Speech Prosody Serves Temporal Prediction of Language via Contextual Entrainment

---

Speech Prosody Serves Temporal Prediction of

Language via Contextual Entrainment

(Prosodic entrainment affects language comprehension)

Yulia Lamekina[a]*, Lorenzo Titone[a], Burkhard Maess[b], and Lars Meyer[a,c]

[a]Research Group Language Cycles, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1a, 04103 Leipzig, DE;

[b]Methods and Development Group Brain Networks, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1a, 04103 Leipzig, DE;

[c]University Clinic Münster, Albert-Schweitzer-Campus 1, 48149 Münster, DE

Email: lamekina@cbs.mpg.de

Number of pages: 63

Number of figures: 8

Number of tables: 5

Number of extended data figures: 1

Number of extended data tables: 1

Number of words

1.   abstract: 214

2.   introduction: 794

3.   discussion: 1510

Abstract

Temporal prediction assists language comprehension. In a series of recent behavioral studies, we have shown that listeners specifically employ rhythmic modulations of prosody to estimate the duration of upcoming sentences, thereby speeding up comprehension. In the current human magnetoencephalography (MEG) study on participants of either sex, we show that the human brain achieves this function through a mechanism termed entrainment. Through entrainment, electrophysiological brain activity maintains and continues contextual rhythms beyond their offset. Our experiment combined exposure to repetitive prosodic contours with the subsequent presentation of visual sentences that either matched or mismatched the duration of the preceding contour. During exposure to prosodic contours, we observed MEG coherence with the contours, which was source-localized to right-hemispheric auditory areas. During the processing of the visual targets, activity at the frequency of the preceding contour was still detectable in the MEG; yet sources shifted to the (left) frontal cortex, in line with a functional inheritance of the rhythmic acoustic context for prediction. Strikingly, when the target sentence was shorter than expected from the preceding contour, an omission response appeared in the evoked potential record. We conclude that prosodic entrainment is a functional mechanism of temporal prediction in language comprehension. In general, acoustic rhythms appear to

endow language for employing the brain's electrophysiological mechanisms of temporal prediction.

Significance statement

Language comprehension benefits from our ability to predict upcoming stimuli. Here, we report on a key neural substrate. We show that electrophysiological brain activity inherits prosodic modulations—the melody of speech—from prior context, allowing listeners to estimate the duration of upcoming language stimuli. By using magnetoencephalography, we find that the brain not only responds to prosody when speech is present, but its activity continues at the prosodic frequency seconds into the future, benefiting behavioral responses. During continuation, activity shifts from auditory to frontal cortex, the epicenter of the brain's predictive abilities. The human brain seems to initiate the top-down prediction of language stimuli by copying sensory rhythms and projecting them into the future.

Introduction

Predicting upcoming linguistic information allows us to process language fast and efficiently (Li & Zhang, 2023; Sohoglu et al., 2012). Predictions are generated across different levels of linguistic content, including phonology (DeLong et al., 2005), syntax (Lau et al., 2016; Van Berkum et al., 2005), and (lexical) semantics (Boston et al., 2011; Demberg & Keller, 2008; Smith & Levy, 2013).

In addition to content, listeners predict the timing of upcoming linguistic units. On slow time scales, listeners employ speech prosody to infer the duration of upcoming sentence segments (Grosjean, 1983). Cues such as pauses, pitch modulations, and duration changes trigger the bottom-up segmentation of sentences into syntactic phrases (Frazier et al., 2006). Crucially, interpretation of prosodic cues depends on a wider prosodic context. For instance, intonational phrase boundaries (IPBs) are not processed in terms of their absolute acoustic magnitude, but relative to the magnitude of preceding IPBs (Clifton et al., 2002; Snedeker & Casserly, 2010). Moreover, even distant prosodic cues can influence the prediction of upcoming material, such that prosody at sentence onset affects subsequent segmentation and word recognition (Brown et al., 2011; Dilley & McAuley, 2008). Furthermore, during reading, where acoustic cues are unavailable, listeners actively construct implicit prosody that facilitates speech processing (Breen, 2014; Breen, Kaswer, et al., 2016; Fodor, 2002). Consequently, context effects have also been reported across perceptual modalities. In previous studies, it was observed that prosodic contours can trigger downstream effects that persist beyond stimulation, affecting the comprehension of upcoming visual sentences devoid of prosody (e.g., Steinhauer & Friederici, 2001).

Recent behavioral findings suggest that temporal prediction, including the effects of prosody, may be triggered by rhythmic or repetitive contexts. For instance, rhythmic amplitude-modulated sounds at a given frequency induce perceptual detectability of subsequent targets that arrive at the same frequency (Hickok, Farahbod, & Saberi 2015). Furthermore, the prosodic syllable rate of a lead-in sentence can affect detection of subsequent target syllables: After a fast-rate sentence, subjects overhear short target syllables that they do perceive when the lead-in sentence is presented at a slow rate (Bosker, 2017; Dilley & Pitt, 2010). Corpus evidence further suggests that language

prosody exhibits sufficient rhythmicity for temporal predictions to emerge (Inbar et al., 2020; Stehwien & Meyer, 2021).

In a recent set of behavioral studies, we found supporting evidence that the facilitatory effect of prosodic context on temporal prediction hinges specifically on the rhythm of prosody (Lamekina & Meyer, 2022). We had participants listen to repetitive prosodic contours, followed by visual target sentences that either matched or not in duration. We showed that a duration match accelerated the processing of the target sentence, indicating that listeners predict sentence duration.

In the current study, utilizing the same paradigm in a magnetoencephalography (MEG) experiment, we report that the behavioral prediction benefit of rhythmic prosodic contexts is driven by brain activity that resonates with such rhythm. Prior work has reported narrow-band electrophysiological activity to tune to rhythmic stimuli and thereby optimize the processing of upcoming input (Henry & Obleser, 2012; Lakatos et al., 2008; Schroeder & Lakatos, 2009). Specifically, activity in the delta range (< 4 Hz) triggers auditory temporal duration predictions through an electrophysiological mechanism termed entrainment, by which the brain inherits a stimulation frequency to persist after stimulus offset (Breska & Deouell, 2017; Herbst et al., 2022; Stefanics et al., 2010). This mechanism could also potentially affect the processing of upcoming language stimuli. In the domain of speech and language, delta-band activity is known to synchronize with prosody (Bourguignon et al., 2013; Luo & Poeppel, 2007), but predictive functions have not been reported.

To investigate whether delta-band prosodic entrainment subserves sustained temporal predictions during subsequent sentence processing, we designed an MEG study that combined an initial repetitive prosodic rhythm (*entrainment phase*) with a subsequent visual sentence presentation (*target phase*; Figure 1A). The visual modality for the target

phase was chosen in order to avoid confounding prosodic characteristics in auditory target stimuli. Similar cross-modal paradigms have been successfully implemented in previous studies (e.g., Steinhauer & Friederici, 2001), which showed that a stable effect is also obtained across modalities. Our experiment employed prosodic contours that either matched or mismatched the duration of a subsequent visual target sentence. Our behavioral results (Lamekina & Meyer, 2022) showed that a duration match indeed associates with a behavioral benefit. On the neural level, we expected this effect to emerge due to rhythmic prosodic entrainment. That is, we first hypothesized that delta-band activity would synchronize with the modulation frequency of the contours. Second, we expected that brain activity at the modulation frequency would still be detectable during the processing of the visual target sentence. Third, we expected an omission ERF response when the duration of the target sentence mismatched the duration of the entraining contours (Ragazzoni et al., 2019; Scharinger et al., 2017).

Methods

Participants

We conducted the experiment on 40 participants (German native speakers, right-handed, 19 females, age range = 18-35 years, mean age = 28 years, SD = 4 years). The MEG data was recorded with different participants and in a separate session from the preceding behavioral study with a similar paradigm (Lamekina & Meyer, 2022). Five participants were further excluded from the analysis due to noise in the MEG; the resulting sample was 35 subjects. Participants had normal or corrected-to-normal vision and no reported history of neurological or hearing disorders. Participation was reimbursed with € 12 per hour. All subjects were naïve to the purpose of the study. Written informed consent was

obtained prior to the experiment. The study conformed to the guidelines of the Declaration of Helsinki and was approved by the local ethics committee of the University of Leipzig, DE (approval N 060/17-ek).

Experimental design and statistical analysis

<u>Stimuli and Paradigm</u>

In order to investigate the potential influence of prosodic entrainment on subsequent sentence comprehension, the trials of our experiment combined an initial prosodic rhythm with a subsequent visual target sentence. This prosodic contour always belonged to one of the conditions—SLOW or FAST—characterised by different length. Importantly, although the main manipulation for the contours was the duration change, the contours were repeated three times, which resulted in three cycles with a period different for every condition; this additional manipulation was introduced in order to induce stimuli rhythmicity and an electrophysiological brain response at the same frequency. Contour repetitions had pauses of 160 ms in between. This duration of the pause has previously been shown to be sufficient in eliciting entrainment to prosodic stimuli (Ghitza, 2017). Together with the pause, one cycle of SLOW contour lasted for 1.73 s, corresponding to a frequency of approximately 0.6 Hz. One cycle of the FAST contour together with the pause lasted for 1.1 s, corresponding to 0.9 Hz. An additional pause matching the difference in duration between the FAST and SLOW contours was added before the FAST contour to equalize duration across conditions.

Prosodic contour exposure was followed by visual word-by-word presentation of a target sentence. Critically, the presentation of the visual words was adjusted to contour duration, whereby word presentation rate was calculated from the same auditory

sentences from which the contour was extracted. Target visual sentences were either LONG or SHORT. The LONG sentences had an exact duration of 1.884 s (314 ms x 6 words) and the SHORT sentences had an exact length of 1.57 s (314 ms x 5 words). The presentation of sentences, therefore, corresponded in timing either to the SLOW contour without pause (5 words) or to the FAST contour without pause (6 = 3+3 words). (cf. Figure 1).

Figure 1 here

After sentence presentation and a delay, some trials were followed by comprehension questions (see Procedure). Stimuli and procedure were adapted from Experiment 1 in our previous behavioral study on prosodic entrainment (Lamekina & Meyer, 2022). Therefore, in the following paragraphs we reiterate the description of the stimulus material used in the previous study.

For examples of target sentences, see below:

(1)   LONG: Max sieht Tom und Karl lacht.

                    "*Max sees Tom and Karl laughs.*"

(2)   SHORT: Max sieht Tom und Karl.

                    "*Max sees Tom and Karl.*"

For sentence construction, we used 32 monosyllabic first names of 3–6 characters to balance word-by-word presentation (New et al., 2006). Noun frequencies were normally distributed (Heister et al., 2011). Since male and female first names differ in length, we used male first names only. We also selected 75 transitive and 75 intransitive German verbs in 3rd person singular present tense. Length was matched (1–2 syllables, 5–8 characters). Verb frequencies were also normally distributed. Pairs of transitive and intransitive verbs were made based on semantic fit (e.g., *expect–come, wake up–sleep*). Combination of verb pairs and names yielded 6,000 sentences. A different name triplet was used for each of these. Name triplets were selected to not contain similar-sounding names (e.g., *Frank* and *Franz*).

Prosodic contours of two different rates were presented before the target sentences. Contours were made by averaging the pitch tracks of the visual sentences, which were stripped off synthetic recordings (Oord et al., 2016) in Praat (Boersma & van Heuven, 2001). We used a female voice (minimum pitch = 116 Hz, maximum pitch = 267 Hz, average pitch = 191.5 Hz) for its broad pitch range and high variability. The two entrainment rates were SLOW (based on the 3,000 5-word SHORT sentences; e.g., *Max sieht Tom und Karl*) and FAST, based on 3,000 additional partial 3-word sentences (e.g., *Max sieht Tom*).

For averaging, contour durations were adjusted to the average duration of the respective sentence recordings (SLOW: 1,570 ms; FAST: 942 ms). The contours within each condition had an identical duration in order to increase signal-to-noise ratio. For the purpose of our experiment, the contours needed to be *delexicalized*, i.e. stripped of any lexical content/acoustic fine structure, so that only prosodic characteristics remain. To achieve this aim, PURR (Prosody Unveiling through Restricted Representation) method was used (Sonntag & Portele, 1998). This pipeline was recommended in previous studies

for constructing a delexicalised prosodic contour (Steinhauer & Friederici, 2001). The method involves extracting the pitch values from the original contours and constructing a sound by adding a sine wave at pitch, its second harmonic of ¼ of the amplitude, and its third harmonic of 1/16 of the amplitude (suggested by Klasmeyer, 1997). Therefore, out of the original spectral characteristics of the speech signal, only the pitch modulations are retained, which permits to disentangle prosodic modulations from other speech components. Prosodic pitch has proved to provide a substantial contribution to entrainment, separately from other acoustic and phonetic features (Teoh et al., 2019). PURR has been tested extensively and compared to other methods, proving over a variety of experiments to have the best functionality and acceptability (listeners recognizing the signal as coming from natural human speech) for speech delexicalisation (Kotz et al., 2003; M. Meyer et al., 2002; Pannekamp et al., 2005). Contours were further normalised to 65 dB and lowered in pitch by 55 Hz to ensure a comfortable hearing level. Average word duration for timed visual presentation calculated from the synthesised contours length was 314 ms.

The 6,000 sentences and contours were combined into 20 experimental lists of 300 trials each. Within list, every verb pair was used 4 times, once within each condition (i.e., SHORT–FAST, SHORT–SLOW, LONG–FAST, and LONG–SLOW). Pairs and conditions did not repeat across subsequent trials. We disallowed adjacent name triplets with identical or similar names. Identical triplets did not repeat within list.

Procedure

Each trial started with a visual fixation cross and auditory presentation of one of the two prosodic contours, repeated three times (Figure 1B). After a contour, a target sentence

(either SHORT or LONG) was presented word by word. The words were shown in a rapid serial visual presentation (RSVP; Young, 1984). Word presentation rate was identical across conditions (314 ms/word); critically, it was adjusted to contour duration to allow for assessing prosodic entrainment. In contrast to our behavioral study (Lamekina & Meyer, 2022), we adapted the procedure to better suit the purposes of the current experiment: finals words of the sentence were no longer self-paced, that is, RSVP was used for all words of the sentence. This was done to avoid muscle artifacts associated with the button presses. It would also be worthwhile to note that single words could induce their own neural responses in the target phase at a frequency of around 3 Hz; however, this doesn't match our frequencies of interest (0.6 and 0.9 Hz) and, in our view, presents no obstacles or interest for further analysis.

After visual sentence presentation and a jittered delay of 500 + 0-250 ms, comprehension questions were presented in 75% of the trials. To avoid strategy build-up, questions requiring a "yes" and a "no" answer were both included for each condition (see (Lamekina & Meyer, 2022) for details). Both entrainment conditions were matched on the types of questions and the amount of correct "yes" and "no" answers. The "yes/no" button assignment was matched across the participants. There was a response timeout of 2,000 ms. In case of timeout, a screen stating "Please answer faster" appeared, and the experiment advanced to the next trial. Participants were instructed to listen to the audio contour, read the sentences presented word-by-word and then answer a comprehension question after some trials. The 300 trials of the experiment ran in five blocks, with self-regulated pauses in between (every block took approximately 10 minutes). The whole experiment lasted approximately 60-70 minutes, depending on the duration of the pauses.

Data recording and preprocessing

MEG recording

The experiment was conducted in an electromagnetically shielded room in a single session. Stimuli were presented using the Presentation software (Neurobehavioral Systems, Inc.). Participants heard the prosodic contours through air-conduction earplugs (ER3-14A/B, Etymotic Research Inc., Elk Grove Village IL, U.S.A.) connected via a 50-cm plastic tube to piezo-phones (TIP-300, Nicolet, Biomedical Madison, WI, U.S.A.). Visual stimuli were back-projected on a semi-transparent screen from a projector located outside of the magnetic shielding room (Panasonic PT-D7700E, Matsushita Electric Industrial Co., Japan). The screen was located ~90 cm from each participant. MEG signals were measured at a sampling rate of 1,000 Hz within a passband of 0-330 Hz from 306 sensors, including 102 magnetometers and 204 gradiometers (Vectorview, Elekta-Neuromag, Helsinki, Finland). Head position inside the helmet was continuously monitored using five head-tracking coils.

MEG preprocessing

Each block of the data was separately corrected for head movements and external noise with the Signal Space Separation (SSS) method (Taulu et al., 2005) via the MaxFilter™ software (MaxfilterVersion 2.2.15, Elekta Oy, FI) utilizing spherical functions up to eleventh order for the head field model and up to the second order for the environmental field model. Movements were compensated using a 500 ms window while forwarding with steps of 250 ms. The temporal correction was computed using a 10 s time window; correlations higher than 0.98 between inside and outside field components were projected out. Line frequency was set to 50 Hz. Further analyses were conducted in MATLAB (The MathWorks, Inc.) via

the FieldTrip toolbox (Oostenveld et al., 2011). Since previous research has shown that results obtained from different types of MEG sensors after SSS are highly correlated (Garcés et al., 2017) we restricted analysis to magnetometers only. The data were first low-pass filtered at 90 Hz (1-pass, zero-phase-shift finite impulse response (FIR) filter with windowed sinc, Kaiser window), detrended and demeaned. Line noise (50 Hz) and its harmonics were removed from the signal using ZapLine (de Cheveigné, 2020). Noisy channels were identified semi-automatically: Candidates in every block were chosen when the maximum absolute value in the channel exceeded a threshold of 6 pT; those candidates were further confirmed by visual inspection. We excluded 5 participants who had more than 10% of noisy channels; all further analyses were conducted on 35 subjects. In this final sample, mean number of removed channels per subject was 2 out of 102 channels (SD = 2). After this procedure, the data were down-sampled to 250 Hz and segmented into epochs, each comprising one contour and the following visual sentence (ranging from 6,480 to 8,530 ms depending on the 4 combinations of prosodic and sentence conditions). For coherence, power, and ERF analyses, smaller segments were cut out separately at later analyses stages (see below). Epoched data were further subjected to artifact rejection. First, trials containing SQUID jump artifacts were detected using z-transformation of the data at each time point. These z-values were calculated from the 9th-order median-filtered data. The z-value threshold of the jump detection was set to 75; trials with z-values exceeding this threshold were excluded. A similar method was used for automatic detection of trials with muscle artifacts (fronto-temporal channels, prior band-pass filtering from 110 to 124 Hz, z-value threshold = 20). For this step, trials were zero-padded prior to filtering to avoid edge artifacts. Finally, trials that contained time points with values exceeding ± 6 pT were excluded. The whole artifact detection procedure led to removal of a mean of 4 trials per block (SD = 4). The preprocessed data were further

subjected to independent component analysis (ICA) to remove ocular and cardiac artifacts. Artifact components were identified based on visual inspection of component waveform and topography. The MEG data were then reconstructed in the original sensor space, excluding the artifactual components. Excluded channels were then interpolated within block using spline interpolation and a template neighborhood structure for Elekta-Neuromag magnetometers. Finally, the preprocessed data from the five blocks per subject were concatenated. Further analyses in sensor space were conducted on the subject level. In the source space, we ran the analysis on the block level due to differences in MEG sensor positions and therefore, different lead fields for every recording block.

MRI preprocessing

In order to identify neural sources underlying the prosodic entrainment mechanism, anatomically constrained source localization was used. First, individual T1-weighted MRI images obtained with a 3T MRI scanner (Magnetom Trio, Siemens AG, Germany) were segmented using the Freesurfer software (http://surfer.nmr.mgh.harvard.edu/). For every participant, MEG data were co-registered with the individual MRI scan via realignment of the fiducial (nasion, left and right pre-auricular) and digitized head surface points (acquired with Polhemus FASTRAK 3D digitizer). To this end, we used a semi-automated iterative procedure (Besl & McKay, 1992) implemented in the MNE software (MGH, Boston, US, http://www.nmr.mgh.harvard.edu/martinos/userInfo/data/). Next, using MNE, volume conductors were constructed as boundary element models (BEMs), resulting in individual inner-skull surfaces comprising 2,562 vertices each. Based on the volume conductors, individual head models were created in Fieldtrip using the single-shell model (Nolte, 2003). In parallel, source spaces consisting of 10,242 vertices per hemisphere were constructed

in MNE. To arrive at results that are comparable between subjects in terms of neuroanatomical function and structure, every subject's source space was parcellated into the regions of the HCPMMP1 atlas (Glasser et al., 2016); all source-space analyses were carried out on atlas-defined regions (Fischl et al., 1999). Based on the individual head models, source spaces, and MEG sensor positions in the current recording block, separate lead fields for every subject and block were calculated in Fieldtrip. Lead fields were normalized in order to remove depth bias (Van Veen et al., 1997).

Sensor space analysis

Frequency data extraction

Sensor-space analyses were performed on separate epochs for the entrainment and target phases to avoid spectral leakage affecting the results. First, in order to assess neural tracking of speech prosody at delta-band frequency in the entrainment phase, we calculated pitch–MEG coherence at 0.6 Hz (rate of the SLOW contour) and 0.9 Hz (rate of the FAST contour). For this analysis, we cut out epochs starting from contour onset and corresponding to the duration of entrainment (5.19 s for SLOW and 3.32 s for FAST). All epochs had a high signal-to-noise ratio (SNR > 2). The F_0 envelope was calculated using Praat (Boersma & van Heuven, 2001). Since F_0 is undefined for pauses, those were interpolated using spline interpolation. Both the speech envelope and the MEG data underwent a fast Fourier transform (FFT) with a Hann window.

Since pitch–MEG coherence values were no longer accessible in the target phase (due to the acoustic stimulus not being present), we followed our hypothesis of sustained entrainment by quantifying power at the two frequencies of interest for both conditions in this phase. To this end, we segmented the data into epochs starting from visual sentence

onset and corresponding in duration to a single prosodic contour (1.73 s for SLOW and 1.10 s for FAST). These time windows were chosen in accordance with the hypothesis that cortical entrainment is sustained for the duration of at least one period at the respective frequency. Data from these epochs was also subjected to a fast Fourier transform with a Hann window. For both phases, FFT was restricted to a frequency range of 0-3 Hz. Because frequency comparison between 0.6 Hz and 0.9 Hz required a resolution of 0.1 Hz, we zero-padded each trial from every phase up to a duration of 10 seconds. Coherence (entrainment phase) and power (target phase) were then calculated for 31 bins with a step size of 0.1 Hz (Rosenberg et al., 1989). Visual inspection suggested that both the coherence and power spectra for the SLOW contour exhibited a peak at 0.6 Hz (the coherence spectrum also showed a peak at 1.2 Hz, which is the harmonics of 0.6 Hz), while the spectra for the FAST contour peaked at 0.9 Hz (Figure 2). The power spectrum in sensor space, however, could have been obscured by the presence of a 3 Hz rhythm induced by visual words (see Procedure); yet the peaks around the target frequencies are still clearly visible. The peaks become more clearly defined in the spectrum on the source level (see *Source analysis*).

Statistical frequency analysis

Statistical analysis for coherence and power was conducted on the two frequencies of interest (0.6 and 0.9 Hz) using nonparametric cluster-based permutation tests (Maris & Oostenveld, 2007), with 5,000 permutations (alpha = 0.05, ≥ 3 channels minimum cluster size). We used a template neighborhood structure for Elekta-Neuromag magnetometers. Because implementation of a two-way test is not straightforward in the cluster-permutation framework, we instead conducted multiple 2-tailed whole-brain paired-samples t-tests: 0.6

Hz vs. 0.9 Hz in SLOW, 0.6 Hz vs. 0.9 Hz in FAST, and a comparison of the difference 0.6 Hz vs. 0.9 Hz in SLOW to the difference 0.6 Hz vs. 0.9 Hz in FAST.

ERF data extraction

In order to investigate our hypothesis on the omission ERF in SHORT sentences after SLOW contours, we analyzed only the SHORT sentences at the onset of the missing verb (which was never presented) as a function of the two conditions of prior prosodic contour — SLOW or FAST; that is, the physical stimulus that was present during the measured ERPs was identical (i.e., no stimulus / omission), only the prior prosodic context differed. Therefore, the ERF omission effect could not be attributed to differences in the visual sentences themselves (as we used the identical sentences), but rather to the differences in prior prosodic condition. To this aim, we epoched the data from SHORT sentences in both prosodic conditions around the theoretical onset of the verb (-250 ms pre-onset, 1 s post-onset). Baseline correction was applied, demeaning the waveforms with the 250 ms interval preceding the onset of the verb. All epochs had a high signal-to-noise ratio (SNR > 2). ERFs were calculated within subject and within the FAST and SLOW conditions and subjected to statistical analysis.

Statistical ERF analysis

For statistical analysis, we employed a cluster-based permutation test on the time interval from 300 to 600 ms after the onset of the missing verb (5,000 permutations, alpha = 0.05, ≥ 3 channels minimum cluster size). The time window was chosen based on the previous literature regarding the timing of omission response (Penney, 2004; Ragazzoni et al.,

2019; Scharinger et al., 2017), as well as on a general consideration that omission response can bear characteristics of both P300 and P600 components (Nakano et al., 2014; cf. Sassenhagen & Fiebach, 2019).

Source-space analysis

Frequency data extraction

Source space analyses of coherence during the entrainment phase and power during the target phase were performed within block; reconstructions were averaged across blocks. To estimate source-level tracking of prosody in the entrainment phase, cross-spectral density (CSD) matrices were computed for all combinations of MEG channels and the pitch track using Fast Fourier Transforms (FFT) with Hann tapering; data were zero-padded to 20 s to achieve sufficient frequency resolution. To source-localize individual conditions while avoiding single-condition bias (Gross et al., 2001; van Vliet et al., 2018), we employed common spatial filters. First, CSD matrices were calculated separately for each frequency of interest (0.6 / 0.9 Hz), but common to the SLOW and FAST conditions. Based on the two CSD matrices, two spatial filters were then constructed using Dynamic Imaging of Coherent Sources (Gross et al., 2001) within the subject's volume conductor and source space. To source-localize the individual conditions, we then calculated individual CSD matrices for each of the four frequency × condition combinations in the entrainment phase. Next, we applied each of the two spatial filters (0.6 / 0.9 Hz) to the SLOW and FAST CSD matrices within that filter's frequency. To additionally localize the sustained entrainment effect from the sensor-space power analysis, source analysis of power in the target phase was conducted using an analogous pipeline.

Subsequently, single-node coherence values (entrainment phase) and power values (target phase) obtained from the beamformer analyses for every frequency and condition were averaged across each area provided by the individual atlas parcellation (360 areas, 180 per hemisphere). Finally, values were averaged across blocks for every participant and subjected to statistical analysis on the a-priori-defined regions of interest (ROI).

The ROI pattern comprised four regions, based on previous literature. Firstly, we looked at superior temporal gyrus (STG), known for its involvement in prosodic (Bourguignon et al., 2013; M. Meyer et al., 2002, 2004; Sammler et al., 2015) as well as general linguistic (Friederici, 2011; Friederici et al., 2003; Vigneau et al., 2006) processing. Secondly, we focused on early auditory cortex (EAC; Heschl's gyrus), which is responsible for general acoustic processing and is known to synchronize to speech prosody (M. Meyer et al., 2002, 2004). Thirdly, we investigated inferior frontal gyrus (IFG), closely related to general linguistic (Friederici et al., 2006; Goucha & Friederici, 2015; Van der Burght et al., 2021; Zaccarella & Friederici, 2015 etc.), prosodic (Frühholz & Grandjean, 2013; Sammler et al., 2015), as well as predictive (Avenanti et al., 2018; Jakuszeit et al., 2013; W. G. Matchin et al., 2017) processing. Finally, we looked at frontal operculum (FOC), also known to be involved in prosodic (Kotz et al., 2003; M. Meyer et al., 2002, 2004) and predictive (Bubic et al., 2009) processing. For the complete list of the atlas sub-areas that were combined into each respective ROI, please refer to Table 3-1.

For visual inspection of the coherence and power spectra in order to confirm the corresponding peaks at the frequencies of interest, we additionally reiterated the same source analysis procedure for each frequency in a limited spectrum of 0-3 Hz with a step of 0.1 Hz.

Statistical frequency analysis

To investigate statistical differences amongst frequencies and conditions within the ROIs, we used a series of linear mixed-effects regression models, implemented in R (R Core Team, 2019) in the lme4 package (Bates et al., 2015). The source reconstruction data was averaged across the above-mentioned ROIs and split by hemispheres. Two experimental phases were analyzed separately. First, to investigate the entrainment effect in every phase, we ran separate models on every ROI with CONDITION (SLOW vs. FAST) and FREQUENCY (0.6 vs. 0.9) and their interaction as predictors. The dependent variables were coherence (entrainment phase) and power (target phase). Since the entrainment effect proved to be significant in all ROIs (see Results), we further proceeded to estimate in which area/hemisphere the effect was statistically higher. To this end, we required a measure which permits to quantify the overall strength of entrainment in a particular region. For this purpose, we adapted an index termed "rate-specific response" from previous literature (RSR; (van Bree et al., 2021)). This index is calculated as follows:

$$RSR = (R_{0.6,\ SLOW} - R_{0.6,\ FAST}) + (R_{0.9,\ FAST} - R_{0.9,\ SLOW})$$

where R is the response measure (coherence or power) and 0.6/0.9 Hz and SLOW/FAST are respectively the frequencies of interest and conditions, for which this response measure is calculated. An RSR is larger than 0 when electrophysiological brain activity follows the initial speech rate for both conditions (stronger 0.6 Hz coherence/power for the SLOW condition and stronger 0.9 Hz coherence/power for the FAST condition). That is, higher RSR indicates a brain response that is specific, or entrained, to the stimulus rate. We subjected the resulting RSR indices for every area and participant to new linear mixed-

effects regression models (a separate model for every phase), where the predictors were now AREA, HEMISPHERE and their interaction. The depending variables were now corresponding RSR indices, reflecting the entrainment strength.

For all models, predictors were coded using mean-centered effects coding. Random intercepts were included for subjects. The models were further subjected to an ANOVA analysis in order to determine the significance of the predictors.

ERF data extraction

To source-localize the omission ERF effect, we used the Linear Constrained Minimum Variance beamformer (LCMV; Van Veen et al., 1997). As the omission can only be studied for SHORT sentences, we first constructed a covariance matrix for the SHORT sentences from both entrainment conditions in the ERF time window (-250 to 1000 ms around the potential onset of the critical word; baseline corrected). This matrix, together with individual volume conductors and lead fields, was used to compute the common spatial filter for every block of data. Next, source reconstruction was performed separately for every condition using the precomputed common filter and the individual covariance matrix corresponding to the condition. The resulting single-voxel time-courses were averaged across every atlas area and further averaged across blocks for every subject. From these subject-level time-courses, we then selected and averaged data across the time window of interest (300-600 ms).

Statistical ERF analysis

Since we did not have a strict pre-defined hypothesis for the source localization of the omission ERF, we did not pre-select any ROIs. Instead, for every atlas area we performed a two-tailed paired-samples t-test in order to determine the differences between SLOW and FAST entrainment conditions (alpha = 0.05). For illustration, we report the ten regions across hemispheres, where the difference between conditions was maximally significant (p-values uncorrected).

Results

Delta-band activity synchronizes with speech prosody

<u>*Sensor space – frequency analysis*</u>

During the entrainment phase, analysis for the 0.6 vs. 0.9 Hz contrast in SLOW revealed a significant positive cluster (cluster-sum t(34) = 258, cluster-level p < 0.001, corrected; peak at sensor MEG2441, peak-level t(34) = 10, peak-level p < 0.001; for MEG sensors layout, see Figure 2-1). This cluster comprised all but six sensors. In turn, the 0.6 vs. 0.9 Hz contrast in FAST revealed a significant negative cluster (cluster-sum t(34) = - 291, cluster-level p < 0.001, corrected; peak at sensor MEG1421, peak-level t(34) = - 11.91, peak-level p < 0.001). This cluster also included all but nine sensors. Finally, the comparison of difference maps (i.e., SLOW vs. FAST) yielded a significant positive cluster including all but one sensors (cluster-sum t(34) = 396, cluster-level p < 0.001, corrected; peak at sensor MEG2221, peak-level t(34) = 12.33, peak-level p < 0.001). Peak sensors for SLOW and difference clusters were located in the right parietal region, while the peak sensor for the FAST cluster was located in the right fronto-temporal region. Overall, these results suggest that brain activity synchronizes to the rate of the external prosodic contour (Figure 3A).

Figure 3 here

Source space – frequency analysis

Source space analysis revealed a significant CONDITION × FREQUENCY interaction, as well as a main effect of CONDITION in every ROI. Post-hoc tests within every condition (0.6 vs. 0.9 contrast) showed higher coherence at 0.6 Hz in SLOW condition and at 0.9 Hz in FAST condition (Table 1; Figure 4A). These results show that brain synchronization to the pitch contour was significant in all ROIs. To further quantify differences in the entrainment effect between areas and hemispheres, we computed the rate-specific response (RSR) index, which reflects overall entrainment strength for every area (van Bree et al., 2021). Using this index as an outcome, we ran a new linear mixed-effects model with AREA, HEMISPHERE and the AREA × HEMISPHERE interaction as predictors. Our model showed a significant AREA × HEMISPHERE interaction (F = 10.5, df = 3,245, p < 0.001), as well as main effects for both AREA (F = 31.18, df = 3,245, p < 0.001) and HEMISPHERE (F = 72.68, df = 1,245, p < 0.001). Post-hoc testing revealed that entrainment was stronger in the right EAC (T = 5.25, df = 245, p < 0.001), and STG (T = 2.83, df = 245, p < 0.001). These results are consistent with previous findings on prosodic tracking in right temporal regions (Bourguignon et al., 2013).

Figure 4 here

Table 1 here

Visual inspection of the coherence spectrum in the corresponding ROIs (Figure 5) confirm the coherence peaks at 0.6 Hz and 1.2 Hz (harmonics of 0.6) for the SLOW condition, and at 0.9 Hz for the FAST condition. The coherence values are higher in right STG and EAC, which corresponds to our statistical analysis results.

Figure 5 here

Delta-band activity carriers the contextual rhythm into the future

Sensor space – frequency analysis

During the target phase, sensor-space statistics for the 0.6 vs. 0.9 Hz contrast in SLOW revealed a significant positive cluster with a fronto-central-occipital topography, shifted to the right (cluster-sum t(34) = 66, cluster-level p < 0.001, corrected; peak at sensor MEG2641, peak-level t(34) = 5.18, peak-level p < 0.001; Figure 3B; for MEG sensors layout, see Figure 2-1). In turn, the 0.6 vs. 0.9 Hz contrast in FAST exposed a negative cluster (cluster-sum t(34) = -942, cluster-level p < 0.001, corrected; peak at sensor MEG2621, peak-level t(34) = -18.58, peak-level p < 0.001). Finally, the comparison of difference maps (i.e., SLOW vs. FAST) revealed a positive cluster (cluster-sum t(34) = 706, cluster-level p < 0.001, corrected; peak at sensor MEG1731, peak-level t(34) = 13.3, peak-level p < 0.001). The latter two clusters comprised all sensors. Peak sensors for SLOW and FAST clusters were located in the right temporal region, while the peak sensor for the difference cluster was located in the left parieto-occipital region. Taken together,

sensor-space results in the target phase strongly suggest that entrainment to prosody persists beyond stimulus offset.

Source space – frequency analysis

Source-level analyses showed a CONDITION × FREQUENCY interaction, as well as a main effect of CONDITION for every ROI. Post-hoc t-tests for frequencies within every condition (0.6 vs. 0.9 contrast) revealed that power was significantly higher at 0.6 Hz in SLOW and at 0.9 Hz in FAST (Table 2; Figure 4B). Linear mixed-effects model for the differences between areas and hemispheres revealed a main effect of AREA ($F = 38.82$, $df = 3,245$, $p < 0.001$) and an AREA × HEMISPHERE interaction ($F = 13.76$, $df = 3,245$, $p < 0.001$). Post-hoc testing showed that entrainment power was higher in left EAC ($T = 2.82$, $df = 245$, $p = 0.005$) and left FOP ($T = 4.52$, $df = 245$, $p < 0.001$). This arrangement of source-space activity is compatible with our hypothesis, as frontal regions are known to be associated with predictions and cognitive control (Alexander & Brown, 2018; Dürschmid et al., 2019). Our study further confirms evidence from the literature concerning a trend for the left lateralization of those functions (Avenanti et al., 2018; Jakuszeit et al., 2013; W. Matchin et al., 2017).

Table 2 here

Visual inspection of the power spectrum in the corresponding ROIs (Figure 6) confirm the power peaks at 0.6 Hz for the SLOW condition and at 0.9 Hz for the FAST condition. The

power values are higher in left FOP and EAC, which corresponds to our statistical analysis results.

Figure 6 here

Falsified predictions associate with an omission effect

Sensor space – ERF analysis

The ERF analysis showed a significant positive right central-parietal cluster (cluster-sum t(34) = 23, cluster-level p < 0.001, corrected; peak at sensor MEG2231, peak-level t(34) = 5.29, peak-level p < 0.001, corrected), as well as a significant neighboring negative cluster (cluster-sum t(34) = -7.9, cluster-level p = 0.019, corrected; peak at sensor MEG0221, peak-level t(34) = -3.57, peak-level p = 0.001, corrected) (Figure 7B left; for MEG sensors layout, see Figure 2-1). Note that the polarity of the cluster in ERF analysis is not informative about the direction of the effect: In both clusters, SLOW condition induced higher-amplitude ERF waveforms than FAST (Figure 7B right). Therefore, we can infer that participants did indeed predict longer sentences under the SLOW entrainment and the incongruent condition caused an omission ERF effect.

Source space – ERF analysis

Source space analysis revealed multiple areas where the activation for the SLOW condition was significantly different (uncorrected) from the FAST condition; we here report 10 regions with maximum differences (Table 3). The areas where the activation for SLOW

condition was significantly higher than for the FAST condition (positive T-values) include temporo-parietal-occipital junction and a number of frontal and premotor regions (Figure 7B middle). Overall, the pattern of source level activity suggests that entrainment is originating in right temporal cortices and maintained in left frontal cortex, which generates temporal predictions; in case those predictions are incongruent, an M300 ERF is emerging from a variety of cortical generators.

Figure 7 here

Table 3 here

Additional analyses

Source space – frequency analysis

In addition to temporal prediction processing being associated with IFG and FOP, previous literature has also demonstrated the potential involvement of motor cortex in temporal predictions. In particular, evidence suggests that temporal predictions in the motor cortex, subserved by beta-band oscillations, can upregulate auditory-motor interaction via top-down phase reset of the oscillations in the auditory cortex (Assaneo et al., 2021; Keitel et al., 2018; Morillon & Baillet, 2017). These oscillations are thought to function as an endogenous temporal constraint, facilitating bottom-up processing (Rimmele et al., 2018). Furthermore, in the current paradigm, predictive processing as a stage of decision-making in order to answer the comprehension question could potentially involve motor preparation. In order to investigate these hypotheses, we conducted a control analysis of the entrainment effect in pre-motor (Brodmann area 6) and motor (Brodmann area 4) areas (for the list of the atlas sub-areas, please refer to Table 3-1). The results demonstrated significant entrainment in both areas and hemispheres (Figure 8).

Figure 8 here

Statistical analyses showed a CONDITION × FREQUENCY interaction, as well as a main effect of CONDITION for every ROI (Table 5). Post-hoc t-tests for frequencies within every condition (0.6 vs. 0.9 contrast) also confirmed that power was significantly higher at 0.6 Hz in SLOW and at 0.9 Hz in FAST, explaining the interaction effect.

ERF analysis

In addition to later effects, prosodic processing has also been demonstrated at very early stages (Tomasello et al., 2022). In order to investigate whether those effects persist in the target phase of prosodic entrainment, we conducted an additional ERF analysis in the time window of 0-120 ms after the critical word onset on the same data as used for the M300 analysis. The results revealed significant differences in a central positive cluster, with higher activation for the SLOW condition (Figure 7A left). This effect can be interpreted as early omission response, akin to prosody expectation violation (Kakouros et al., 2018; Paulmann et al., 2012) induced by prosodic entrainment.

We further attempted a source reconstruction on the early ERF response (Figure 7A middle). We report 10 regions where the ERF difference between the SLOW and FAST conditions was maximal (uncorrected) (Table 4). The pattern appears quite similar to the source pattern obtained for the 300-600 ms omission ERF (Figure 7B middle), with higher activation for the SLOW condition in frontal areas on the left and temporo-parietal-occipital junction on the right. The TPOJ activation could be potentially be related to visual omission effects (the omitted word was expected to be presented in a visual mode). However, we here need to note that source reconstruction of ERF with beamforming can potentially be problematic, since this approach underestimates highly covariant sources, and does not consider the degree to which estimated source signals explain observed sensor signals (Kuznetsova et al., 2021; O'Reilly et al., 2023). The current results, therefore, should be treated with caution.

Discussion

Our results suggest that speech prosody subserves temporal predictions in language via electrophysiological inheritance of contextual temporal patterns. This converges on recent evidence for similar mechanisms supporting temporal predictions in the domain of auditory perception more generally (Breska & Deouell, 2017; Herbst et al., 2022; Stefanics et al., 2010). The facilitatory effect of delta-band entrainment on non-linguistic auditory perception (e.g. tone sequences) has been shown earlier (Henry & Obleser, 2012; Hickok et al., 2015). While in the preceding studies the effect was demonstrated for auditory tones, we here report delta-band entrainment in the context of language processing. To our knowledge, earlier reports of behaviorally relevant sustained linguistic entrainment have been published only for higher frequencies only (Kösem et al., 2018; van Bree et al., 2021). The setup of the current study allows us to adopt a strong interpretation in terms of entrainment proper (L. Meyer et al., 2020b, 2020a; Obleser & Kayser, 2019; Rimmele et al., 2018) — that is, electrophysiological neural activity does not just mimic stimulation frequencies at the time of stimulation, maintains these for a subsequent period (Kösem et al., 2018; Stefanics et al., 2010; van Bree et al., 2021).

The present findings stress the necessity of incorporating mechanisms of temporal prediction into current neurobiological, psycholinguistic and computational models of human speech and language (Alexander & Brown, 2018; Friederici & Alter, 2004; Hickok & Poeppel, 2007; Ten Oever & Martin, 2021). Our source-space results are consistent with the notion that such temporal predictions operationalize auditory-motor interaction (Assaneo et al., 2021; Keitel et al., 2018; Kern et al., 2021; Lubinus et al., 2023). Recent

studies have proposed models where temporal predictions operate through top-down phase resetting of electrophysiological activity mediated by the motor system or higher-order language– or attention-related systems (for a review, see Rimmele et al., 2018). This mechanism facilitates bottom-up perceptual processing, while intrinsic temporal characteristics of neural activity might serve as intrinsic temporal constraint. Furthermore, motor cortices have also been related to predictive processing on a broader scale (Grisoni et al., 2017, 2021). Interestingly, it has been shown that their engagement is dependent on the specific nature of the predictions being made—for instance, frontocentral-sensorimotor activation for tool-related words (Grisoni et al., 2021), dorsolateral hand motor activation for hand-related words (e.g., "write"), ventral motor activation for face-related words ("talk") (Grisoni et al., 2017). Our additional analysis confirmed a sustained entrainment effect in the motor cortices.

Moreover, studies on lateralization of predictive processes indicate that bottom-up auditory processing is primarily associated with auditory cortices in the right hemisphere, while top-down control tends to be left-lateralized (Assaneo et al., 2019; Keitel et al., 2018). This is in line with our results: In the entrainment phase, right STG and EAC showed significantly higher coherence with the pitch contour. These findings are also consistent with the assumed neural substrate of prosody (Bourguignon et al., 2013; M. Meyer et al., 2004; Sammler et al., 2015). In the target phase, however, activity at the conditioning frequency moved to left frontal cortices, involved in temporal prediction in language (Jakuszeit et al., 2013; W. Matchin et al., 2017). Left FOP activity is also generally known to correspond to syntactic processing or prediction (Bubic et al., 2009; Friederici et al., 2003, 2006), which in our paradigm is engaged by prior entrainment. Sustained entrainment in frontal areas could also potentially reflect the broader engagement of the frontal cortex in decision-making regarding the experimental task. Temporal prediction

could be the first and integral stage of decision-making processing, facilitating the upcoming sentence analysis and, therefore, forming preliminary answers to the comprehension question.

Our interpretation of the omission response in terms of a temporal prediction error is in line with previous reports from auditory perception (Herbst et al., 2022; Stefanics et al., 2010), where temporal predictions carried by delta-band activity modulated the amplitude of the P300 component on target tones. The current ERF effect is likely an omitted stimulus potential (OSP), bearing characteristics of both P300 and P600 components (or M300/M600, as their MEG counterparts; Bullock et al., 1994; Karamürsel & Bullock, 2009; Nakano et al., 2014; Penney, 2004; Ragazzoni et al., 2019). This interpretation conforms to the idea that the P600 component in the language domain is a member of the greater P300 family (Sassenhagen & Fiebach, 2019). The widely distributed array of underlying cortical sources observed here is in line with prior reconstructions of the P300, which include diverse distributed generators (Bocquillon et al., 2011; Brázdil et al., 2005; Linden, 2005; Ragazzoni et al., 2019). The temporal-parietal-occipital junction as a generator of P300 was previously reported in multiple deviating stimuli studies, together with dorsolateral prefrontal areas (Daffner et al., 2003; Strobel et al., 2008; Volpe et al., 2007). Moreover, a study on semantic predictions has also previously reported temporo-parieto-occipital activation (Grisoni et al., 2021). A visual oddball study (Bledowski et al., 2004) also reports activation in the frontal eye field, similar to our results. These findings are in line with our with our conclusions, since, in our paradigm, the stimulation in the target phase was in the visual modality. Moreover, multiple omitted-target studies also find activation in the inferior frontal cortex (Bledowski et al., 2004; Ragazzoni et al., 2019; Volpe et al., 2007), which is corroborated by our

results. This evidence further strengthens our claim of the left frontal cortex involvement in predictive processing (Jakuszeit et al., 2013; W. Matchin et al., 2017).

Our complementary analysis also revealed a significant earlier omission response in the time window of 0-120 ms. This could be potentially related to prosodic entrainment inducing early omission effects, similar to prosody expectation violations (Kakouros et al., 2018; Paulmann et al., 2012; Tomasello et al., 2022). The source localization pattern of this earlier ERF resembles that of the M300/M600 omission effect; however, it is important to note that source localization procedures for ERFs should be regarded with caution (Kuznetsova et al., 2021; O'Reilly et al., 2023).

The current study bears important implications for the general conceptualization of the neurobiological underpinnings of interpersonal communication. Prosodic entrainment could potentially be used as a facilitatory mechanism in dialogue, enhancing mutual comprehension (Breen, Weidman, et al., 2016; Edlund, 2011; Lehnert-LeHouillier et al., 2020). In line with the implicit prosody account (Breen, 2014; Fodor, 2002; Perrone-Bertolotti et al., 2014), in the current experiment, based on prior prosodic entrainment listeners would silently generate the same prosodic structures during the reading of subsequent sentence, which would influence their temporal prediction regarding the length of the sentence. This mechanism can be applied to communication in a dialogue: the listener could potentially entrain to the speaker's prosody, generating silent prosodic structures and forming temporal predictions as to what the speaker is going to say next.

These findings could be incorporated into broader models of alignment in dialogue (Pickering & Garrod, 2004, 2021). This account postulates that interlocutors in a dialogue utilize linguistic representations which automatically align on many levels, facilitating production and comprehension processes. In principle, this general mechanism could operate over various acoustic features, including rate, intensity, voice quality, and pitch

(Brennan, 1996; Teoh et al., 2019). These features need to exhibit sufficient rhythmicity to facilitate conversational entrainment. Although ubiquitous rhythmicity is not characteristic of human speech, previous research has demonstrated that speech is rhythmic enough on the prosodic level to trigger entrainment in dialogue (Inbar et al., 2020; Stehwien & Meyer, 2021); furthermore, rhythmic intonational units in speech evoke a neural response (Inbar et al., 2023). Further research quantifying the rhythmicity of naturalistic speech and investigating the differences in prosodic entrainment for rhythmic and non-rhythmic parts could provide additional evidence for this claim.

Work on interpersonal communication also supports the notion that prosody is involved in conversational entrainment (Levitan et al., 2012; Levitan & Hirschberg, 2011; Reichel et al., 2018). It has previously been demonstrated that prosody conveys communicative intentions in speech irrespective of emotion (Hellbernd & Sammler, 2016). Neurobiologically these effects have been localized to auditory ventral stream and articulatory-motor regions (Hellbernd & Sammler, 2018; Tomasello et al., 2022). On a more general level, it has been proposed that human brains synchronize communicative activities through an oscillatory signal (Hasson et al., 2012; Wilson & Wilson, 2005). Furthermore, it has not only been found that listener's brain activity is spatially and temporally coupled with the speaker's, but that listener's frontal areas also exhibit predictive anticipatory responses during dialogue (Stephens et al., 2010). Our present results, showing prosodic entrainment as a mechanism of temporal predictions in listening comprehension, could therefore extend to scenarios in dialogue. Altogether, prosodic entrainment appears to play an important role in enhancing communication processes via temporal prediction.

The focus of current study has been the investigation of purely prosodic entrainment. However, it is pertinent to note that potentially similar effects could be evoked by

presenting non-prosodic stimuli of fixed rhythmic duration (e.g., acoustic stimuli with a flat pitch or visual stimulation). While the timing of the experiment and the necessity of a substantial number of trials did not permit us to include a control non-prosodic condition in the current paradigm, future work would be beneficial to distinguish between prosodic vs. merely acoustic/visual entrainment effects.

References

Alexander, W. H., & Brown, J. W. (2018). Frontal cortex function as derived from hierarchical predictive coding. *Scientific Reports*, *8*(1), 1–11. https://doi.org/10.1038/s41598-018-21407-9

Assaneo, M. F., Rimmele, J. M., Sanz Perl, Y., & Poeppel, D. (2021). Speaking rhythmically can shape hearing. *Nature Human Behaviour*, *5*(1), 71–82. https://doi.org/10.1038/s41562-020-00962-0

Assaneo, Rimmele, J. M., Orpella, J., Ripollés, P., de Diego-Balaguer, R., & Poeppel, D. (2019). The lateralization of speech-brain coupling is differentially modulated by intrinsic auditory and top-down mechanisms. *Frontiers in Integrative Neuroscience*, *13*(July), 1–11. https://doi.org/10.3389/fnint.2019.00028

Avenanti, A., Paracampo, R., Annella, L., Tidoni, E., & Aglioti, S. M. (2018). Boosting and decreasing action prediction abilities through excitatory and inhibitory tDCS of inferior frontal cortex. *Cerebral Cortex*, *28*(4), 1282–1296. https://doi.org/10.1093/cercor/bhx041

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v067.i01

Besl, P. J., & McKay, N. D. (1992). A Method for Registration of 3-D Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *14*(2), 239–256. https://doi.org/10.1109/34.121791

Bledowski, C., Prvulovic, D., Hoechstetter, K., Scherg, M., Wibral, M., Goebel, R., & Linden, D. E. J. (2004). Localizing P300 generators in visual target and distractor processing: A combined event-related potential and functional magnetic resonance imaging study. *Journal of Neuroscience*, *24*(42), 9353–9360. https://doi.org/10.1523/JNEUROSCI.1897-04.2004

Bocquillon, P., Bourriez, J. L., Palmero-Soler, E., Betrouni, N., Houdayer, E., Derambure, P., & Dujardin, K. (2011). Use of swLORETA to localize the cortical sources of target- and

distracter-elicited P300 components. *Clinical Neurophysiology*, *122*(10), 1991–2002. https://doi.org/10.1016/j.clinph.2011.03.014

Boersma, P., & van Heuven, V. (2001). Speak and unSpeak with Praat. *Glot International*, *5*(9–10), 341–347.

Boston, M. F., Hale, J. T., Vasishth, S., & Kliegl, R. (2011). Parallel processing and sentence comprehension difficulty. *Language and Cognitive Processes*, *26*(3), 301–349. https://doi.org/10.1080/01690965.2010.492228

Bourguignon, M., De Tiège, X., De Beeck, M. O., Ligot, N., Paquier, P., Van Bogaert, P., Goldman, S., Hari, R., & Jousmäki, V. (2013). The pace of prosodic phrasing couples the listener's cortex to the reader's voice. *Human Brain Mapping*, *34*(2), 314–326. https://doi.org/10.1002/hbm.21442

Brázdil, M., Dobšík, M., Mikl, M., Hluštík, P., Daniel, P., Pažourková, M., Krupa, P., & Rektor, I. (2005). Combined event-related fMRI and intracerebral ERP study of an auditory oddball task. *NeuroImage*, *26*(1), 285–293. https://doi.org/10.1016/j.neuroimage.2005.01.051

Breen, M. (2014). Empirical investigations of the role of implicit prosody in sentence processing. *Linguistics and Language Compass*, *8*(2), 37–50. https://doi.org/10.1111/lnc3.12061

Breen, M., Kaswer, L., Van Dyke, J. A., Krivokapić, J., & Landi, N. (2016). Imitated prosodic fluency predicts reading comprehension ability in good and poor high school readers. *Frontiers in Psychology*, *7*(JUL), 1026. https://doi.org/10.3389/fpsyg.2016.01026

Breen, M., Weidman, S., Haydon, K. C., Breen, M., & Haydon, K. C. (2016). Prosodic speech entrainment in romantic relationships. *Proceedings of the International Conference on Speech Prosody, 2016-Janua*, 508–512. https://www.researchgate.net/publication/305684684

Brennan, S. E. (1996). Lexical Entrainment in Spontaneous Dialog. *Proceedings, 1996 International Symposium on Spoken Dialogue, ISSD-96*, 41–44.

Breska, A., & Deouell, L. Y. (2017). Neural mechanisms of rhythm-based temporal prediction:

Delta phase-locking reflects temporal predictability but not rhythmic entrainment. *PLoS Biology*, *15*(2), 1–30. https://doi.org/10.1371/journal.pbio.2001665

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychonomic Bulletin and Review*, *18*(6), 1189–1196. https://doi.org/10.3758/s13423-011-0167-9

Bubic, A., Von Yves Cramon, D., Jacobsen, T., Schröger, E., & Schubotz, R. I. (2009). Violation of expectation: Neural correlates reflect bases of prediction. *Journal of Cognitive Neuroscience*, *21*(1), 155–168. https://doi.org/10.1162/jocn.2009.21013

Bullock, T. H., Karamürsel, S., Achimowicz, J. Z., McClune, M. C., & Başar-Eroglu, C. (1994). Dynamic properties of human visual evoked and omitted stimulus potentials. *Electroencephalography and Clinical Neurophysiology*, *91*(1), 42–53. https://doi.org/10.1016/0013-4694(94)90017-5

Clifton, C., Carlson, K., & Frazier, L. (2002). Informative prosodic boundaries. *Language and Speech*, *45*(2), 87–114. https://doi.org/10.1177/00238309020450020101

Daffner, K. R., Scinto, L. F. M., Weitzman, A. M., Faust, R., Rentz, D. M., Budson, A. E., & Holcomb, P. J. (2003). Frontal and Parietal Components of a Cerebral Network Mediating Voluntary Attention to Novel Events. *Journal of Cognitive Neuroscience*, *15*(2), 294–313. https://doi.org/10.1162/089892903321208213

de Cheveigné, A. (2020). ZapLine: A simple and effective method to remove power line artifacts. *NeuroImage*, *207*, 116356. https://doi.org/10.1016/j.neuroimage.2019.116356

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience 2005 8:8*, *8*(8), 1117–1121. https://doi.org/10.1038/nn1504

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of

syntactic processing complexity. *Cognition*, *109*(2), 193–210.

https://doi.org/10.1016/J.COGNITION.2008.07.008

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and

lexical processing. *Journal of Memory and Language*, *59*(3), 294–311.

https://doi.org/10.1016/j.jml.2008.06.006

Dürschmid, S., Reichert, C., Hinrichs, H., Heinze, H. J., Kirsch, H. E., Knight, R. T., & Deouell, L.

Y. (2019). Direct Evidence for Prediction Signals in Frontal Cortex Independent of Prediction

Error. *Cerebral Cortex*, *29*(11), 4530–4538. https://doi.org/10.1093/cercor/bhy331

Edlund, J. (2011). *In search for the conversational homunculus: serving to understand spoken*

human face-to-face interaction. http://www.diva-

portal.org/smash/record.jsf?pid=diva2:402997

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis. II: Inflation,

flattening, and a surface-based coordinate system. *NeuroImage*, *9*(2), 195–207.

https://doi.org/10.1006/NIMG.1998.0396

Fodor, J. D. (2002). Psycholinguistics cannot escape prosody. *Proceedings of the 1st*

International Conference on Speech Prosody, *April 2002*, 83–88.

Frazier, L., Carlson, K., & Clifton, C. (2006). Prosodic phrasing is central to language

comprehension. *Trends in Cognitive Sciences*, *10*(6), 244–249.

https://doi.org/10.1016/j.tics.2006.04.002

Friederici, A. D. (2011). The brain basis of language processing: From structure to function.

Physiological Reviews, *91*(4), 1357–1392. https://doi.org/10.1152/physrev.00006.2011

Friederici, A. D., & Alter, K. (2004). Lateralization of auditory language functions: A dynamic dual

pathway model. *Brain and Language*, *89*(2), 267–276. https://doi.org/10.1016/S0093-

934X(03)00351-1

Friederici, A. D., Fiebach, C. J., Schlesewsky, M., Bornkessel, I. D., & Von Cramon, D. Y. (2006).

Processing linguistic complexity and grammaticality in the left frontal cortex. *Cerebral Cortex*, *16*(12), 1709–1717. https://doi.org/10.1093/cercor/bhj106

Friederici, A. D., Rüschemeyer, S. A., Hahne, A., & Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: Localizing syntactic and semantic processes. *Cerebral Cortex*, *13*(2), 170–177. https://doi.org/10.1093/cercor/13.2.170

Frühholz, S., & Grandjean, D. (2013). Processing of emotional vocalizations in bilateral inferior frontal cortex. In *Neuroscience and Biobehavioral Reviews* (Vol. 37, Issue 10, pp. 2847–2855). https://doi.org/10.1016/j.neubiorev.2013.10.007

Garcés, P., López-Sanz, D., Maestú, F., & Pereda, E. (2017). Choice of Magnetometers and Gradiometers after Signal Space Separation. *Sensors (Basel, Switzerland)*, *17*(12). https://doi.org/10.3390/S17122926

Ghitza, O. (2017). Acoustic-driven delta rhythms as prosodic markers. *Language, Cognition and Neuroscience*, *32*(5), 545–561. https://doi.org/10.1080/23273798.2016.1232419

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., Smith, S. M., & Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171. https://doi.org/10.1038/NATURE18933

Goucha, T., & Friederici, A. D. (2015). The language skeleton after dissecting meaning: A functional segregation within Broca's Area. *NeuroImage*, *114*, 294–302. https://doi.org/10.1016/J.NEUROIMAGE.2015.04.011

Grisoni, L., Miller, T. M., & Pulvermüller, F. (2017). Neural Correlates of Semantic Prediction and Resolution in Sentence Processing. *The Journal of Neuroscience*, *37*(18), 4848–4858. https://doi.org/10.1523/JNEUROSCI.2800-16.2017

Grisoni, L., Tomasello, R., & Pulvermüller, F. (2021). 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

Grosjean, F. (1983). How long is the sentence? prediction and prosody in the on-line processing of language. *Linguistics*, *21*(3), 501–530. https://doi.org/10.1515/ling.1983.21.3.501

Gross, J., Kujala, J., Hämäläinen, M., Timmermann, L., Schnitzler, A., & Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(2), 694–699. https://doi.org/10.1073/pnas.98.2.694

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). Brain-to-brain coupling: A mechanism for creating and sharing a social world. In *Trends in Cognitive Sciences* (Vol. 16, Issue 2, pp. 114–121). NIH Public Access. https://doi.org/10.1016/j.tics.2011.12.007

Heister, J., Würzner, K. M., Bubenzer, J., Pohl, E., Hanneforth, T., Geyken, A., & Kliegl, R. (2011). dlexDB - Eine lexikalische datenbank für die psychologische und linguistische forschung. *Psychologische Rundschau*, *62*(1), 10–20. https://doi.org/10.1026/0033-3042/a000029

Hellbernd, N., & Sammler, D. (2016). Prosody conveys speaker's intentions: Acoustic cues for speech act perception. *Journal of Memory and Language*, *88*, 70–86. https://doi.org/10.1016/j.jml.2016.01.001

Hellbernd, N., & Sammler, D. (2018). Neural bases of social communicative intentions in speech. *Social Cognitive and Affective Neuroscience*, *13*(6), 604–615. https://doi.org/10.1093/scan/nsy034

Henry, M. J., & Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(49), 20095–20100.

https://doi.org/10.1073/pnas.1213390109

Herbst, S. K., Stefanics, G., & Obleser, J. (2022). Endogenous modulation of delta phase by expectation–A replication of Stefanics et al., 2010. *Cortex*, *149*, 226–245. https://doi.org/10.1016/j.cortex.2022.02.001

Hickok, G., Farahbod, H., & Saberi, K. (2015). The Rhythm of Perception: Entrainment to Acoustic Rhythms Induces Subsequent Perceptual Oscillation. *Psychological Science*, *26*(7), 0956797615576533. https://doi.org/10.1177/0956797615576533

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

Inbar, M., ..., & Landau, A. N. (2023). Intonation Units in spontaneous speech evoke a neural response Maya. *BioRxiv*, 2023.01.26.525707. https://doi.org/10.1101/2023.01.26.525707

Inbar, M., Grossman, E., & Landau, A. N. (2020). Sequences of Intonation Units form a ~ 1 Hz rhythm. *Scientific Reports*, *10*(1), 1–23. https://doi.org/10.1038/s41598-020-72739-4

Jakuszeit, M., Kotz, S. A., & Hasting, A. S. (2013). Generating predictions: Lesion evidence on the role of left inferior frontal cortex in rapid syntactic analysis. *Cortex*, *49*(10), 2861–2874. https://doi.org/10.1016/j.cortex.2013.05.014

Kakouros, S., Salminen, N., & Räsänen, O. (2018). Making predictable unpredictable with style – Behavioral and electrophysiological evidence for the critical role of prosodic expectations in the perception of prominence in speech. *Neuropsychologia*, *109*, 181–199. https://doi.org/10.1016/j.neuropsychologia.2017.12.011

Karamürsel, S. (2000). Human auditory fast and slow omitted stimulus potentials and steady-state responses. *International Journal of Neuroscience*, *100*(1–4), 1–20. https://doi.org/10.3109/00207450008999674

Keitel, A., Gross, J., & Kayser, C. (2018). Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLoS Biology*, *16*(3), e2004473.

https://doi.org/10.1371/journal.pbio.2004473

Kern, P., Assaneo, M. F., Endres, D., Poeppel, D., & Rimmele, J. M. (2021). Preferred auditory temporal processing regimes and auditory-motor synchronization. *Psychonomic Bulletin and Review*, *28*(6), 1860–1873. https://doi.org/10.3758/s13423-021-01933-w

Klasmeyer, G. (1997). Perceptual importance of selected voice quality parameters. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *3*, 1615–1618. https://doi.org/10.1109/ICASSP.1997.598808

Kösem, A., Bosker, H. R., Takashima, A., Meyer, A., Jensen, O., & Hagoort, P. (2018). Neural Entrainment Determines the Words We Hear. *Current Biology*, *28*(18), 2867-2875.e3. https://doi.org/10.1016/j.cub.2018.07.023

Kotz, S. A., Meyer, M., Alter, K., Besson, M., Von Cramon, D. Y., & Friederici, A. D. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain and Language*, *86*(3), 366–376. https://doi.org/10.1016/S0093-934X(02)00532-1

Kuznetsova, A., Nurislamova, Y., & Ossadtchi, A. (2021). Modified covariance beamformer for solving MEG inverse problem in the environment with correlated sources. *NeuroImage*, *228*, 117677. https://doi.org/10.1016/j.neuroimage.2020.117677

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, *320*(5872), 110–113. https://doi.org/10.1126/science.1154735

Lamekina, Y., & Meyer, L. (2022). Entrainment to speech prosody influences subsequent sentence comprehension. *Language, Cognition and Neuroscience*. https://doi.org/10.1080/23273798.2022.2107689

Lau, E. F., Weber, K., Gramfort, A., Hämäläinen, M. S., & Kuperberg, G. R. (2016). Spatiotemporal Signatures of Lexical–Semantic Prediction. *Cerebral Cortex*, *26*(4), 1377–1387. https://doi.org/10.1093/CERCOR/BHU219

Lehnert-LeHouillier, H., Terrazas, S., & Sandoval, S. (2020). Prosodic Entrainment in Conversations of Verbal Children and Teens on the Autism Spectrum. *Frontiers in Psychology*, *11*, 2718. https://doi.org/10.3389/fpsyg.2020.582221

Levitan, R., Gravano, A., Willson, L., Beňuš, Š., Hirschberg, J., & Nenkova, A. (2012). Acoustic-prosodic entrainment and social behavior. *NAACL HLT 2012 - 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference*, 11–19. http://www.mturk.com

Levitan, R., & Hirschberg, J. (2011). Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 3081–3084. https://doi.org/10.21437/interspeech.2011-771

Li, Z., & Zhang, D. (2023). How does the human brain process noisy speech in real life? Insights from the second-person neuroscience perspective. *Cognitive Neurodynamics 2023*, 1–12. https://doi.org/10.1007/S11571-022-09924-W

Linden, D. E. J. (2005). The P300: Where in the brain is it produced and what does it tell us? In *Neuroscientist* (Vol. 11, Issue 6, pp. 563–576). Sage PublicationsSage CA: Thousand Oaks, CA. https://doi.org/10.1177/1073858405280524

Lubinus, C., Keitel, A., Obleser, J., Poeppel, D., & Rimmele, J. M. (2023). Explaining flexible continuous speech comprehension from individual motor rhythms. *Proceedings of the Royal Society B: Biological Sciences*, *290*(1994). https://doi.org/10.1098/rspb.2022.2410

Luo, H., & Poeppel, D. (2007). Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. *Neuron*, *54*(6), 1001–1010. https://doi.org/10.1016/J.NEURON.2007.06.004

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *The Journal of Neuroscience Methods*, *164*(1), 177–190.

https://doi.org/10.1016/j.jneumeth.2007.03.024

Matchin, W. G., Hammerly, C., & Lau, E. F. (2017). The role of the IFG and pSTS in syntactic prediction: Evidence from a parametric study of hierarchical structure in fMRI. *Cortex*, *88*, 106–123. https://doi.org/10.1016/j.cortex.2016.12.010

Matchin, W., Hammerly, C., & Lau, E. (2017). The role of the IFG and pSTS in syntactic prediction: Evidence from a parametric study of hierarchical structure in fMRI. *Cortex*, *88*, 106–123. https://doi.org/10.1016/j.cortex.2016.12.010

Meyer, L., Sun, Y., & Martin, A. E. (2020a). "Entraining" to speech, generating language? In *Language, Cognition and Neuroscience* (Vol. 35, Issue 9, pp. 1138–1148). Routledge. https://doi.org/10.1080/23273798.2020.1827155

Meyer, L., Sun, Y., & Martin, A. E. (2020b). Synchronous, but not entrained: exogenous and endogenous cortical rhythms of speech and language processing. In *Language, Cognition and Neuroscience* (Vol. 35, Issue 9, pp. 1089–1099). Routledge. https://doi.org/10.1080/23273798.2019.1693050

Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., & Von Cramon, D. Y. (2002). FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Human Brain Mapping*, *17*(2), 73–88. https://doi.org/10.1002/hbm.10042

Meyer, M., Steinhauer, K., Alter, K., Friederici, A. D., & Von Cramon, D. Y. (2004). Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain and Language*, *89*(2), 277–289. https://doi.org/10.1016/S0093-934X(03)00350-X

Morillon, B., & Baillet, S. (2017). Motor origin of temporal predictions in auditory attention. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(42), E8913–E8921. https://doi.org/10.1073/pnas.1705373114

Nakano, H., Rosario, M. A. M., Oshima-Takane, Y., Pierce, L., & Tate, S. G. (2014). Electrophysiological response to omitted stimulus in sentence processing. *NeuroReport*,

25(14), 1169–1174. https://doi.org/10.1097/WNR.0000000000000250

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in
visual word recognition: New evidence from the English Lexicon Project. *Psychonomic
Bulletin and Review*, *13*(1), 45–52. https://doi.org/10.3758/BF03193811

Nolte, G. (2003). The magnetic lead field theorem in the quasi-static approximation and its use for
magnetoenchephalography forward calculation in realistic volume conductors. *Physics in
Medicine and Biology*, *48*(22), 3637–3652. https://doi.org/10.1088/0031-9155/48/22/002

O'Reilly, J. A., Zhu, J. D., & Sowman, P. F. (2023). Localized estimation of electromagnetic
sources underlying event-related fields using recurrent neural networks. *Journal of Neural
Engineering*, *20*(4), 046035. https://doi.org/10.1088/1741-2552/ACEF94

Obleser, J., & Kayser, C. (2019). Neural Entrainment and Attentional Selection in the Listening
Brain. *Trends in Cognitive Sciences*, *23*(11), 1–14. https://doi.org/10.1016/j.tics.2019.08.004

Oord, A. van den, Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N.,
Senior, A., & Kavukcuoglu, K. (2016). *WaveNet: A Generative Model for Raw Audio*.
http://arxiv.org/abs/1609.03499

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software
for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational
Intelligence and Neuroscience*, *2011*. https://doi.org/10.1155/2011/156869

Pannekamp, A., Toepel, U., Alter, K., Hahne, A., & Friederici, A. D. (2005). Prosody-driven
sentence processing: An event-related brain potential study. *Cognitive Neuroscience,
Journal Of*, *17*(3), 407–421. https://doi.org/10.1162/0898929053279450

Paulmann, S., Jessen, S., & Kotz, S. A. (2012). It's special the way you say it: An ERP
investigation on the temporal dynamics of two types of prosody. *Neuropsychologia*, *50*(7),
1609–1620. https://doi.org/10.1016/j.neuropsychologia.2012.03.014

Penney, T. B. (2004). Electrophysiological correlates of interval timing in the Stop-Reaction-Time

task. *Cognitive Brain Research*, *21*(2), 234–249.

https://doi.org/10.1016/j.cogbrainres.2004.07.005

Perrone-Bertolotti, M., Rapin, L., Lachaux, J. P., Baciu, M., & Lœvenbruck, H. (2014). What is

that little voice inside my head? Inner speech phenomenology, its role in cognitive

performance, and its relation to self-monitoring. In *Behavioural Brain Research* (Vol. 261,

pp. 220–239). Elsevier. https://doi.org/10.1016/j.bbr.2013.12.034

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral

and Brain Sciences*, *27*(2), 169–190. https://doi.org/10.1017/s0140525x04000056

Pickering, M. J., & Garrod, S. (2021). UNDERSTANDING DIALOGUE: LANGUAGE USE AND

SOCIAL INTERACTION. *Applied Linguistics*. https://doi.org/10.1093/applin/amab025

Ragazzoni, A., Di Russo, F., Fabbri, S., Pesaresi, I., Di Rollo, A., Perri, R. L., Barloscio, D., Bocci,

T., Cosottini, M., & Sartucci, F. (2019). "Hit the missing stimulus". A simultaneous EEG-fMRI

study to localize the generators of endogenous ERPs in an omitted target paradigm.

Scientific Reports 2019 9:1, *9*(1), 1–15. https://doi.org/10.1038/s41598-019-39812-z

Reichel, U. D., Mády, K., & Cole, J. (2018). *Prosodic entrainment in dialog acts*.

https://doi.org/10.48550/arxiv.1810.12646

Rimmele, J. M., Morillon, B., Poeppel, D., & Arnal, L. H. (2018). Proactive Sensing of Periodic

and Aperiodic Auditory Patterns. In *Trends in Cognitive Sciences* (Vol. 22, Issue 10, pp.

870–882). Elsevier Ltd. https://doi.org/10.1016/j.tics.2018.08.003

Rosenberg, J. R., Amjad, A. M., Breeze, P., Brillinger, D. R., & Halliday, D. M. (1989). The Fourier

approach to the identification of functional coupling between neuronal spike trains. *Progress

in Biophysics and Molecular Biology*, *53*(1), 1–31. https://doi.org/10.1016/0079-

6107(89)90004-7

Sammler, D., Grosbras, M. H., Anwander, A., Bestelmeyer, P. E. G., & Belin, P. (2015). Dorsal

and ventral pathways for prosody. *Current Biology*, *25*(23), 3079–3085.

https://doi.org/10.1016/j.cub.2015.10.009

Sassenhagen, J., & Fiebach, C. J. (2019). Finding the P3 in the P600: Decoding shared neural mechanisms of responses to syntactic violations and oddball targets. *NeuroImage*, *200*, 425–436. https://doi.org/10.1016/j.neuroimage.2019.06.048

Scharinger, M., Steinberg, J., & Tavano, A. (2017). Integrating speech in time depends on temporal expectancies and attention. *Cortex*, *93*, 28–40. https://doi.org/10.1016/j.cortex.2017.05.001

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, *32*(1), 9–18. https://doi.org/10.1016/J.TINS.2008.09.012

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319. https://doi.org/10.1016/j.cognition.2013.02.013

Snedeker, J., & Casserly, E. (2010). Is it all relative? effects of prosodic boundaries on the comprehension and production of attachment ambiguities. *Language and Cognitive Processes*, *25*(7), 1234–1264. https://doi.org/10.1080/01690960903525499

Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *Journal of Neuroscience*, *32*(25), 8443–8453. https://doi.org/10.1523/JNEUROSCI.5069-11.2012

Sonntag, G. P., & Portele, T. (1998). PURR - A method for prosody evaluation and investigation. *Computer Speech and Language*, *12*(4), 437–451. https://doi.org/10.1006/csla.1998.0107

Stefanics, G., Hangya, B., Hernádi, I., Winkler, I., Lakatos, P. P., & Ulbert, I. (2010). Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. *Journal of Neuroscience*, *30*(41), 13578–13585. https://doi.org/10.1523/JNEUROSCI.0703-10.2010

Stehwien, S., & Meyer, L. (2021). Rhythm Comes, Rhythm Goes: Short-Term Periodicity of

Prosodic Phrasing. *PsyArXiv*. doi:10.31234/osf.io/c9sgb

Steinhauer, K., & Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research*, *30*(3), 267–295. https://doi.org/10.1023/A:1010443001646

Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(32), 14425–14430. https://doi.org/10.1073/pnas.1008662107

Strobel, A., Debener, S., Sorger, B., Peters, J. C., Kranczioch, C., Hoechstetter, K., Engel, A. K., Brocke, B., & Goebel, R. (2008). Novelty and target processing during an auditory novelty oddball: A simultaneous event-related potential and functional magnetic resonance imaging study. *NeuroImage*, *40*(2), 869–883. https://doi.org/10.1016/J.NEUROIMAGE.2007.10.065

Taulu, S., Simola, J., & Kajola, M. (2005). Applications of the signal space separation method. *IEEE Transactions on Signal Processing*, *53*(9), 3359–3372. https://doi.org/10.1109/TSP.2005.853302

Ten Oever, S., & Martin, A. E. (2021). An oscillating computational model can track pseudo-rhythmic speech by using linguistic predictions. *ELife*, *10*. https://doi.org/10.7554/ELIFE.68066

Teoh, E. S., Cappelloni, M. S., & Lalor, E. C. (2019). Prosodic pitch processing is represented in delta-band EEG and is dissociable from the cortical tracking of other acoustic and phonetic features. *European Journal of Neuroscience*, *50*(11), 3831–3842. https://doi.org/10.1111/ejn.14510

Tomasello, R., Grisoni, L., Boux, I., Sammler, D., & Pulvermüller, F. (2022). Instantaneous neural processing of communicative functions conveyed by speech prosody. *Cerebral Cortex*, *32*(21), 4885–4901. https://doi.org/10.1093/cercor/bhab522

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(3), 443–467. https://doi.org/10.1037/0278-7393.31.3.443

van Bree, S., Sohoglu, E., Davis, M. H., & Zoefel, B. (2021). Sustained neural rhythms reveal endogenous oscillations supporting speech perception. In *PLoS Biology* (Vol. 19, Issue 2). https://doi.org/10.1371/journal.pbio.3001142

Van der Burght, C. L., Friederici, A. D., Goucha, T., & Hartwigsen, G. (2021). Pitch accents create dissociable syntactic and semantic expectations during sentence processing. *Cognition*, *212*, 104702. https://doi.org/10.1016/j.cognition.2021.104702

Van Veen, B. D., Van Drongelen, W., Yuchtman, M., & Suzuki, A. (1997). Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering*, *44*(9), 867–880. https://doi.org/10.1109/10.623056

van Vliet, M., Liljeström, M., Aro, S., Salmelin, R., & Kujala, J. (2018). Analysis of functional connectivity and oscillatory power using DICS: From Raw MEG data to group-level statistics in python. *Frontiers in Neuroscience*, *12*(SEP). https://doi.org/10.3389/fnins.2018.00586

Vigneau, M., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., Mazoyer, B., & Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. In *NeuroImage* (Vol. 30, Issue 4, pp. 1414–1432). Academic Press. https://doi.org/10.1016/j.neuroimage.2005.11.002

Volpe, U., Mucci, A., Bucci, P., Merlotti, E., Galderisi, S., & Maj, M. (2007). The cortical generators of P3a and P3b: a LORETA study. *Brain Research Bulletin*, *73*(4–6), 220–230. https://doi.org/10.1016/J.BRAINRESBULL.2007.03.003

Wilson, M., & Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. In *Psychonomic Bulletin and Review* (Vol. 12, Issue 6, pp. 957–968). Springer.

https://doi.org/10.3758/BF03206432

Young, S. R. (1984). RSVP: A task, reading aid, and research tool. In *Behavior Research Methods, Instruments, & Computers* (Vol. 16, Issue 2).

Zaccarella, E., & Friederici, A. D. (2015). Merge in the human brain: A sub-region based functional investigation in the left pars opercularis. *Frontiers in Psychology*, *6*(NOV), 1818. https://doi.org/10.3389/fpsyg.2015.01818

Figure captions

Figure 1. Paradigm and procedure. (A) Paradigm. Prosodic contour repeated 3 times to induce entrainment, followed by time-matched rapid serial visual presentation of the target sentence. FAST entrainment facilitates comprehension of SHORT sentences, SLOW entrainment facilitates comprehension of LONG sentences (Lamekina & Meyer, 2022). (B) Procedure. Participants first listened to the audio contour. Then the sentence was presented word-by-word. In 75% of the trials, a comprehension question followed.

Figure 2. Sensor-level spectra. (A) Coherence spectrum in the entrainment phase. Distinct peaks are visible: 0.9 Hz for FAST condition, and 0.6 and 1.2 (harmonics of 0.6) for the SLOW condition. These peaks correspond to the respective occurrence frequencies of the contours. X-axis limited to 1.5 Hz for display purposes. (B) Power spectrum in the target phase. Peaks visible around frequencies of interest: 0.6 Hz for SLOW and 0.9 Hz for FAST conditions. X-axis adjusted to 3 Hz for display purposes.

Figure 3. Sensor space results. (A) Entrainment phase. Left: coherence values, averaged over participants and sensors (mean ± standard error). Coherence at 0.6 Hz is higher in SLOW condition, while at 0.9 Hz it is higher in FAST condition. Right: Topographic distribution of coherence t-values of the cluster permutation tests. (B) Target phase. Left: power values, averaged over participants and sensors (mean ± standard error). Power at 0.6 Hz is higher in SLOW condition, while at 0.9 Hz it is higher in FAST condition. Right: Topographic distribution of power t-values of the cluster permutation tests (scales adjusted for demonstrative purposes). For MEG sensors layout, see Figure 2-1.

Figure 4. Source space results. (A). Entrainment phase. Bar plots indicate coherence values for ROIs (mean ± standard error) in left and right hemisphere. Entrainment effect (CONDITION × FREQUENCY interaction) was significant in all ROIs, but higher in right STG and EAC. (B). Target phase. Bar plots indicate power values for ROIs (mean ± standard error) in left and right hemisphere. Entrainment effect (CONDITION × FREQUENCY interaction) was significant in all ROIs, but higher in left EAC and FOP. For the complete list of the sub-areas in each ROI, please refer to Table 3-1.

Figure 5. Coherence spectrum in the source space in the entrainment phase for different ROIs. Distinct peaks are visible: 0.9 Hz for FAST condition, and 0.6 and 1.2 (harmonics of 0.6) for the SLOW condition. These peaks correspond to the respective occurrence frequencies of the contours. The coherence values are higher in right STG and EAC, in line with statistical analysis results. X-axis limited to 1.5 Hz for display purposes.

Figure 6. Power spectrum in the source space in the target phase for different ROIs. Distinct peaks are visible: 0.9 Hz for FAST condition, and 0.6 for the SLOW condition. These peaks correspond to the respective occurrence frequencies of the contours. The power values are higher in left FOP and EAC, in line with statistical analysis results.

Figure 7. ERF analysis. (A) Earlier window (0-120 ms). Left: Topographical distribution of the ERF effect; cluster with significant differences highlighted. Right: Grand average ERF wave across the cluster. SLOW condition associates with increased ERF amplitude at onset of missing verb in SHORT sentences relative to FAST condition. Time window of interest (0-120 ms) marked in grey. Middle: source space. Stronger red color indicates areas where SLOW condition induced higher activation than FAST (see also Table 4), while the reverse is indicated by stronger blue color. (B) Later window (300-600 ms). Left: Topographical distribution of omission ERF effect; clusters with significant differences highlighted (positive cluster – black, negative cluster – white). Right: Grand average ERF waves across the clusters of sensors (left – positive cluster, right – negative cluster). For both clusters, SLOW condition associates with increased ERF amplitude at onset of missing verb in SHORT sentences relative to FAST condition. Time window of interest (300-600 ms) marked in grey. Middle: source space. Stronger red color indicates areas where SLOW condition induced higher activation than FAST (see also Table 3), while the reverse is indicated by stronger blue color.

Figure 8. Source space results – entrainment in motor cortex in the target phase. Bar plots indicate power values (mean ± standard error) in left and right hemisphere. Entrainment effect (CONDITION × FREQUENCY interaction) was significant in both motor and pre-motor regions (Brodmann areas 4 and 6).

Table 1. Source space statistics for coherence in entrainment phase.

| | Frequency[1] | | Condition[1] | | Frequency × Condition[1] | | Post-hoc t-tests within condition | | | |
| | | | | | | | Slow | | Fast | |
	F	p	F	p	F	p	T	p	T	p
Left										
EAC	0.008	0.92	6.18	< 0.05	68	<0.001	6	<0.001	-5.86	<0.001
STG	0.05	0.82	5.44	< 0.05	69	<0.001	5.7	<0.001	-5.79	<0.001
IFG	2.41	0.12	5.54	< 0.05	34	<0.001	3.57	0.001	-4.32	<0.001
FOP	1.75	0.18	10	< 0.05	47	<0.001	4.6	<0.001	-5.17	<0.001
Right										
EAC	0.21	0.65	6.75	< 0.05	163	<0.001	7.89	<0.001	-8.84	<0.001
STG	0.04	0.85	9.5	< 0.05	203	<0.001	8.98	<0.001	-11	<0.001
IFG	2.16	0.15	10	< 0.05	68	<0.001	5.42	<0.001	-5.85	<0.001
FOP	<0.001	0.98	9.4	< 0.05	110	<0.001	7.93	<0.001	-6.48	<0.001

[1] Degrees of freedom for all reported LME models were equal to 1 (numerator) and 105 (denominator).

Table 2. Source space statistics for power in target phase.

| | Frequency[2] | | Condition[2] | | Frequency × Condition[2] | | Post-hoc t-tests within condition | | | |
| | | | | | | | Slow | | Fast | |
	F	p	F	p	F	p	T	p	T	p
Left										
EAC	0.94	0.34	482	< 0.001	43	< 0.001	11	< 0.001	-12	< 0.001
STG	0.84	0.36	535	< 0.001	39	< 0.001	15	< 0.001	-12	< 0.001
IFG	1.93	0.17	552	< 0.001	44	< 0.001	13	< 0.001	-11	< 0.001
FOP	1.14	0.34	501	< 0.001	38	< 0.001	14	< 0.001	-11	< 0.001
Right										
EAC	0.04	0.84	518	< 0.001	35	< 0.001	8.87	< 0.001	-10	< 0.001
STG	2.3	0.13	724	< 0.001	51	< 0.001	18	< 0.001	-14	< 0.001
IFG	1.99	0.16	449	< 0.001	31	< 0.001	12	< 0.001	-14	< 0.001
FOP	1	0.32	333	< 0.001	24	< 0.001	10	< 0.001	-13	< 0.001

[2] Degrees of freedom for all reported LME models were equal to 1 (numerator) and 105 (denominator).

Table 3. Regions with significant differences in the ERF 300-600 ms contrast (source space).

	T	df	p
Left			
Area i6-8 (dorsolateral prefrontal cortex)	2.87	34	< 0.05
Area 9m (medial prefrontal cortex)	2.67	34	< 0.05
Area IFSa (inferior frontal cortex)	2.37	34	< 0.05
Area p24 (medial prefrontal cortex)	2.32	34	< 0.05
Area 47l (inferior frontal cortex)	2.29	34	< 0.05
Area 8C (dorsolateral prefrontal cortex)	2.3	34	< 0.05
Area PGs (inferior parietal cortex)	-2.22	34	< 0.05
Right			
Temporal-parietal-occipital junction (area 2)	3.07	34	< 0.05
Frontal eye field	2.98	34	< 0.05
Area i6-8 (dorsolateral prefrontal cortex)	2.48	34	< 0.05

Table 4. Regions with significant differences in the ERF 0-120 ms contrast (source space).

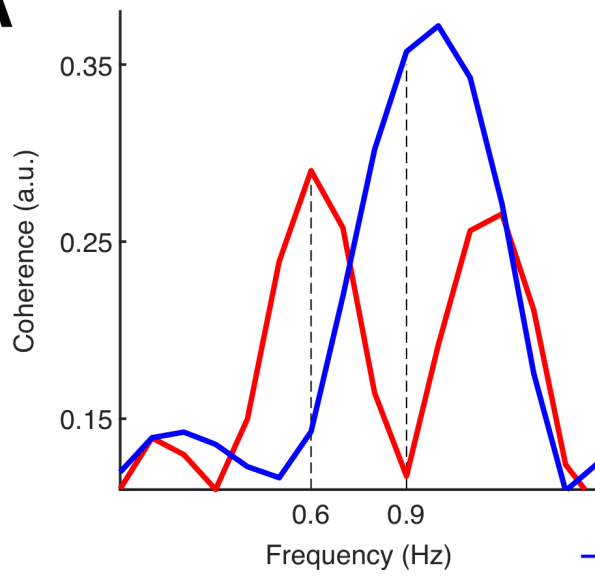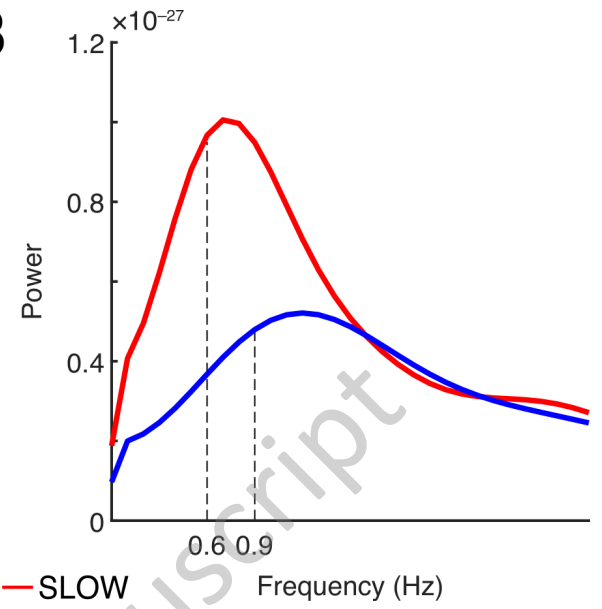|	T	df	p
Left			
Area 9p (dorsolateral prefrontal cortex)	2.73	34	< 0.05
Area PeEc (medial temporal cortex)	-2.38	34	< 0.05
Area 8BL (dorsolateral prefrontal cortex)	2.2	34	< 0.05
Area 8BM (medial prefrontal cortex)	2.13	34	< 0.05
Right			
Middle temporal cortex	2.88	34	< 0.05
Area AAIC (anterior insular cortex)	-2.71	34	< 0.05
Area Pir (anterior insular cortex)	-2.61	34	< 0.05
Area V7 (visual cortex)	-2.22	34	< 0.05
Temporal-parietal-occipital junction (area 2)	2.11	34	< 0.05
Posterior orbitofrontal complex	-1.98	34	< 0.05

Table 5. Source space statistics for power in target phase (motor and pre-motor areas).

| | Frequency[3] | | Condition[3] | | Frequency × Condition[3] | | Post-hoc t-tests within condition | | | |
| | | | | | | | Slow | | Fast | |
	F	p	F	p	F	p	T	p	T	p
Left										
BA4	4.48	< 0.05	732	< 0.001	62	< 0.001	16.8	< 0.001	-11.57	< 0.001
BA6	5.77	< 0.05	751	< 0.001	64	< 0.001	14.19	< 0.001	-12.07	< 0.001
Right										
BA4	4.26	< 0.05	467	< 0.001	41	< 0.001	13.05	< 0.001	-13.65	< 0.001
BA6	3.03	0.085	453	< 0.001	37	< 0.001	12.86	< 0.001	-13.5	< 0.001

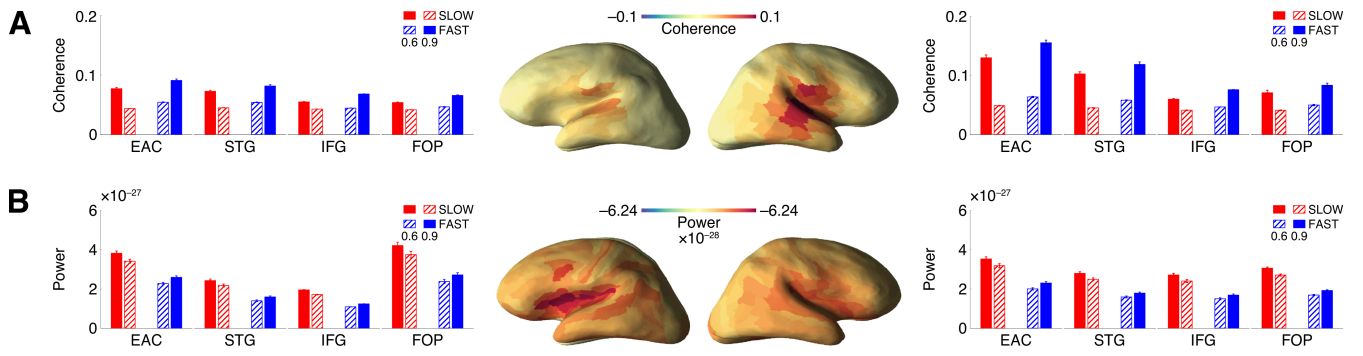[3] Degrees of freedom for all reported LME models were equal to 1 (numerator) and 105 (denominator).

A

FAST
SLOW

Pitch (Hz)
200
100

0   Time (s)   1   × 3

Max sieht Tom und Karl lacht.   **LONG**
Max sieht Tom und Karl.   **SHORT**

Pitch (Hz)
200
100

0   Time (s)   1.5   × 3

Max sieht Tom und Karl lacht.   **LONG**
Max sieht Tom und Karl.   **SHORT**

Entrainment phase   Target phase

B

+   Ma   sie   Tor   un   Ka   lac   Sieht Max Karl?
                    Weit   Weit   Ja   Nein

Audio contour   Rapid word-by-word visual presentation   Question answering

A

Coherence (a.u.)

0.35

0.25

0.15

0.6   0.9

Frequency (Hz)

B

×10⁻²⁷

Power

1.2

0.8

0.4

0

0.6 0.9

Frequency (Hz)

— FAST   — SLOW

A   B

Entrainment phase

EAC STG IFG FOP

Left hemisphere — Frequency (Hz)

Right hemisphere — Frequency (Hz)

SLOW FAST
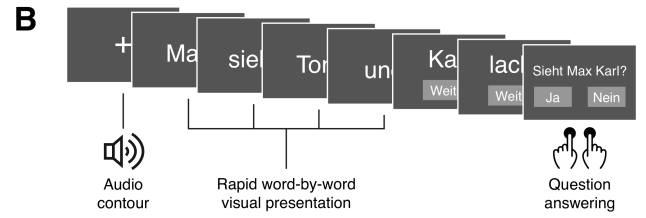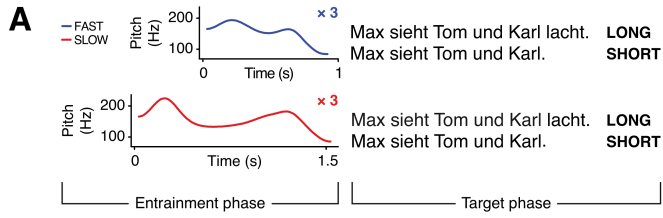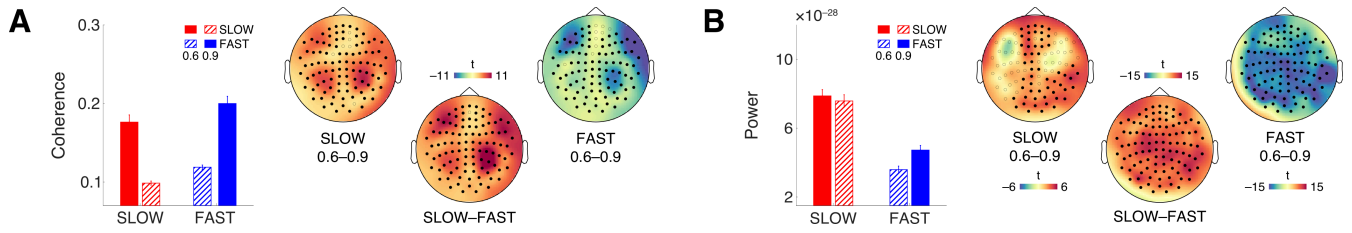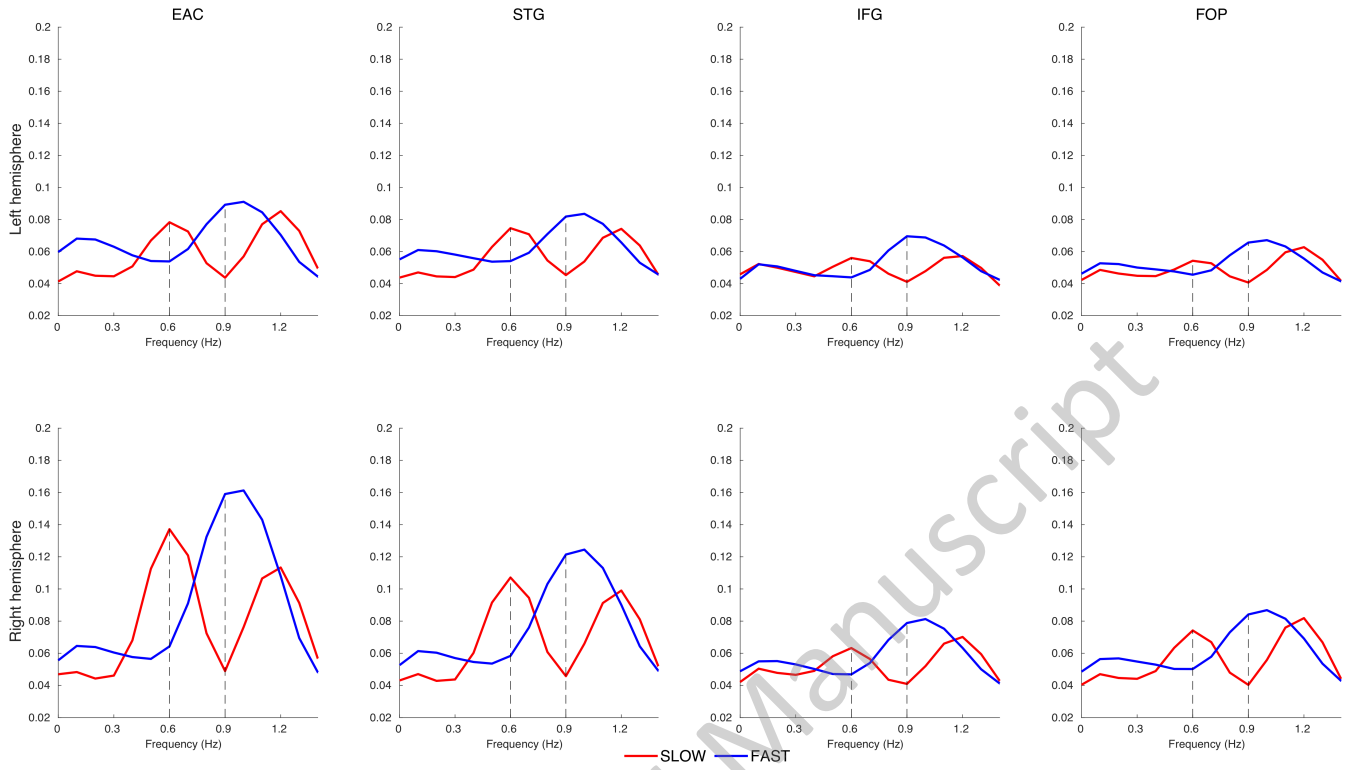
Target phase
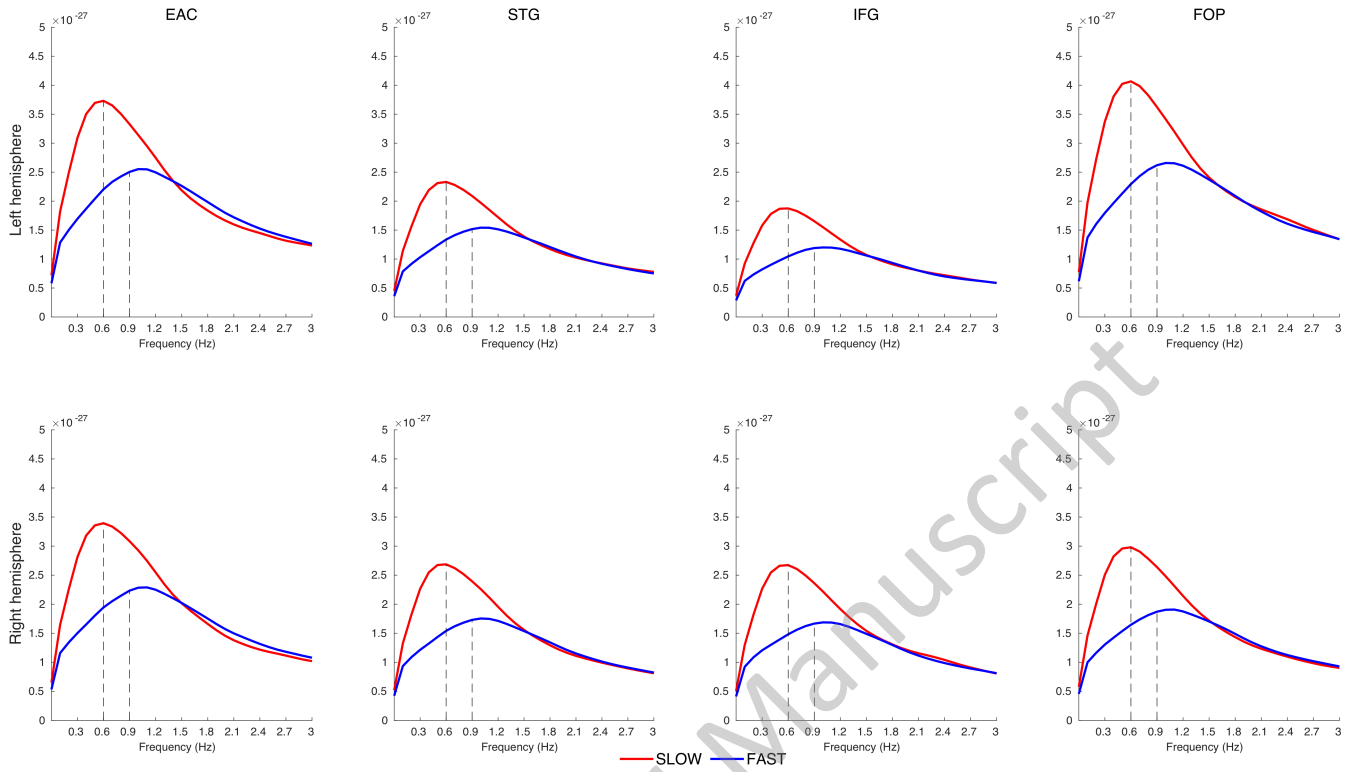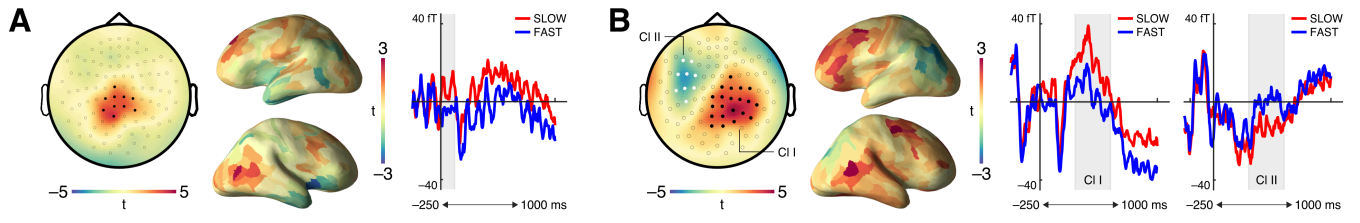
A Left hemisphere

B Right hemisphere