brain and at ROIs. However, since the hypothesis was to observe an overlap of responses between the Late PP and the Late MMN-like responses, three ROIs were defined based on the results of source analysis of the Late PP regions. Therefore, the ROIs were defined as the three significant clusters which survived the FWE correction for the Late PP ROIs analysis (i.e., Pre and postcentral left gyrus: x = −52, y = −16, z = 34 number of voxels: 1419; Inferior left prefrontal cortex: x = −40, y = 48, z = 12 number of voxels: 521; Superior temporal cortex: x = −56, y = −32, z = 14 number of voxels: 169, see last raw of

Table 2). As for the previous analysis, also for this *t*-test against zero on whole brain and ROIs *p* values were thresholded at *p* = .05 and corrected for multiple comparisons with the Family-wise error (FWE) procedure, only those clusters larger 20 voxels were considered.

For all the ANOVAs, Partial eta-squares (ηp^2) are reported as indices of effect size (Cohen, 1973). Furthermore, to all the significant main effects and interactions with factors involving more than two levels Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) was applied and corrected *p* values along with epsilon (ε) values were reported when sphericity violations were detected. Post hoc analyses were conducted by means of Fisher Least Significant Difference (LSD) test which was then corrected for multiple comparisons by multiplying the uncorrected *p* value by the number of relevant comparisons (Bonferroni correction). All the *p* values reported with the wording "Bonferroni corrected" refers to the *p* value already corrected for (i.e., already multiplied by) the number of comparisons.

4. Results

4.1. Standard sounds: pre-stimulus anticipatory activity (PP)

One expectation of this study was to observe anticipatory brain activity emerging before phoneme presentation (hypothesis i). Consistent with this hypothesis, before standard sound onset, both labial and coronal phonemes showed a slow-wave potential whose features (i.e., negative polarity, latency and slowly growing shape) were consistent with the Prediction Potential previously reported (Pulvermüller & Grisoni, 2020) (see Fig. 2a—e). The PPs for all four phonemes from the whole dataset at fronto-central electrodes were reliable as confirmed by the *t*-tests against zero (/bƏ/: t = −3.49, Bonferroni corrected *p* = .012;/pƏ/: t = −3.61, Bonferroni corrected *p* = .008;/dƏ/: t = −2.79, Bonferroni corrected *p* = .048;/tƏ/: t = −3.57, Bonferroni corrected *p* = .008) (hypothesis i confirmed). Similarly, the PPs emerging before standard stimuli appearing "Early" and "Late" in the standard train sequences (see Methods and Fig. 1e) were both reliable (Early PPs: t = −2.46, Bonferroni corrected *p* = .048; Late PPs: t = −3.66, Bonferroni corrected *p* = .004) (hypothesis i confirmed). The repeated measures ANOVA on the mean amplitudes extracted from the whole and partitioned dataset from the last 100 msec before sound onset did not show any significant main effects or interactions of the factors Place-of-Articulation (i.e., labial *vs* coronal), Phonemes (i.e.,/bƏ/,/pƏ/,/dƏ/and/tƏ/) and Position (i.e., Early *vs* Late), thus suggesting that the PPs recorded at fronto-central electrodes were not affected by any of these factors (hypothesis iii not supported).

Another important expectation of this study was to observe modulations of the topographical distribution of the PP mediated by the PoA (hypothesis iii). To test this hypothesis, we restricted the analysis to the mean amplitudes extracted from the last 50 msec before (Early and Late) phoneme presentations. This choice was determined by the well-known dynamic of activations underlying the anticipatory ERPs − including Readiness Potential (RP) and PP − which sharpen towards their

Table 1 – Results: ANOVAs. The table displays the significant main effects and interactions of the PP and the MMN-like responses. The F values, the *p* values, and the Partial Eta Squared (ηp^2) are reported.

STANDARD STIMULI: PP	Significant effects	F-value	p-value	Effect size
Mean amplitudes from the last 100 msec before phoneme presentations (Early and Late trials collapsed): 2 × 2 x 5 × 5 Factors: Place-of-Articulation (Labial *vs* Coronal) Position (Early *vs* Late) Gradient (Frontal-Parietal, 5 levels) Laterality (Left-Central-Right, 5 levels)	Interaction of Gradient and Laterality	$F_{16,288} = 3.7$	Adjusted $p = .006$	$\eta p^2 = .17$
Mean amplitudes from the last 50 msec before phoneme presentations (Early and Late trials collapsed): 2 × 2 x 5 × 5 Factors: Place-of-Articulation (Labial *vs* Coronal) Position (Early *vs* Late) Gradient (Frontal-Parietal, 5 levels) Laterality (Left-Central-Right, 5 levels)	Main effect of Gradient Main effect of Laterality Interaction of Gradient and Laterality Interaction of Place-of-Articulation, Gradient and Laterality	$F_{4,72} = 4.1$ $F_{4,72} = 3.5$ $F_{16,288} = 3.6$ $F_{16,288} = 2.3$	Adjusted $p = .046$ Adjusted $p = .047$ Adjusted $p = .006$ Adjusted $p = .048$	$\eta p^2 = .18$ $\eta p^2 = .16$ $\eta p^2 = .17$ $\eta p^2 = .11$
Late trials: Mean amplitudes from the last 50 msec before phoneme presentations (Early and Late trials collapsed): 2 × 5 x 5 Factors: Place-of-Articulation (Labial *vs* Coronal) Gradient (Frontal-Parietal, 5 levels) Laterality (Left-Central-Right, 5 levels)	Main effect of Gradient Interaction of Gradient and Laterality Interaction of Place-of-Articulation, Gradient and Laterality	$F_{4,72} = 4.4$ $F_{16,288} = 2.7$ $F_{16,288} = 2.1$	Adjusted $p = .033$ Adjusted $p = .024$ Adjusted $p = .062$	$\eta p^2 = .20$ $\eta p^2 = .13$ $\eta p^2 = .10$
Last trials: Mean amplitudes from the last 50 msec before phoneme presentations (Early and Late trials collapsed): 2 × 5 x 5 Factors: Place-of-Articulation (Labial *vs* Coronal) Gradient (Frontal-Parietal, 5 levels) Laterality (Left-Central-Right, 5 levels)	Interaction of Place-of-Articulation, Gradient	$F_{4,72} = 3.8$	Adjusted $p = .05$	$\eta p^2 = .17$
DEVIANT STIMULI: MMN-like				
Whole Dataset 2 × 2 Factors: Place-of-Articulation (Labial *vs* Coronal) Context (Labial *vs* Coronal)	Interaction of Place-of-Articulation, Context	$F_{1,18} = 5.2$	$p = .03$	$\eta p^2 = .22$
Partitioned Dataset 2 × 2 x 2 Factors: Position (Early *vs* Late) Place-of-Articulation (Labial *vs* Coronal) Context (Labial *vs* Coronal)	Interaction of Position, Place-of-Articulation, Context	$F_{1,18} = 11.2$	$p = .003$	$\eta p^2 = .38$
Partitioned Dataset 2 × 2 x 2 × 5 x 5 Factors: Position (Early *vs* Late) Place-of-Articulation (Labial *vs* Coronal) Context (Labial *vs* Coronal) Gradient (Frontal-Parietal, 5 levels) Laterality (Left-Central-Right, 5 levels)	Main effect of Gradient Main effect of Laterality Interaction of Gradient and Laterality Interaction of Place-of-Articulation and Gradient Interaction of Position, Place-of-Articulation and Context Interaction of Position, Place-of-Articulation, Context and Laterality	$F_{4,72} = 31.6$ $F_{4,72} = 11.4$ $F_{16,288} = 3.5$ $F_{4,72} = 6.2$ $F_{1,18} = 10.9$ $F_{4,72} = 7.3$	Adjusted $p < .001$ Adjusted $p < .001$ Adjusted $p < .01$ Adjusted $p = .02$ $p = .004$ Adjusted $p = .001$	$\eta p^2 = .64$ $\eta p^2 = .39$ $\eta p^2 = .16$ $\eta p^2 = .25$ $\eta p^2 = .38$ $\eta p^2 = .29$

Table 1 − *(continued)*

STANDARD STIMULI: PP	Significant effects	F-value	p-value	Effect size
Early trials: 2 × 2 x 5 × 5 Factors: Place-of-Articulation (Labial *vs* Coronal) Context (Labial *vs* Coronal) Gradient (Frontal-Parietal, 5 levels) Laterality (Left-Central-Right, 5 levels)	Main effect of Gradient Main effect of Laterality Interaction of Place-of-Articulation and Gradient Interaction of Gradient and Laterality	$F_{4,72} = 32.1$ $F_{4,72} = 8.6$ $F_{4,72} = 6.0$ $F_{16,288} = 3.9$	Adjusted $p < .001$ Adjusted $p < .001$ Adjusted $p = .02$ Adjusted $p = .003$	$\eta p^2 = .64$ $\eta p^2 = .32$ $\eta p^2 = .25$ $\eta p^2 = .18$
Late trials: 2 × 2 x 5 × 5 Factors: Place-of-Articulation (Labial *vs* Coronal) Context (Labial *vs* Coronal) Gradient (Frontal-Parietal, 5 levels) Laterality (Left-Central-Right, 5 levels)	Main effect of Gradient Main effect of Laterality Interaction of Place-of-Articulation and Context Interaction of Place-of-Articulation, Context and Laterality	$F_{4,72} = 24.6$ $F_{4,72} = 8.3$ $F_{1,18} = 9.7$ $F_{4,72} = 5.2$	Adjusted $p < .001$ Adjusted $p = .001$ Adjusted $p = .006$ Adjusted $p = .01$	$\eta p^2 = .58$ $\eta p^2 = .32$ $\eta p^2 = .35$ $\eta p^2 = .22$

ends revealing surprising local specificity (Grisoni, Mohr, et al., 2019; Pulvermüller & Grisoni, 2020; Shibasaki & Hallett, 2006). To this end, the mean amplitudes from the last 50 msec before sound onset were analysed by means of a 2×2 x 5×5 repeated measures ANOVA with the factors Place-of-Articulation (i.e., labial *vs* coronal), Position (Early *vs* Late), Gradient (five levels, frontal-parietal) and Laterality (five levels, left-central-right). Results revealed main effects of Gradient ($F_{4,72} = 4.1$, adjusted $p = .046$, $\varepsilon = .33$, $\eta p^2 = .18$), Laterality ($F_{4,72} = 3.5$, adjusted $p = .047$, $\varepsilon = .48$, $\eta p^2 = .16$), and the interaction of the factors Gradient and Laterality ($F_{16,288} = 3.6$, adjusted $p = .006$, $\varepsilon = .40$, $\eta p^2 = .17$). Interestingly, the topography of the PP was modulated by the Place-of-Articulation (i.e., labial *vs* coronal) at this shorter 50 msec time window (but not at the longer 100 msec time window, see Table 1) immediately before sound onset as documented by the interaction of the factors Place-of-Articulation, Gradient and Laterality ($F_{16,288} = 2.28$, adjusted $p = .048$, $\varepsilon = .50$, $\eta p^2 = .11$). Therefore, the place of articulation of the standard phoneme modulated the PP topography (hypothesis iii confirmed). Assuming that subjects tend to expect and implicitly predict the continuation of a monotonous standard stimulus train, it becomes possible to relate these PP dynamics to anticipation and prediction. As mentioned, some accounts (Schomers & Pulvermüller, 2016) predict different mechanisms for speech sounds with different place-of-articulation and are therefore consistent with the obtained results.

We further tested whether the topographical modulations were affected by the position of the stimulus in a sequence. Specifically, we expected that activity in cortical, including phonological representations accumulates with larger numbers of repetitions of the stimulus, so that sharper (i.e., more precise) and more distinct activation patterns emerge in Late than Early trials (hypothesis iv). To address this issue, we split the same ANOVA to see whether the observed interaction effects of the factors Place-of-Articulation (labial *vs* coronal), Gradient (five levels, frontal-parietal) and Laterality (five levels, left-central-right) interaction was equally present at both Early and Late positions. Whereas we did not observe any significant effect at Early trials, we observed a marginally significant interaction of these factors when investigating the anticipatory PP of late trials in the standard sequences ($F_{16,288} = 2.08$, adjusted $p = .06$, $\varepsilon = .57$, $\eta p^2 = .10$) (see Table 1 for further results). Although we do not wish to strongly interpret such a non-significant result, this pattern, together with the pooled results across all standards taken together, may indicate that the latter significant effect is primarily carried by the anticipatory signals emerging before late standard stimuli within long standard trains. This hypothesis was further confirmed by a separate analysis of only the Last standard position (i.e., 8th presentations of a given sound). Despite the small number of trains available for this analysis (N = 16 per each phoneme), there was a significant interaction of the factors Place-of-Articulation and Gradient ($F_{4,72} = 3.77$, adjusted $p = .05$, $\varepsilon = .37$, $\eta p^2 = .17$) (see Table 1). Overall, this pattern is consistent with the hypothesis of a sharper active memory at Late than at Early standard presentations (hypothesis iv; see also **Discussion**).

4.2. *Source analysis: pre-standard anticipatory activity (PP)*

The current study was aimed at investigating also whether articulatory motor activations take place when participants predict or perceive phonemes. Therefore, the main aim of this source analysis was to assess any possible activation of articulatory motor areas underlying the PP responses (hypothesis ii) and whether phonemes with a different PoA activates distinct articulatory motor and superior temporal sub-areas (hypothesis iii).

First, we contrasted the neuronal sources preceding expected phonemes during the last 50 msec before standard sound onset against the baseline. The whole-brain analysis PP > baseline (Phoneme type, that is labial and coronal, and Position, that is Early and Late, here collapsed) revealed significant clusters in left and right motor areas which, due to their ventral position, are consistent with face motor and somatosensory areas. Furthermore, additional significant clusters were observed in left and right inferior prefrontal areas and in right inferior temporal-occipital regions (see Table 2 and Fig. 2); a pattern overall consistent with hypothesis ii. The same contrast performed on Early and Late trials (Phoneme type, labial and coronal, collapsed), however, showed a slightly different pattern of results. Indeed, whereas Early trials showed clusters of activations in somatosensory and motor ventral, face, motor areas along with the involvement of right prefrontal areas (see Table 2 and Fig. 2), the Late

Table 2 – Results: Source analysis. The table displays the significant clusters of the PP and the MMN-like responses. The MNI coordinates of the voxel with highest t value, its t value, the number of significant voxels per each significant cluster, the Brodmann area labels and a description of the area in which the cluster ("peak voxel") was observed, are reported.

STANDARD SOUNDS: PP	x	y	z	t-values (peak-level)	Number of voxels	Brodmann Areas	Cortical Areas
Whole brain: whole dataset PP > Baseline	−48	−16	34	7.72	3251	1/4	Pre and postcentral left gyrus
t-test against zero	52	−18	36	7.42	3508	1/4	Pre and postcentral right gyrus
p < .05 FWE	38	48	20	6.80	3354	45/46	Inferior right prefrontal cortex
	−44	18	30	6.20	1280	44	Inferior left prefrontal cortex
	36	−68	−18	5.00	723	19/37	Posterior temporal cortex
	−34	50	12	4.57	79	46	Inferior left prefrontal cortex
ROIs: whole dataset PP > Baseline t-test	−48	−14	34	7.68	2901	1/4	Pre and postcentral left gyrus
against zero	−44	18	30	6.20	718	44	Inferior left prefrontal cortex
p < .05 FWE							
Whole brain: Early trials PP > Baseline t-	40	40	20	6.36	1425	45	Inferior right prefrontal cortex
test against zero	−46	−18	36	6.16	2384	1/4	Pre and postcentral left gyrus
p < .05 FWE	50	−18	38	5.83	2342	1/4	Pre and postcentral right gyrus
ROIs: Early trials PP > Baseline t-test	−48	−18	38	6.16	2111	1/4	Pre and postcentral left gyrus
against zero	−44	18	30	4.19	52	44	Inferior left prefrontal cortex
p < .05 FWE							
Whole brain: Late trials PP > Baseline t-test	50	−18	34	6.58	2439	1/4	Pre and postcentral right gyrus
against zero	−48	−16	32	6.48	2013	1/4	Pre and postcentral left gyrus
p < .05 FWE	38	48	18	5.82	1628	45	Inferior right prefrontal cortex
	−36	46	12	5.18	749	44/45	Inferior left prefrontal cortex
	56	−34	14	4.65	35	42	Right superior temporal cortex
ROIs: Late trials PP > Baseline t-test	−48	−14	32	6.56	1846	1/4	Pre and postcentral left gyrus
against zero	−40	48	12	5.19	705	45/46	Inferior left prefrontal cortex
p < .05 FWE	−54	−32	14	4.38	335	42	Left superior temporal cortex
Whole brain: Labial PP > Baseline t-test	50	−18	38	6.23	2450	1/4	Pre and postcentral right gyrus
against zero	−44	−14	42	6.02	2019	1/4	Pre and postcentral left gyrus
p < .05 FWE	46	36	16	5.83	1764	45/46	Inferior right prefrontal cortex
	−42	20	30	4.82	424	44/45	Inferior left prefrontal cortex
ROIs: Labial PP > Baseline t-test against	−44	−14	44	6.05	1980	1/4	Pre and postcentral left gyrus
zero	−42	28	26	5.18	685	44/45	Inferior left prefrontal cortex
p < .05 FWE							
Whole brain: Coronal PP > Baseline t-test	36	46	18	6.32	1498	45/46	Inferior right prefrontal cortex
against zero	−48	−20	32	5.16	1372	1/4	Pre and postcentral left gyrus
p < .05 FWE	52	−20	34	5.03	1250	1/4	Pre and postcentral right gyrus
ROIs: Coronal PP > Baseline t-test against	−48	−14	30	5.16	1941	1/4	Pre and postcentral left gyrus
zero	−44	18	28	4.57	117	44	Inferior left prefrontal cortex
p < .05 FWE							
DVIANTS LATE INCONGRUENT - CONGRUENT: MMN-like	x	y	z	t-values (peak-level)	Number of voxels	Brodmann Areas	Cortical Areas
Whole brain: Incongruent - Congruent t-	−54	−16	−22	7.27	8963	19/37	Posterior temporal left cortex
test against zero	46	−66	−6	7.04	7933	20/37	Inferior temporal right cortex
p < .05 FWE	−60	−14	22	6.70	4619	43	Left superior temporal & Pre
	−56	−36	22	6.30		1/4	and postcentral gyrus
	50	32	8	6.24	4593	45/46	Inferior right prefrontal cortex
	−20	4	56	5.58	744	6	Left Dorsal prefrontal cortex
	−38	46	0	5.37	617	45/46	Inferior left prefrontal cortex
	−42	−70	28	4.67	66	39	Left temporo-parietal cortex
ROIs: Incongruent - Congruent t-test	−58	−28	16	8.58	335	42	Left superior temporal cortex
against zero	−44	−28	48	8.15	1846	1/4	Left pre and postcentral gyrus
p < .05 FWE	−56	−8	32	7.76			
	−44	48	4	5.29	705	44/45/46	Inferior prefrontal cortex
	−46	24	22	4.85			
	−42	56	8	4.64			

trials showed additional activity in left prefrontal areas and ROI analysis further showed activity also in superior temporal regions, in proximity of primary auditory areas. Finally, both labial and coronal phonemes (Position, Early and Late, collapsed) showed similar cluster of activations in face related motor and prefrontal areas (see Fig. 2 and Table 2) (hypothesis ii supported; hypothesis iii not confirmed).

4.3. Post-deviant stimulus potentials: MMN-like responses

Consonant voicing did not affect the MMN-like amplitudes as revealed both by the t-test and the repeated measures ANOVA testing the MMN-like modulations for the whole and the partitioned dataset, respectively. Vice versa, consistent with

hypothesis v, the 2 × 2 repeated measures ANOVA with the factors Place-of-Articulation (i.e., labial and coronal) and Context (i.e., labial and coronal) carried out on the whole dataset revealed that the MMN-like responses were affected by the places of articulation of each stimulus initial consonant and its context (see Fig. 3) as documented by the significant interaction of the factors Place-of-Articulation and Context ($F_{1,18} = 5.22$, $p = .03$, $\varepsilon = 1$, $\eta p^2 = .22$). Although both the labial and coronal phonemes showed smaller averaged MMN-like responses in place-of-articulation congruent (i.e., labial in the context of labial and coronal in the context of coronal) as compared to the incongruent (i.e., labial in the context of coronal and coronal in the context of labial) conditions (see Fig. 3), the planned comparisons did not show statistically significant double dissociation (hypothesis v not fully supported).

The same analysis carried out with the partitioned dataset revealed an interaction of the factors Position, Place-of-Articulation and Context ($F_{1,18} = 11.25$, $p = .003$, $\varepsilon = 1$, $\eta p^2 = .38$). Here, planned comparisons revealed the expected priming-like modulation for deviants appearing after long (Late) standard sequences, but not (at Early position) after short ones. Particularly, after 6, 7 and 8 labial standard sounds the MMN-like responses elicited by labial deviants were smaller as compare to the MMN-like responses elicited by the same sound after 6, 7 and 8 coronal standard sounds (corrected $p = .034$) and, vice versa, the MMN-like responses elicited by the coronal phonemes were smaller after 6, 7 and 8 coronal standard sound as compared to the condition in which the same phonemes were presented in the context of labial phonemes (corrected $p = .045$) (Fig. 3e). Importantly, when the MMN was computed as the difference between the
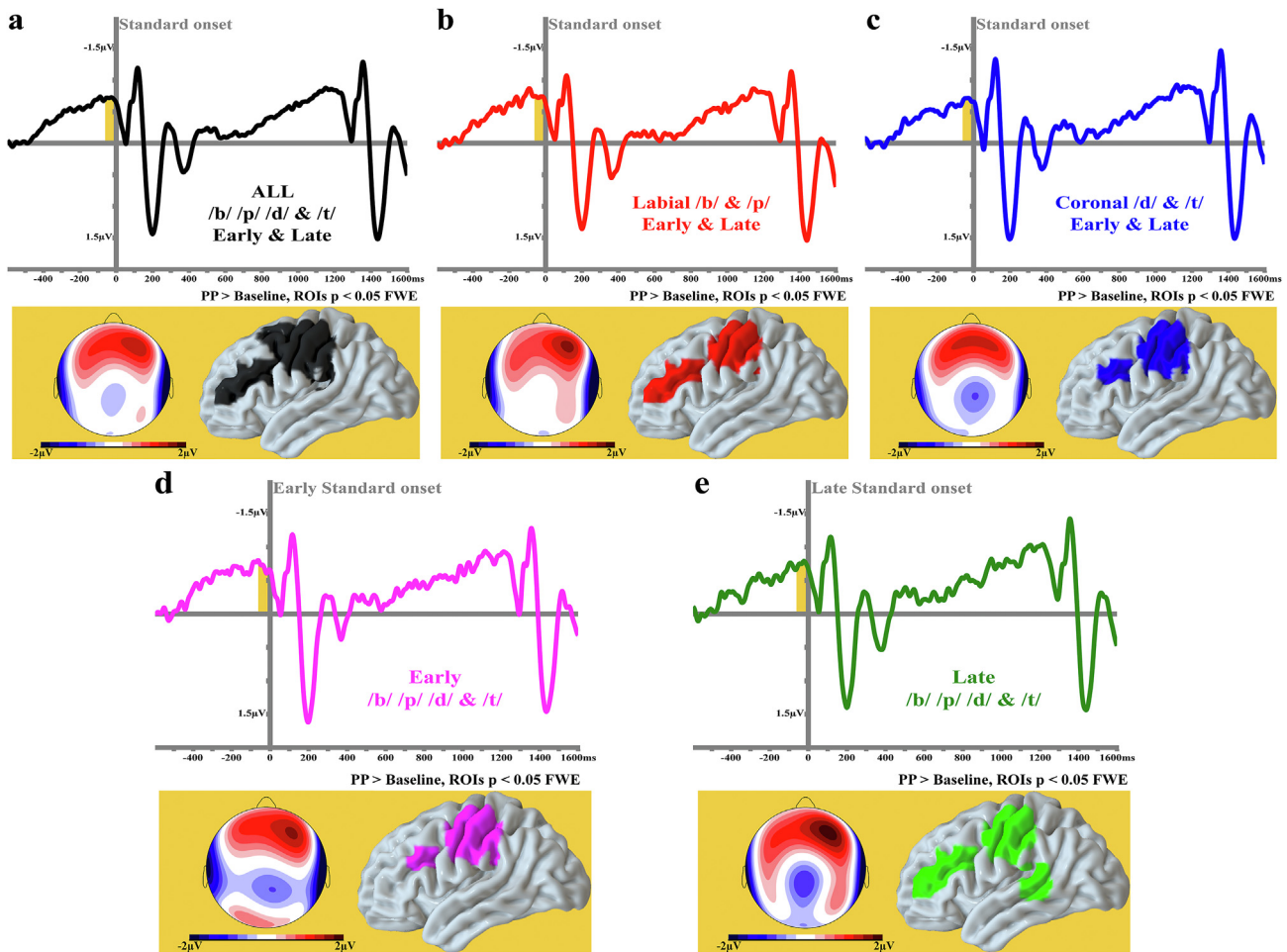

Fig. 2 − PP responses to spoken phonemes. The baseline period for the ERPs was from −600 msec to −500 msec before stimulus onset (i.e., the first 100 msec of the epochs); all the ERPs showed in this picture were extracted from fronto-central left electrodes (i.e., average of: F3, F5, F7, FC3, FC5, FT7, C3, C5, T7, CP3, CP5, TP7). The shaded yellow areas highlight the last 50 msec before stimulus onset; this time window was used for the voltage maps and to compute source estimations presented below each ERP. a) PP observed before all the standard sounds (labial and coronal phonemes and Early and Late positions collapsed) along with their voltage map and sources (PP > Baseline, $p < .05$ FWE). PP observed before labial (b) and coronal (c) standard sounds onset (Early and Late positions collapsed) along with their sources (PP > baseline, $p < .05$ FWE). PP observed before Early (d) and Late (e) sounds onset (labial and coronal phonemes collapsed) along with their sources (PP > baseline, $p < .05$ FWE). All the PP signals here reported were recorded from fronto-lateral left electrodes.
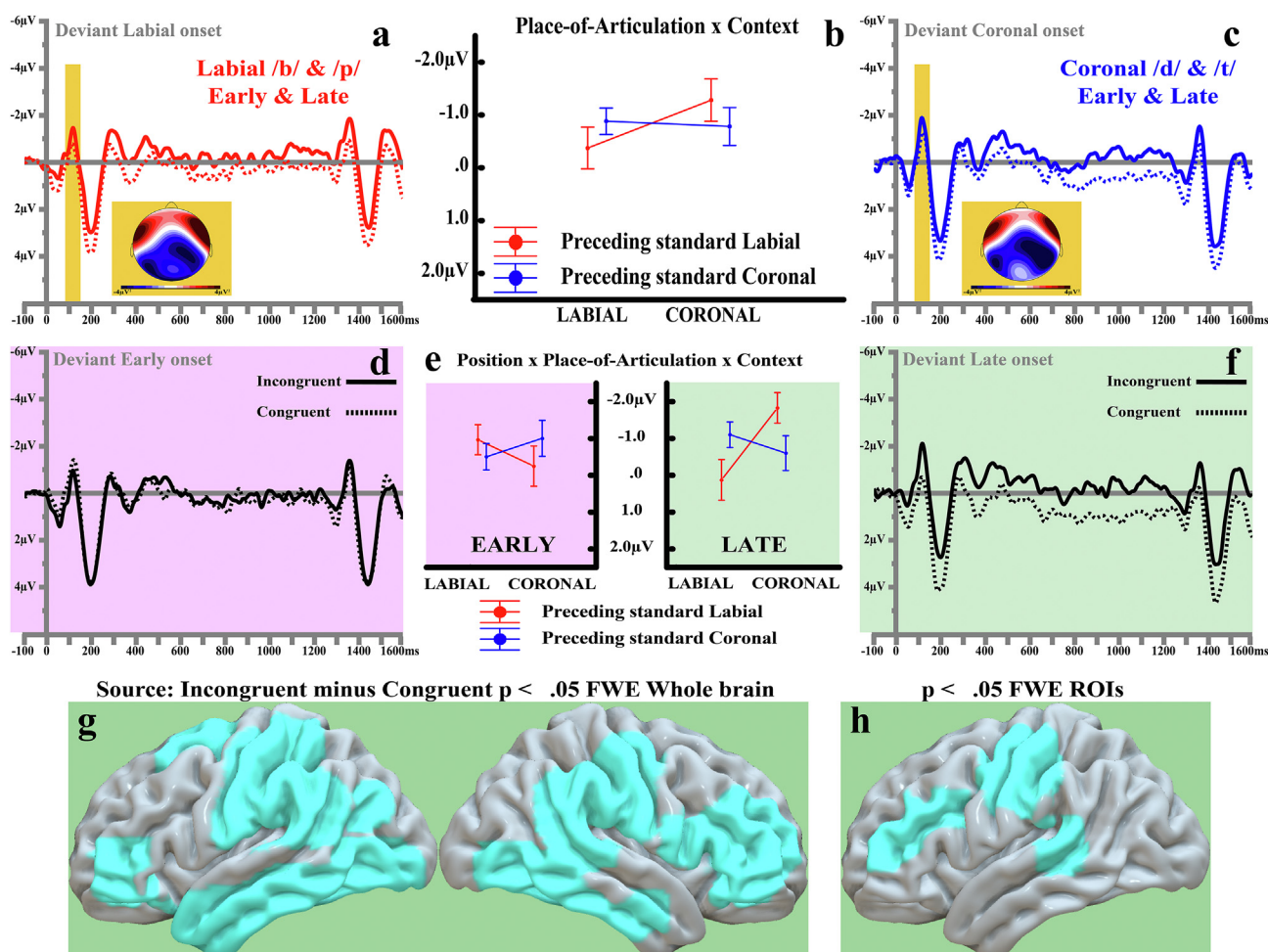
Fig. 3 — MMN-like responses to spoken phonemes. The baseline period for the stimulus-related responses was the last 100 msec before spoken sound onset; the ERPs showed in this picture were extracted from the same fronto-central electrodes (i.e., average of: FC1, FCz, FC2, C1, Cz, C2, CP1, CPz, CP2). The shaded yellow areas highlight the 50 msec time window centered at 120 msec (i.e., the latency of the local maximum of the ERP obtained by collapsing all the deviants); this time window was also used for the statistical evaluations, the voltage maps and to compute source estimations (bottom of the figure). MMN-like responses elicited by labial (a, red) and coronal (c, blue) phonemes in Articulatory-incongruent (thick lines) and congruent (dotted lines) along with the significant interaction (b) from the whole dataset (Early and Late collapsed). MMN-like responses elicited by Articulatory-incongruent (thick lines) and congruent (dotted lines) at Early (magenta) and Late (green) trials pairs are represented in panels (d and f, respectively), whereas panel (e) shows the significant interaction Position x Place-of-Articulation x Context (see Results). Source analysis results calculated on the difference signal (Late Incongruent minus Late Congruent) are shown on panels g (t-test against zero, $p < .05$ FWE, whole brain) and h (t-test against zero, $p < .05$ FWE, on ROIs defined as the significant clusters observed at PP latency for the Late standard sounds, see Fig. 2e).

deviant and the preceding standard, the same interaction of the factors Position, Place-of-Articulation and Context ($F_{1,18} = 16.37$, $p < .001$, $\varepsilon = 1$, $\eta p^2 = .47$) was observed. Furthermore, planned comparisons confirmed the results observed with the unsubtracted data for both the labial (corrected $p = .032$) and coronal (corrected $p = .026$) phonemes (see Fig. 3). Thus, deviant phonemes presented in the context of place-of-articulation-congruent contexts elicited smaller ERPs than the same stimuli in incongruent contexts. However, this priming-like modulation was revealed only after long sequences of standard sounds, that is by those trials for which the PP topography before syllable onset was

likewise modulated by the PoA (see above; hypothesis iii confirmed).

Unsubtracted, MMN-like, responses were further evaluated by examining the larger array of electrodes by performing a 2 × 2 × 2 × 5 × 5 repeated measures ANOVA with the factors Position (Early, Late), Place-of-Articulation (labial, coronal), Context (labial, coronal), Gradient (fronto-parietal, five levels) and Laterality (left-right, five levels). Here, we observed main effects of Gradient ($F_{4,72} = 31.60$, adjusted $p < .001$, $\varepsilon = .33$, $\eta p^2 = .64$) and Laterality ($F_{4,72} = 11.39$, adjusted $p < .001$, $\varepsilon = .54$, $\eta p^2 = .39$), along with the following interactions: Gradient and Laterality ($F_{16,288} = 3.48$, adjusted $p < .01$, $\varepsilon = .33$, $\eta p^2 = .16$),

Place-of-Articulation and Gradient ($F_{4,72} = 6.19$, adjusted $p = .02$, $\varepsilon = .29$, $\eta p^2 = .25$), and, crucially, Position, Place-of-Articulation and Context ($F_{1,18} = 10.99$, $p = .004$, $\varepsilon = 1$, $\eta p^2 = .38$) and Position, Place-of-Articulation, Context and Laterality ($F_{4,72} = 7.32$, adjusted $p = .001$, $\varepsilon = .65$, $\eta p^2 = .29$). The latter two interactions have been confirmed also with the canonical MMN, namely as the difference between the deviant and the preceding standard sound: Position, Place-of-Articulation and Context ($F_{1,18} = 21.73$, $p < .001$, $\varepsilon = 1$, $\eta p^2 = .55$) and Position, Place-of-Articulation, Context and Laterality ($F_{4,72} = 4.14$, adjusted $p = .02$, $\varepsilon = .59$, $\eta p^2 = .19$). To further explore the latter complex 4-way interactions (i.e., Position, Place-of-Articulation, Context and Laterality), we ran two independent repeated measures ANOVAs, one for each deviant position. This further analysis revealed that the 3-way interaction (i.e., Place-of-Articulation, Context and Laterality) was significant for Late (i.e., after long standard trains) but not for Early (i.e., after short standard trains) deviant for both the MMN-like, unsubtracted ($F_{4,72} = 5.17$, adjusted $p = .01$, $\varepsilon = .59$, $\eta p^2 = .22$) and the MMN, subtracted ($F_{4,72} = 3.54$, adjusted $p = .03$, $\varepsilon = .59$, $\eta p^2 = .16$) responses (hypothesis v confirmed; see Table 1 for further results).

4.4. Source analysis: MMN-like

To investigate the cortical loci of origin of the relative enhancement of the MMN-like responses to the place-of-articulation-incongruent conditions following long standard trains (as compared to congruent context), we first subtracted the Late congruent ERPs from the Late incongruent responses; subsequently, the sources of these new ERPs were estimated using the "Greedy Search" algorithm (Friston et al., 2008; Litvak & Friston, 2008). However, since the Late responses included only 48 trials, we collapsed across both phoneme types (i.e., labial and coronal) to obtain a better signal to noise ratio. Whole brain t-tests against zero revealed clusters of activation located in left and right inferior temporal areas, left and right inferior prefrontal areas, superior temporal (auditory) areas and, crucially, in ventral somatosensory and motor areas (see Fig. 3 and Table 2). To ascertain whether predictive and perceptual mechanisms were mediated by the same motor and auditory areas (hypothesis vi), we calculated the activation in specific ROIs (see also PP sources above); this analysis revealed that the areas active in anticipation of Late standard presentation were also active when an unexpected and incongruent phoneme was perceived, thus confirming the relevant role of distributed circuits spread out over superior temporal and prefrontal, motor, areas in both phoneme anticipation and perception (see Fig. 3 and Table 2) (hypothesis vi confirmed).

5.   Discussion

Previous work highlighted the cortical manifestation of semantic predictions. Here, we address for the first time the brain basis of phonological prediction and priming. Our findings demonstrate a brain signature of phonological prediction which was specific to phonological features of the first phoneme of the CV syllable stimuli and preceded the appearance (onset) of CV syllable by ca. 400 msec.

Phonological specificity was also seen for the ERPs elicited by unexpected "deviant" syllables, which differed from the expected speech unit by one phonetic distinctive feature. If expected and unexpected items differed with regard to their Place-of-Articulation (PoA), there was an enhanced brain response after longer trains of repeated standard stimuli, as compared with a situation where expected and unexpected items were congruent with regard to their PoA.

5.1.   Phonological specificity in predictive brain activity

Participants were binaurally presented with repeated syllables in trains of 2−8 identical stimuli. Stimuli differed in their initial phoneme, which was either labial or coronal, so that subsequent trains included initial sounds which were either congruent or incongruent with regard to their Place-of-Articulation. The first presentation of a train was classified as a surprising "deviant" stimulus, whereas its repetitions were "standard stimuli", which either appeared early (position 3) or late (positions 6−8) in the train.

We observed an anticipatory potential shift from circa 400 msec before sound onset on, which was maximally negative at frontolateral recording sites (see Introduction: hypothesis i). Crucially, the cortical topography of the potential differed between labial and coronal stimuli at Late trials thus demonstrating phonological specificity (see Introduction: hypothesis iii). The fact that this effect was observed during the last 50 msec before phoneme onset is in line, on the one hand, with previous observations of the PP in similar MMN paradigms (Grisoni, Mohr, et al., 2019; Pulvermüller & Grisoni, 2020) and, on the other, with what we know about the dynamics of the anticipatory slow-waves, namely that they sharpen towards their ends when the activity flow into primary areas, thus revealing local specificity (Shibasaki & Hallett, 2006). Although attentional mechanisms may play a role in the elicitation of this anticipatory activity, it is worth noting that participants were instructed to focus their attention on a silent movie and to ignore the acoustic stimuli. Furthermore, and importantly, we found phoneme-related topographical specificity for the anticipatory potentials. Therefore, we interpret these potential shifts as a *phonological prediction potential,* including information about the phonological features of the stimuli expected by the subjects. Consistent with this interpretation, the cortical sources of the phonological PP were found in prefrontal and sensorimotor cortex consistent with an activation of articulator representations (lip and tongue; see Introduction: hypothesis ii), although any differences between cortical sources of labial and coronal phoneme expectation conditions were not-significant (Introduction: hypothesis iii and below for further discussion).

5.2.   Phonological specificity in response to unexpected syllables

Phonological specificity was also found in the brain responses to the unexpected deviant stimuli following standard trains. Deviants presented after longer trains of PoA incongruent standards elicited larger ERPs as compared with Late congruent pairs, a pattern consistent with priming mediated by PoA (see Introduction: hypothesis iv, Fig. 3e and f and Table

1). In the deviant ERPs, there was a clear topographical difference between syllables reflecting aspects of their phonological makeup, in particular the place-of-articulation (see Introduction: hypothesis v). Labial phonemes were associated with relatively stronger negative-going centrolateral ERPs, whereas coronal phonemes seemed to enhance the negativity at inferior central sites. This is reminiscent of the different cortical sources previously reported for labial and coronal speech sound perception, which corresponds to the cortical localization of the articulators — lips and tongue — relevant for producing the sounds (Pulvermüller et al., 2006; Schomers & Pulvermüller, 2016; Strijkers et al., 2017). These results seem to align with the idea that partly the same cortical mechanisms are at work in the prediction and anticipation of speech sounds, which likely plays a role in both speech production/articulation and comprehension, and in speech perception of phonemes and syllables. However, we hasten to add that with deviant ERPs, source localization failed to reveal a significant difference between the sources for labial and coronal speech sounds (see Introduction: hypothesis vi).

5.3. Relationship between brain indexes of phonological prediction and priming

A main result of our study was that both the anticipatory, Prediction Potential (PP), occurring before stimulus presentation, and the calculated post-stimulus brain index of phonological priming (i.e., MMN) produced interactions of topographical factors with the Place-of-Articulation factor. Whereas the former was manifest as triple interaction of the topographical factors (Gradient and Laterality) with PoA, the latter was manifest in the form of a triple interaction of topographical factors with Place-of-Articulation and Context. Interestingly, in both cases the mentioned interactions emerged at Late, and not at Early, trials. This pattern is consistent with the proposal that the precision and specificity of the phonologically related brain responses increased with stimulus repetition, possibly due to accumulation of activity in the memory circuits for phonemes or phonemic features.

Regarding the sources, it is apparent that very complex source maps were obtained for the generator constellation indexing phonological priming. These were spread out across a broad range of cortical areas, including anterior, inferior and posterior temporal lobe, part of occipital cortex, parieto-temporal and central sensorimotor cortex, and even anterior dorsolateral and inferior prefrontal areas in both hemispheres. Although these broad activations can only be interpreted with great care, it can be noted that almost all areas found to be active during phonological prediction were also more strongly engaged at post-stimulus latency during unprimed phoneme processing as compared with activation primed by a PoA congruent preceding phoneme. These were inferior prefrontal, central sensorimotor and posterior superior-temporal cortex. These results are consistent with the view that cell assemblies distributed across these latter areas make functional contributions to both phonological prediction, before the stimulus appears, and neurophysiological manifestation of priming, after stimulus presentation, a claim immanent to the integration model summarized in the introduction (Pulvermüller et al., 2006; Pulvermüller &

Fadiga, 2010). In this context, a specific cell assembly (CA-1) shows predictive activation before the phoneme and syllable appearing as a standard (S1); a second cell assembly, CA-2, processing another syllable and critical phoneme, S2, which shares phonological features with S1, is partly active too, as CA-1 and CA-2 share phonological feature neurons. In this context, when a phoneme appears as a novel deviant stimulus, two parallel processes take place: i) the activation level of the CA-1 associated with the previous standard loses activity over time and ii) the activity of the novel standard becomes stronger due to accumulation of activity in the newly activated CA resulting from its repeated stimulations. This double process (i.e., decrease of the previous memory trace together with an increase in the activation of the current stimulation) can be considered as a sharpening of the prediction that could cause not so much a modulation of the PP amplitudes during phoneme repetition, but rather a specific dynamic of the topographical modulation of the anticipatory, PP, activity. Particularly, as a phoneme is repeated over time, the active CA of the previous memory trace, ideally, disappears, whereas the CA associated with the current stimulation increases so that any phoneme-specificity would emerge for Late as compared with Early trials. In this respect, the PP topographical modulations observed at Late trials would support this hypothesis. (See also introduction, hypothesis iv).

However, it is also important to note that this model can also explain the priming effect observed after Late-deviant syllable onset. Indeed, as reported above, CAs may share phonological feature neurons, a fact which begs consideration of phonological priming as a difference in baseline before critical stimulus appears (Pulvermüller, 2018). Indeed, primed and unprimed conditions also differ because the network representing a given phoneme may (congruent PoA) or may not (incongruent PoA) overlap with the (predictively) preactivated network, thus causing a difference in the observed post-stimulus brain responses (i.e., MMN). Therefore, also the post-stimulus response (MMN) would, in this context, be more affected by the PoA incongruency at Late rather than at Early trials, meaning in those trials in which deviant (congruent or incongruent) sounds were presented in the context of sharper predictions. Indeed, at Late trials a sound which is incongruent with the predictions about the PoA would add, relatively, more activation than the congruent pairs, as in the former case there would be the ignition of two more distinct circuits within the production system, whereas in the latter case there would be more overlap between the predictively and perceptually activated representation. In other words, Early congruent deviants would add relatively less activity than Late incongruent ones, because at Early trials there might still be residual activity from previous standard traces which may then reduce the amount of "novel" activity brought by the deviant sound. In this perspective the data here presented may shed new light on the interaction between prediction and resolution.

5.4. Limitations, outlook

Although topographical modulations of the PP depending on the PoA of the anticipated phoneme were observed, these interactions just reached the statistical significance threshold (e.g., $p = .048$) with moderate effect sizes (e.g., $\eta p^2 = .11$). This

contrasts with previous observations on semantic predictions in which more robust PP's topographical modulations to different semantic categories (e.g., animal *vs* tool nouns) were reported (Grisoni et al., 2020). A possible explanation for this apparent discrepancy is to be found in the spatial proximity of the underlying representations. Whereas words belonging to different semantic categories show preactivations falling into distant modality preferential brain areas (e.g., parieto-temporo-occipital and prefrontal areas for animal and tool nouns, respectively), the PoA feature is codified in specific articulatory motor and auditory temporal sub-areas which are much closer to each other. The EEG spatial resolution is not optimal to fully detect such modulations which requires methods with a much higher spatial resolution. Similarly, the lack of significant source differences for phonemes with different place-of-articulation contrasts with a range of reports documenting such differences (D'Ausilio et al., 2009; Pulvermüller et al., 2006; Schomers et al., 2015; Skipper, van Wassenhove, Nusbaum, & Small, 2007). However, these earlier reports had mostly used functional MRI or even intracranial recordings − (Schomers & Pulvermüller, 2016; Skipper et al., 2017) − that is, methods with much better spatial resolutions than the multi-channel EEG recordings from the top of the head, which we used in the present study. One previous study found differential activation of tongue and lip representations using 306 channel MEG recordings, but the authors discuss the limitations of their source estimation results even with such high spatial resolution (Strijkers et al., 2017). It is very well known that source estimation with EEG or MEG is not only subject to the Helmholtz inverse problem (Helmholtz, 1853); furthermore, the mathematically sophisticated estimation techniques of distributed cortical sources have their well-known limits if it comes to localizing the contributions of multiple close-by cortical sources (Hämäläinen & Ilmoniemi, 1994; Hauk, Stenroos, & Treder, 2019; Hauk, Wakeman, & Henson, 2011). Note that the spatial difference between the cortical lip and tongue tip representations are only 1−2 cm apart, so that it is very likely that a distributed source estimation algorithm may exceed its limits unless the sources are clearly distinct and very strong. Furthermore, we cannot assume that tongue motor activity is entirely at zero when a/pƏ/ is being produced nor can be state that lips entirely rest when pronouncing a/tƏ/and likewise the respective articulator activations may not be entirely selective in speech perception as well (see above). What our source activation data clearly show, however, is that during the expectation of specific speech units and also when encountering unexpected syllables, there is clear activation of inferior and lateral motor areas, which is also seen during articulation. This does not finally prove that the respective motor circuits are involved in both speech production and perception, but the data fit the related proposal of models claiming the involvement of mechanisms that link together action and perception mechanisms in both speech prediction and perception. Intriguingly, even the observed phonological priming effects came with an activation of central articulatory motor areas, along with activity in the superior temporal auditory system.

To further follow up on the questions addressed by this work, it will be necessary to move more closely to the brain generators underlying the processing of phonological information. One way is offered by recordings from the exposed cortical surface in neurosurgery patients. While this strategy comes with the caveat of working with tissue affected by disease of the brain (e.g., epilepsy or tumors), it offers the chance to find more fine-grained local differences in generator constellation. Subdural and intracortical recordings could help to delineate the differences in local brain mechanisms active in the processing of specific phonemes and phonological features in speech perception and production (Bouchard, Mesgarani, Johnson, & Chang, 2013; Khoshkhoo et al., 2018; Mesgarani et al., 2014), and likewise that of priming and prediction.

Open practices

We report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, whether inclusion/ exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study. No part of the study procedures and analyses was pre-registered prior to the research being conducted.

The study in this article earned an Open Materials badge for transparent practices. Materials and data for the study are available at https://osf.io/ej3wx/?view_only=e2697a99dd5d4e a786911b05a27fa422.

Raw data could not be made public as this is not permitted by the conditions of our ethics approval, which prevent sharing of individual subject data under any circumstances with anyone outside the author team.

Credit authors statement

Luigi Grisoni: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing-Original Draft, Writing-Review & Editing, Visualization, Project administration.

Friedemann Pulvermüller: Conceptualization, Methodology, Validation, Resources, Data Curation, Writing − Original Draft, Writing-Review & Editing, Supervision, Project administration, Funding acquisition.

Acknowledgements

REFERENCES

Arsenault, J. S., & Buchsbaum, B. R. (2015). Distributed neural representations of phonological features during speech perception. *The Journal of Neuroscience, 35*(2), 634−642.

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation, 7*(6), 1129—1159. https://doi.org/10.1162/neco.1995.7.6.1129

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology, 60*(4), 343—355. https://doi.org/10.1016/0013-4694(85)90008-2

Bouchard, K. E., Mesgarani, N., Johnson, K., & Chang, E. F. (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature, 495*(7441), 327—332.

Chartier, J., Anumanchipalli, G. K., Johnson, K., & Chang, E. F. (2018). Encoding of articulatory kinematic trajectories in human speech sensorimotor cortex. *Neuron, 98*(5), 1042—1054. https://doi.org/10.1016/j.neuron.2018.04.031. e1044.

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor Anova designs. *Educational and Psychological Measurement, 33*(1), 107—112. https://doi.org/10.1177/001316447303300111

Cowan, N., Winkler, I., Teder, W., & Näätänen, R. (1993). Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP). *Journal of Experimental Psychology. Learning, Memory, and Cognition, 19*(4), 909—921. https://doi.org/10.1037/0278-7393.19.4.909

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology: CB, 19*(5), 381—385. https://doi.org/10.1016/j.cub.2009.01.017

Deecke, L., Scheid, P., & Kornhuber, H. H. (1969). Distribution of readiness potential, pre-motion positivity, and motor potential of the human cerebral cortex preceding voluntary finger movements. *Experimental Brain Research, 7*(2), 158—168.

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

Evans, S., & Davis, M. H. (2015). Hierarchical organization of auditory and motor representations in speech perception: Evidence from searchlight similarity analysis. *Cerebral Cortex, 25*(12), 4772—4788. https://doi.org/10.1093/cercor/bhv136

Friston, K., Harrison, L., Daunizeau, J., Kiebel, S., Phillips, C., Trujillo-Barreto, N., … Mattout, J. (2008). Multiple sparse priors for the M/EEG inverse problem. *Neuroimage, 39*(3), 1104—1120. https://doi.org/10.1016/j.neuroimage.2007.09.048

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. [journal article]. *Psychometrika, 24*(2), 95—112. https://doi.org/10.1007/bf02289823

Grisoni, L., Dreyer, F. R., & Pulvermüller, F. (2016). Somatotopic semantic priming and prediction in the motor system. *Cerebral Cortex, 26*(5), 2353—2366. https://doi.org/10.1093/cercor/bhw026

Grisoni, L., Miller, T. M., & Pulvermüller, F. (2017). Neural correlates of semantic prediction and resolution in sentence processing. *The Journal of Neuroscience: the Official Journal of the Society for Neuroscience, 37*(18), 4848—4858. https://doi.org/10.1523/JNEUROSCI.2800-16.2017

Grisoni, L., Mohr, B., & Pulvermüller, F. (2019a). Prediction mechanisms in motor and auditory areas and their role in sound perception and language understanding. *Neuroimage, 199*, 206—216. https://doi.org/10.1016/j.neuroimage.2019.05.071

Grisoni, L., Moseley, R. L., Motlagh, S., Kandia, D., Sener, N., Pulvermüller, F., … Mohr, B. (2019b). Prediction and mismatch negativity responses reflect impairments in action semantic processing in adults with autism spectrum disorders. [Original research]. *Frontiers in Human Neuroscience, 13*(395). https://doi.org/10.3389/fnhum.2019.00395

Grisoni, L., Tomasello, R., & Pulvermüller, F. (2020). Correlated brain indexes of semantic prediction and prediction error: Brain localization and category specificity. *Cerebral Cortex, 31*(3), 1553—1568. https://doi.org/10.1093/cercor/bhaa308

Hämäläinen, M. S., & Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: Minimum norm estimates. [journal article]. *Medical & Biological Engineering & Computing, 32*(1), 35—42. https://doi.org/10.1007/bf02512476

Hauk, O., Stenroos, M., & Treder, M. (2019). EEG/MEG source estimation and spatial filtering: The linear toolkit. In S. Supek, & C. J. Aine (Eds.), *Magnetoencephalography: From signals to dynamic cortical networks* (pp. 1—37). Cham: Springer International Publishing.

Hauk, O., Wakeman, D. G., & Henson, R. (2011). Comparison of noise-normalized minimum norm estimates for MEG analysis using multiple resolution metrics. *Neuroimage, 54*(3), 1966—1974. https://doi.org/10.1016/j.neuroimage.2010.09.053

Helmholtz, H. (1853). Über einige Gesetze der Vertheilung elektrischer Ströme in körperlichen Leitern mit Anwendung auf die thierisch-elektrischen Versuche. *Annalen der Physik, 165*(6), 211—233. https://doi.org/10.1002/andp.18531650603

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews. Neuroscience, 8*(5), 393—402. https://doi.org/10.1038/nrn2113

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and cognitive processes, 5*(4), 281—312.

Kappenman, E. S., & Luck, S. J. (2012). *The oxford handbook of event-related potential components.* Oxford University Press.

Khoshkhoo, S., Leonard, M. K., Mesgarani, N., & Chang, E. F. (2018). Neural correlates of sine-wave speech intelligibility in human frontal and temporal cortex. *Brain and language, 187*, 83—91.

Kilner, J. M., Vargas, C., Duval, S., Blakemore, S. J., & Sirigu, A. (2004). Motor activation prior to observation of a predicted movement. *Nature Neuroscience, 7*(12), 1299—1301. https://doi.org/10.1038/nn1355

Leon-Cabrera, P., Flores, A., Rodriguez-Fornells, A., & Moris, J. (2019). Ahead of time: Early sentence slow cortical modulations associated to semantic prediction. *Neuroimage, 189*, 192—201. https://doi.org/10.1016/j.neuroimage.2019.01.005

León-Cabrera, P., Rodríguez-Fornells, A., & Morís, J. (2017). Electrophysiological correlates of semantic anticipation during speech comprehension. *Neuropsychologia, 99*, 326—334. https://doi.org/10.1016/j.neuropsychologia.2017.02.026

Litvak, V., & Friston, K. (2008). Electromagnetic source reconstruction for group studies. *Neuroimage, 42*(4), 1490—1498. https://doi.org/10.1016/j.neuroimage.2008.06.022

Litvak, V., Mattout, J., Kiebel, S., Phillips, C., Henson, R., Kilner, J., … Friston, K. (2011). EEG and MEG data analysis in SPM8. *Computational Intelligence and Neuroscience, 2011*, 852961. https://doi.org/10.1155/2011/852961

Luck, S. J. (2014). *An introduction to the event-related potential technique (vol. Second edition).* Cambridge, MA: MIT Press.

Maldjian, J. A., Laurienti, P. J., Kraft, R. A., & Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage, 19*(3), 1233—1239.

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science, 343*(6174), 1006—1010. https://doi.org/10.1126/science.1245994

Obleser, J., & Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory cortex. [Research Support, Non-U.S. Gov't]. *Trends Cogn Sci, 13*(1), 14—19. https://doi.org/10.1016/j.tics.2008.09.005

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

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin, 144*(10), 1002—1044. https://doi.org/10.1037/bul0000158

Pulvermüller, F. (2018). Neurobiological mechanisms for semantic feature extraction and conceptual flexibility. *Topics in Cognitive Science, 10*(3), 590—620. https://doi.org/10.1111/tops.12367

Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. [10.1038/nrn2811]. *Nature Reviews. Neuroscience, 11*(5), 351—360. http://www.nature.com/nrn/journal/v11/n5/suppinfo/nrn2811_S1.html.

Pulvermüller, F., & Grisoni, L. (2020). Semantic prediction in brain and mind. *Trends Cogn Sci.* https://doi.org/10.1016/j.tics.2020.07.002

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proc Natl Acad Sci U S A, 103*(20), 7865—7870. https://doi.org/10.1073/pnas.0509989103

Pulvermüller, F., Tomasello, R., Henningsen-Schomers, M. R., & Wennekers, T. (2021). Biological constraints on neural network models of cognitive function. *Nature Reviews Neuroscience.* https://doi.org/10.1038/s41583-021-00473-5

Schomers, M. R., Kirilina, E., Weigand, A., Bajbouj, M., & Pulvermüller, F. (2015). Causal influence of articulatory motor cortex on comprehending single spoken words: TMS evidence. *Cerebral Cortex, 25*(10), 3894—3902. https://doi.org/10.1093/cercor/bhu274

Schomers, M. R., & Pulvermüller, F. (2016). Is the sensorimotor cortex relevant for speech perception and understanding? An integrative review. *Front Hum Neurosci, 10*, 435. https://doi.org/10.3389/fnhum.2016.00435

Shibasaki, H., & Hallett, M. (2006). What is the Bereitschaftspotential? *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology, 117*(11), 2341—2356. https://doi.org/10.1016/j.clinph.2006.04.025

Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain and Language, 164*, 77—105. https://doi.org/10.1016/j.bandl.2016.10.004

Skipper, J. I., van Wassenhove, V., Nusbaum, H. C., & Small, S. L. (2007). Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cerebral Cortex, 17*(10), 2387—2399. https://doi.org/10.1093/cercor/bhl147

Stokes, R. C., Venezia, J. H., & Hickok, G. (2019). The motor system's [modest] contribution to speech perception. *Psychonomic Bulletin & Review, 26*(4), 1354—1366. https://doi.org/10.3758/s13423-019-01580-2

Strijkers, K., Costa, A., & Pulvermüller, F. (2017). The cortical dynamics of speaking: Lexical and phonological knowledge simultaneously recruit the frontal and temporal cortex within 200 msec. *Neuroimage, 163*, 206—219. https://doi.org/10.1016/j.neuroimage.2017.09.041